COMPREHENSIVE MICROSYSTEMS

EDITORS-IN-CHIEF

YOGESH B. GIANCHANDANI
OSAMU TABATA
HANS ZAPPE

VOLUME ONE

Editors-in-Chief

Yogesh B. Gianchandani is a Professor in the EECS Department and holds a courtesy appointment in the Department of Mechanical Engineering at the University of Michigan, Ann Arbor. He received a Ph.D. in electrical engineering from the same university in 1994. Prior to this, he worked as an IC designer, primarily at Xerox Corporation. His research interests include all aspects of design, fabrication, and packaging of micromachined sensors and actuators and their interface circuits. He has contributed to more than 200 papers or patents in this field, and serves on the editorial boards of several journals. He also served as a General Co-Chair for the IEEE/ASME International Conference on Micro Electro Mechanical Systems (IEEE MEMS) in 2002. At the University of Michigan, Prof. Gianchandani has served as the director of the College of Engineering Interdisciplinary Professional Degree Program in Integrated Microsystems. As of 2007, he is on a temporary assignment at the US National Science Foundation, serving in the Directorate for Engineering as a program manager for nano- and microsystems.

Professor Osamu Tabata Born in 1956. He received a MS degree and a Ph.D. degree from the Nagoya Institute of Technology, Nagoya, Japan, in 1981 and in 1993, respectively. Since 1981 for 15 years, he has been with the Toyota Central Research and Development Laboratories, Japan. In 1996, he joined the Department of Mechanical Engineering, Ritsumeikan University, Shiga, Japan. He was a guest Professor of Institute of Microsystem Technology, University of Freiburg and ETH Zurich from September to December 2000 and from January to March 2001, respectively. In 2003, he joined the Department of Mechanical Engineering, Kyoto University, Japan. He is currently Professor in the Department of Micro Engineering, Kyoto University. He is currently interested in the establishment of a technology to realize a unique and novel nanosystem by assembling the various functional components such as a microchip, and a particle, a microcapsule, and a cell, with sizes ranging from the nanometer to micrometer scale on a MEMS/NEMS substrate. He termed this concept as SENS (synthetic engineering for nanosystems), and is pursuing experimental and theoretical research on the establishment of SENS. He is an associate editor of Journal of Micro Electro Mechanical Systems as well as a member of the editorial board of Advanced Micro- and Nanosystems, and Sensors and Actuators. Also he is a program committee member of many International Conferences.

Professor Hans Zappe was born in Paris, France, in 1961 and raised in New York. He received his B.Sc. and M.Sc. in Electrical Engineering from the Massachusetts Institute of Technology in 1983 and his Ph.D. in the same field from the University of California at Berkeley in 1989. He has worked at IBM (Burlington, VT, USA) on silicon VLSI, at the Fraunhofer Institute for Applied Solid State Physics (Freiburg, Germany) on GaAs electronics and high-speed lasers, and at the Centre Suisse d'Electronique et de Microtechnique (Zurich, Switzerland) on integrated optical microsystems and surface-emitting lasers. Since 2000, he has been Professor in the Department of Microsystems Engineering at the University of Freiburg, Germany. His current research interests focus on the development of novel tunable micro- optical components, including polymer, membrane, and liquid-based optics, variable photonic crystals, and optical microsystems for medical diagnostics and clinical applications.

Preface

Microsystems have emerged from the laboratory and have become essential components in a wide range of medical and industrial products, and research instrumentation. They include not only microelectromechanical systems (MEMS), but all forms of microsensors, microactuators, and interface microelectronics, deployed as enabling components within a larger system or distributed network. Combining high functionality with small dimensions and reduced power consumption, microsystems benefit from mass-fabrication technologies to allow their manufacture in high volumes and, often, at low cost. The inclusion of electronics, when possible, reduces the cost of deployment and use, providing autocalibration and self-testing. The small dimensional scale sometimes allows physical effects to be leveraged in unconventional ways, providing surprisingly high functionality and performance. As a result, microsystems are virtually invisible to most people, yet have become indispensable in many aspects of their lives.

The microsystems field has expanded to embrace a host of technologies. The well-established discipline of microelectronics has now been joined by micromechanics, microfluidics, and microoptics to allow the fabrication of complex, multifunctional integrated microsystems. As a result, the highly interdisciplinary nature of the subject often makes it difficult for researchers to obtain an overview of the technologies and capabilities available in this established yet dynamically growing engineering field. Thanks to a superb collection of authors, reviewers, and Editorial Advisory Board members, we are confident that *Comprehensive Microsystems* represents an authoritative primary reference source that addresses this need.

As its title suggests, the book before you covers virtually all aspects of the microsystems field. In 54 chapters, the work discusses a breadth of topics, which underscores the interdisciplinary nature of research and development in microsystems and MEMS, a spectrum that has few parallels in other technical disciplines. As a readable reference work, *Comprehensive Microsystems* provides engineers, students, and educators with a unified source of information that will prove to be useful for new as well as established microsystems researchers.

The book is thematically divided into sections that cover a variety of topics; the chapters are self-contained, yet cross-referenced, allowing readers to easily obtain relevant related information. The book opens with a section on Materials, including chapters on silicon as well as metals and polymers, and continues with Fabrication and Packaging, in which the topics range from micromachining to self-assembly and packaging. Moving to an overview of some of the basic components used in microsystems, the section Electronics and System Design covers areas including electronic interface circuits and simulation, and Actuation Mechanisms has chapters on electrostatic, magnetic, and thermal actuation.

The exceptionally wide variety of application areas in which microsystems play a role is reflected in the range of sections that follow. Physical Sensing includes contributions on pressure and flow sensors, gyroscopes, and accelerometers, whereas microfluidics, micropumps, and chemical sensors are only a few of the topics found in the section Chemical and Biological Systems. Finally, the rich spectrum of activities discussed in Optical Systems includes micromirrors, the artificial retina, and biophotonics.

Realizing that the MEMS and microsystems fields have led to mature products in a number of industrial applications as well as provided inspiration for research in unexplored areas, the work concludes with a section Industrial Applications, in which chapters discuss radio-frequency MEMS, medical applications, and ink-jets, and finally Emerging Topics, a look toward the future in which MEMS atomic clocks, microcombustion

systems, and molecular machine-based nanoelectromechanical systems (NEMS) will see extensive development.

Comprehensive Microsystems provides an extensive cross section through engineering science, from fundamental physics and chemistry to complex, cross-disciplinary systems. The editors-in-chief have been privileged to work with some of the world's leading researchers as authors, reviewers, and editors of the chapters that follow. We hope that you, the reader, will find studying the work as rewarding as we have in assembling it.

ACKNOWLEDGMENTS

The Editors-In-Chief gratefully acknowledge the valuable editorial advice provided by the following professors:

Narayan Aluru
University of Illinois
Urbana-Champaign, U.S.A.

Kukjin Chun
Seoul National University
South Korea

Dominique Collard
University of Lille
France

Nico de Rooij
University of Neuchatel
Switzerland

Masayoshi Esashi
Tohoku University
Japan

Long-Shen Fan
National Tsing Hua University
Taiwan

Paddy J. French
Delft University of Technology
The Netherlands

Hiroyuki Fujita
University of Tokyo
Japan

Kazuhiro Hane
Tohoku University
Japan

Christofer Hierold
Swiss Federal Institute of Technology
ETH Zurich, Switzerland

Tom Kenny
Stanford University
U.S.A.

Chang-Jin Kim
University of California
Los Angeles, U.S.A.

Liwei Lin
University of California
Berkeley, U.S.A.

Oliver Paul
University of Freiburg
Germany

Kazuo Sato
Nagoya University
Japan

Shuichi Shoji
Waseda University
Japan

Goran Stemme
Royal Institute of Technology
Sweden

Richard Syms
Imperial College
U.K.

Man Wong
The Hong Kong University of Science and Technology
Hong Kong

Euisik Yoon
University of Minnesota
U.S.A.

Ted Zellers
University of Michigan
Ann Arbor, U.S.A.

Roland Zengerle
University of Freiburg
Germany

Foreword

Microsystems promise to play a pervasive role in enhancing the quality of life during the next several decades. Emerging today in a plethora of products, they represent a powerful weapon for tackling some of society's most pressing problems, including many in health care, environmental quality, homeland security, energy generation, manufacturing competitiveness, and food production. In a very real sense, microsystems represent the culmination of work in three areas, each of which has sparked a revolution of its own: microelectronics, wireless communications, and sensing technology. Bringing these three areas together, microsystems will gather data from the physical world, interpret those data, and then communicate the results via distributed information networks. Increasingly based on nanotechnology, they will control the cars we drive (and monitor the roads and bridges we drive on), enable new generations of high-density information storage and display devices, revolutionize the diagnosis and treatment of debilitating disorders, and provide interfaces to the cellular, molecular, and atomic levels. Just as microelectronics transformed data processing and communications during the past half-century, microsystems are now reaching out to tackle the problems of the nonelectronic world.

Although microelectronics began with the invention of the transistor shortly after World War II, it really took off with the development of the planar process for integrated circuits in the early 1960s. It was quickly apparent that microelectronics was going to permit great progress in the processing of electronic signals, but the world we live in is not electronic. It is mechanical, optical, chemical, and thermal. Thus, researchers began exploring whether the same lithographic processes used to form integrated circuits could also be used to make devices for converting nonelectronic parameters into electronic form (integrated sensors). By the mid-1960s, researchers were realizing visible imaging devices using integrated circuit techniques, and by the late 1960s they were developing cellular probes, pressure sensors, and other devices using the same technology, augmented by selective silicon etching (later christened "micromachining"). The 1970s saw the development of anisotropic micromachining technology and wafer bonding, both necessary for creating the three-dimensional microstructures needed for transducing real-world parameters into electronic signals. They also saw the first efforts to put integrated sensors into high-volume production as the automotive industry sought to meet new exhaust emissions and fuel efficiency requirements. By the late 1970s, the vision of highly integrated digital-output self-testing auto-calibrating "smart" sensors was in place, but neither the supporting technologies nor the marketplace was quite ready for them.

The 1980s saw other important technologies added to the arsenal of sensor processes, including surface micromachining, LIGA, and deep reactive ion etching. The first serious efforts at commercialization began, and conferences dedicated to integrated sensors were launched. By the end of the decade, integrated sensors were becoming known as microelectromechanical systems (MEMS) and were generating great excitement by realizing actuators as well as sensors using the technology. Uncooled infrared sensors, flowmeters, micromotors, and self-testing accelerometers were being reported. The 1990s saw MEMS become a major worldwide effort and proliferate into a number of subfields, including optical-MEMS, inertial-MEMS, RF-MEMS, bio-MEMS, and microfluidics. Integrated gyros, microvalves, scanning surface probes, ink-jet print heads, DNA chips, and optical projection systems provided the focus for new efforts, fueled in part by new materials and processes. The present decade has seen continued proliferation into new materials, rapid growth in chemical and biological devices, and the joining of sensing (and actuation) technology with microelectronics and wireless interfaces.

These developments have taken sensors well beyond simple transducers to application-driven microsystems that solve complete problems.

Comprehensive Microsystems is a remarkably comprehensive and state-of-the-art look at the microsystems field, assembled by editors who are world leaders in microsystems and by subject authors that are expert in its many constituent areas. It begins by reviewing the diverse set of materials used in microsystems, a set that is steadily expanding to take advantage of properties ranging far beyond those found in silicon alone. The critical issues associated with fabrication and packaging are examined next. High throughput and high yield in forming complex three-dimensional microstructures are important, but unlike microelectronics, where complementary metal oxide semiconductors (CMOS) have become dominant, the diversity of microsystem structures has, so far, resisted process standardization. Both fabrication flows and unit processes still involve considerable innovation. Microsystem packaging is also more difficult than in microelectronics because, by definition, many sensors must directly contact the environment they are trying to measure, and for many devices, packaging at the wafer level is essential both for fabrication yield and for operating performance. The complex issues associated with design in the interdisciplinary multiparameter world of microsystems are discussed next, followed by chapters that examine the state of the art in optical, chemical, biological, and physical microsystems. Available techniques for microactuation are discussed, both for use in their own right (e.g., in ink-jet printers, digital light processors) and for sensor self-test. Important emerging topics are then explored, and the book concludes with comments on interface circuitry for MEMS and a look at three high-volume application areas: automotive, medical, and environmental.

Comprehensive Microsystems is an outstanding resource and window on an exciting new frontier that will be key in solving many of the most important problems facing us in the 21st century, improving the quality of life for our children, our children's children, and ourselves.

Kensall D. Wise
Ann Arbor, Michigan

1.01 Silicon and Related Materials

Toshiyuki Tsuchiya, Kyoto University, Kyoto, Japan

Nomenclature

a_0	Interatomic distance (m)
c	Elasticity tensor (Pa)
c_{ij}	(i, j) component of elasticity tensor (Pa^{-1})
d_s	Atomic distance along the slipping direction on slip plane (m)
d_m	Interval of the atom row (m)
E	Young's modulus (Pa)
K	Gauge factor
G	Shear modulus (Pa)
l, m, n	Direction cosine of a given direction with respect to each orthogonal base vector
R	Resistance (Ω)
S	Strain tensor
s	Elastic compliance tensor (Pa^{-1})
s_{ij}	(i, j) component of elastic compliance tensor (Pa^{-1}), in engineering notation

$\mathbf{T}$	Stress tensor (Pa)
T	Temperature (K)
T_{ij}	Component of stress tensor (Pa) in physical tensor notation
T_{λ}	Component of stress tensor (Pa) in engineering notation
α	Coefficient of thermal expansion (K^{-1})
ε	Tensile strain
γ	Surface energy (J)
ν	Poisson's ratio
Π	Piezoresistive coefficient tensor (Pa^{-1})
π_l	Piezoresistive coefficient along longitudinal axis (Pa^{-1})
π_{ijkl}	Component of piezoresistive constant tensor Π (Pa^{-1}) in physical tensor notation
$\pi_{\omega\lambda}$	Component of piezoresistive constant tensor Π (Pa^{-1}), in engineering notation
$\pi_{<lmn>}$	Longitudinal piezoresistive constant along <*lmn*> direction (Pa^{-1})
ρ	Resistivity (Ω m)
ρ_{ij}	Component of resistivity tensor (Ω m) in physical tensor notation
ρ_{ω}	Component of resistivity tensor (Ω m) in engineering notation
σ_{m}	Fracture strength (Pa)
τ_{m}	Shear strength (Pa)

Glossary

BHF Buffered Hydrofluoric Acid
BPSG Borophosphosilicate Glass
CMOS Complementary Metal Oxide Semiconductor
CVD Chemical Vapor Deposition
CZ method Czochralski method, a crystal growth method of silicon
DRIE Deep Reactive Ion Etching
ELTRAN® an SOI wafer fabricated using epitaxial layer transfer technique
FZ method Floating Zone method, a crystal growth method of silicon
HF Hydrofluoric Acid
LPCVD Low-pressure Chemical Vapor Deposition
NSG Nondoped Silicate Glass
PECVD Plasma-enhanced Chemical Vapor Deposition
PSG Phosphosilicate Glass
SCS Single-crystal Silicon
SDB Silicon Direct Bonding
SIMOX an SOI wafer fabricated using separation by implanted oxygen technique
SOI Silicon on Insulator
TEOS Tetraethylorthosilicate, $Si(C_2H_5O)_4$
TMAH Tetramethylammonium Hydroxide
TMB Trimethylborate $(CH_3O)_3B$
TMP Trimethylphosphite, $(CH_3O)_3P$
UNIBOND® an SOI wafer fabricated using wafer separation technique at hydrogen-implanted layer

1.01.1 Introduction

Silicon and related materials, such as crystalline and amorphous silicon (a-Si), silicon alloys, and compounds, are the most important materials in microsystems. The main reason is their use in semiconductor device structures and the strong dependence of fabrication technologies of microsystems on semiconductor microfabrication, such as thin-film deposition, photolithography, and etching (Plummer *et al.* 2000). In addition, silicon also acts not only as an electrical material but also as a thermal and a mechanical material (Petersen 1982). Especially, as silicon has high relative elastic constant and yield strength, many mechanical structures in microsystems are made of silicon.

Single-crystal silicon (SCS) is used in both bulk and surface micromachined structures. In the bulk micromachining process, an SCS wafer is processed using crystallographic anisotropic etching or deep reactive ion etching (DRIE) of silicon to form relatively thick (~500 μm thick) and large (0.5–10 mm^2) structures, such as pressure sensors, inertial sensors, microfluidic devices, and optical mirrors. In the surface micromachining process, relatively thin (1–10 μm thick) and small (50–1000 μm^2) microstructures are fabricated on silicon substrates using silicon-on-insulator (SOI) wafers by sacrificial etching of silicon dioxide or by using bulk silicon wafers with special etching techniques, which combines a high aspect anisotropic plasma etching method and a lateral or isotropic etching method. Inertial sensors

and optical mirrors are applications using surface micromachined SCS structures.

Polycrystalline silicon (polysilicon) used in a gate electrode and as an interconnection in a complementary metal oxide semiconductor (CMOS) device is now one of the main structural materials in surface micromachined devices. Because polysilicon is highly compatible with the CMOS fabrication process, the polysilicon device is used as an integrated inertial sensor, such as an accelerometer and a gyroscope. In addition, polysilicon is used as a sacrificial layer for the fabrication of dielectric structures, such as infrared sensors, pressure sensors, and gas sensors, in which thermal and electrical insulations to the substrate or passivation of the underlying structures is required. a-Si used in thin-film transistors for flat panel displays and solar cells is also used as a sacrificial layer in surface micromachined devices but is rarely used as a structural material because of its low resistance against chemical and mechanical attacks.

In microsystems, silicon is not only a structural material but also a functional material. Thermal-type infrared sensors detect heat due to the absorption of infrared radiation using thermopiles and bolometers, in which silicon is used as a thermoelectric material and a thermistor, respectively. As an optical material, the photoconductive effect and the photovoltaic effect of silicon are applied to sensing devices such as solar cells and photodiodes. In mechanical structures, the piezoresistive effect of silicon is widely employed for strain detection in pressure and force sensors and accelerometers.

Silicon germanium alloy has recently been used in wireless communication devices because of its low power consumption, high speed, and low noise due to its low electrical conductivity. In microsystems, because of the fabrication capability at low temperature, polycrystalline silicon germanium structures are being used these days. Recent CMOS devices can withstand temperatures of up to 520°C using a barrier metal between the aluminum interconnection and the polysilicon gate electrode. Low-stress and small-stress-gradient films can be obtained below 520°C. In addition, polycrystalline germanium (poly-Ge) film is used as a sacrificial layer, because poly-Ge is etched by a hydrogen peroxide solution, which is one reason why germanium is not widely used for semiconductor devices, although germanium has the better electrical properties as a semiconductor material when compared to silicon.

Dielectric compounds of silicon, nitride, and oxide are also widely used in microsystems. Silicon nitride films have been used as a passivation layer in semiconductor devices because of their high stability and inertness, which is also important and useful in microsystems. The deformable membranes of pressure sensors, supporting membranes, or beams of thin-film-type gas sensors are fabricated using silicon nitride films. Nitride films have excellent properties as an optical material and are also used in dielectric mirrors.

Oxide film is used as an excellent electrical insulating material, such as a gate insulating layer and an interlayer insulation of the CMOS device. Oxide is used as a sacrificial layer in the surface micromachining process where silicon structures are employed. Because the thermal conductivity is low, i.e., lesser than silicon nitride, it is used as a membrane support material in a thermal-type infrared sensor.

Thus, silicon and related materials are indispensable for microsystems. In this chapter, the process technologies and properties of these materials, especially SCS, are discussed with priority being given to mechanical properties, which is the most important property in designing, fabricating, and evaluating microsystems.

1.01.2 Single-Crystal Silicon

1.01.2.1 Structure

Silicon occupies 25.8% of the Earth's crust mass, which is the highest mass ratio next to oxygen, and exists mostly as silicate glass materials. Silicon is the material most widely used as a semiconductor. It is a covalent crystal that has a diamond cubic lattice structure, as shown in **Figure 1(a)**. The lattice constant of silicon is 5.43 Å and the distance between the neighboring covalently bonded silicon atoms is 2.35 Å. The unit cell of the cubic diamond lattice structure can be considered as two interleaved face-centered cubic lattices, with their origins in (0, 0, 0) and (1/4, 1/4, 1/4) directions. In the unit cell, eight silicon atoms are arranged at the corners, six at each face center and four at the midpoint between the center and the four corners that are not adjacent to each other. The unit cell contains eight atoms. The filling ratio of the unit cell is only 34%, which is much lower than the 74% of the close-packed structure, which contributes to easy dopant diffusions in the CMOS fabrications. The atom density in the unit cell is $4.995 \times 10^{22}\,\mathrm{cm}^{-3}$, and surface atom density on the (100) surface is $6.78 \times 10^{14}\,\mathrm{cm}^{-2}$.

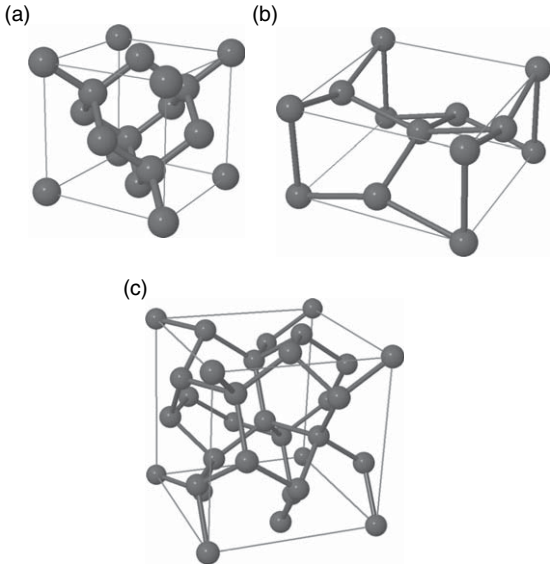


Figure 1 Structures of single-crystal silicon (SCS): (a) diamond cubic (SiI), (b) body-centered tetragonal (β-tin, SiII), and (c) body-centered cubic (BC8, SiIII) structures.

There have been many researches on the high-pressure phases of silicon (Hu *et al.* 1986), because silicon has very high strength against compressive stress. A phase transition from diamond cubic (SiI) to body-centered tetragonal (β-tin), which is a metallic phase (SiII), occurs at about 11 GPa. The β-tin silicon shows a drop of several orders of magnitude in the resistivity. Above 14 GPa, the β-tin phase shows a phase transition to a primitive hexagonal phase. This hexagonal phase changes to the close-packed structure (hexagonal close-packed structure, hcp) at about 40 GPa. The β-tin silicon transfers to body-centered cubic structure (BC8), as shown in **Figure 1(c)**, by quenching at the ambient temperature. The phase transitions from the diamond cubic to β-tin structure and from the β-tin to the BC8 structure are also observed in the microscale by a microindentation. The penetration curve using a spherical indenter shows noncontinuous points during both loading and unloading, which are identical to each of the above-mentioned phase transitions (Williams *et al.* 1999).

1.01.2.2 Crystal Growth

SCS is grown from molten polysilicon, which is reduced from silicon dioxide and purified. There are two major methods for growing single crystals: the Czochralski (CZ) and the floating zone (FZ) method. In the CZ method, which is the most commonly used growth method in the industry, polysilicon granules are put into a quartz crucible and heated in vacuum up to about 1420°C, just above the melting point of silicon. Then, a seed crystal of silicon is dipped into the molten silicon; it is then pulled up quickly at first and then slowly to solidify silicon on the seed crystal by rotating both the seed crystal and the crucible. The dopant impurities can be put into the crucible with polysilicon to control the conductive type and the concentration of the grown silicon ingot. Boron is a p-type dopant and phosphor, antimony, and arsenic are n-type dopants. This method allows a wide range control of electrical conductivity, but it is difficult to obtain a high-resistivity wafer because the impurities (mainly oxygen) accumulate during the crystal growth.

The main source of oxygen impurities in CZ silicon wafers is the crucible. Oxygen in the quartz crucible dissolves into molten silicon and diffuses into the grown silicon ingot. By applying magnetic field to the crucible, the deviation in the dopant concentration as well as the oxygen concentration is reduced and the temperature uniformity is improved because the convective flow in the molten silicon is suppressed. This method is known as the magnetic field applied Czochralski (MCZ) method.

The FZ method is used especially for making high-resistivity wafers, since the maximum resistivity of the CZ wafer is 100–1000 Ω cm because of the impurities (e.g., oxygen). In the FZ method, a polysilicon ingot is placed on top of a seed crystal. The bottom end of the ingot is molten in vacuum or in an argon gas environment using an inductive heating coil. The seed crystal is dipped into the molten polysilicon and the coil is slowly moved upward by rotating both the seed crystal and the polysilicon ingot. The molten silicon solidifies on the seed crystal. This method does not use a crucible and the impurities are segregated into the molten area. Therefore, high-purity and uniformly doped crystal can be grown. High-resistivity silicon wafer up to 20 kΩ cm is possible using this growth method.

The single-crystal ingot is sliced and then lapped and etched to remove the damaged layer during slicing. Finally, the sliced silicon wafer is polished. The crystallographic orientation of the ingot and thus the wafer is determined by the seed crystal. In CMOS device and microsystems, a (100) oriented wafer whose primary orientation flat shows <110> direction is often used. A (111) oriented wafer is used in bipolar transistor devices and microsystem devices. In bipolar transistors, a slightly (~3°)

misoriented (111) wafer is employed to make atomic steps, which act as nucleation sites for epitaxy. In microsystems, a (111) wafer is used for the in-plane isotropy of the elastic constant, which is discussed later. An (110) oriented wafer is used for the high aspect ratio trench fabrication using crystallographic anisotropic wet etching using alkaline solutions.

Another single crystal growth method is epitaxial growth. Vapor-phase epitaxy on SCS substrate using chemical vapor deposition (CVD) is the typical deposition method. Silicon atoms from the pyrolysis of silane (SiH_4) or chlorosilanes deposit on a silicon substrate and single crystal is epitaxially grown on the substrate. *In situ* doping can be easily obtained by mixing phosphine (PH_4) or diborane (B_2H_6) with the source gas during the epitaxial growth. This method is used when an SCS layer of different conductivity type, doping material, or concentration on a silicon wafer is required. For example, the highly boron-doped silicon layer is used for a p^+ etch stop layer in anisotropic etching using an alkaline solution (Seidel *et al.* 1990) or an n-type layer on a p-type substrate is used for the electrochemical etch stop for fabricating bulk-micromachined pressure sensors and accelerometers.

In the surface micromachining process, the term epitaxial polysilicon (epi-poly) is often used for a thick polysilicon layer deposited by means of the same tools as used for epitaxial SCS. This is not an accurate expression but it distinguishes this deposition method from the low-pressure chemical vapor deposition (LPCVD) method, which is the method normally used for polycrystalline silicon thin films.

1.01.2.3 Silicon on Insulator

SOI wafer, where the top SCS layer (device layer) on an oxide film (buried oxide) (Celler and Cristoloveanu 2003) is fabricated on a silicon wafer (handle wafer), is widely used in microsystems as well as in semiconductor devices. In microsystems, the most attractive feature is that the SCS can be used as structural materials in surface micromachined devices. Relatively thick SOI wafers, whose device layer thickness is more than a few microns, are widely used in accelerometers, vibrating gyroscopes, and micromirrors.

The SOI wafer is classified by its fabrication methods, such as silicon direct bonding SOI (SDB-SOI) (Lasky 1986), separation by implanted oxygen (SIMOX) (Izumi *et al.* 1978), cleavage of bonded wafer at hydrogen-implanted layer (UNIBOND®) (Bruel 1995), and epitaxial layer transfer (ELTRAN®) (Yonehara and Sakaguchi 2001). For an SOI with a thick device layer, which is more than a few micrometers thick, an SDB-SOI wafer is most commonly used.

In the SDB-SOI wafer, two silicon wafers, one or both of which are oxidized, are bonded at temperatures higher than 800°C. One silicon layer of the bonded wafer for the device layer is thinned by grinding, etching, and polishing into the designated thickness. The minimum thickness of the device layer is 1 μm, because the best uniformity of the thinning process is ±0.5 μm. The properties, such as the crystallographic orientation, conductive type, dopant material, and the concentration of two SCS layers, can be arbitrarily selected. The crystallinity of the device layer is better than that of other SOI wafers because the device layer is made of a bulk wafer without any irradiation.

The other SOI wafers, on the other hand, are not so often used for microsystems applications because it is difficult to fabricate thick device layers. For the fabrication of the SIMOX silicon wafer, the buried SiO_2 layers were formed by oxygen ion implantation into a silicon wafer. The thickness of the device and buried oxide layers can be determined from the implantation energy and dose. For example, the 150 keV and $1.2 \times 10^{18}\,cm^{-2}$ oxygen implantation and annealing at 1150°C of silicon wafer resulted in the 280-nm-thick silicon layer on the 210-nm-thick oxide layer (Izumi *et al.* 1980). The thickness of the device layer of SIMOX wafers is limited by the ion penetration depth of oxygen.

The fabrication of the UNIBOND SOI wafers is similar to that for SDB-SOI but is different in the device layer thinning process, which is known as the smart cut process. The raw silicon wafer for the device layer is first oxidized and then hydrogen is implanted. The wafer is bonded at the ambient temperature to a wafer for substrate. The bonded wafer is annealed at 400–600°C to split into two wafers at the hydrogen-implanted layer (Bruel 1995). The thickness of the device layer of UNIBOND SOI wafers is limited to 1.5 μm due to the ion penetration depth of hydrogen (Celler and Cristoloveanu 2003).

The ELTRAN SOI wafer is also characterized by the wafer thinning process. A raw silicon wafer of the device layer is first electrochemically etched to form a porous silicon layer as the splitting layer (Yonehara *et al.* 1994). The wafer is bonded to the substrate wafer at the ambient temperature. After heat treatment, the bonded wafer is split using a water-jet method. The device layer thickness of the ELTRAN SOI wafer

depends on the tools used for the epitaxial deposition. Therefore, a thick device layer (~100 μm) is possible, but is not commercially available due to its fabrication cost.

1.01.2.4 Processing

SCS is processed using various techniques in the wafer manufacturing, microfabrication process, micromachining, and assembly process. Even the mechanical processing of silicon is important for microfabrication. For example, lapping using alumina abrasive and chemical mechanical polishing using silica slurry are very important in the wafer fabrication. The chip dicing process using a diamond blade and the laser dicing are used in the packaging process. Wet chemical etching and dry etching processes are the fundamental techniques in microfabrication.

Isotropic wet etching of SCS is done using a mixture of hydrofluoric acid (HF), nitric acid, and acetic acid (Schwarz and Robbins 1976). Because etching is a two-step process, which involves oxidization of silicon by nitric acid and the etching of the oxide by HF, the etching profile can be controlled by the mixing ratio of these acids. In the high HF concentration region, the etched surface is rough, and sharply peaked corners and edges appear because of the autocatalytic mechanism in the oxidization process. When HNO_3 concentration increases, the etched surfaces become smooth and have a rounded shape. In this region, the etching reaction is dominated by oxide removal. In addition, the etch rate of silicon in a solution with the mixing ratio $HF:HNO_3:CH_3COOH = 1:3:8$ decreases steeply as dopant concentration of silicon increases, which can be used for the stiction-free sacrificial layer etching process of SOI structures (Fujitsuka and Sakata 2002).

Silicon is etched with alkaline solution, such as potassium hydroxide (KOH), ammonia (NH_3), and tetramethylammonium hydroxide (TMAH) ($(CH_3)_4NOH$) solution. The etch rate strongly depends on the crystal orientation, and the crystallographic anisotropic etching of silicon is a key technology in the bulk micromachining process. Details are described in Chapter **1.08**.

Electrochemical etching of silicon is another important wet etching method for microsystems. When a positive voltage relative to a reference electrode is applied to silicon in HF aqueous solution with or without ethanol, silicon is etched by anodic dissolution (Smith and Collins 1992). The etch rate is about $1\,\mu m\,min^{-1}$. The etching properties are significantly changed by the etching conditions, such as the current density, etchant concentrations, dopant type of silicon, its concentration, and illumination during etching. At a low current density, highly porous structures with up to 80% porosity are formed, whereas the smooth etching surface is formed at a high current density, which is called electropolishing, and it is used in wafer fabrication (Zhang *et al.* 1989). Since silicon dissolves in HF solution if holes are present, p-type silicon can be etched irrespective of its dopant concentration. In electrochemical etching, lightly-doped n-type silicon needs illumination, because holes may be generated in an n-type silicon electrode by illuminations (photoelectrochemical etching) (Lehmann 1993). The pore size can be controlled from 1 to 20 μm by the dopant concentration. The micrometer-sized pore formation can be applied to periodic structures like photonic crystals by making etch pits on the surfaces before etching (Grüning *et al.* 1995). However, a highly doped n-type silicon can be etched electrochemically (without illumination) because a positively charged depletion layer is formed by the positive bias on the n-type silicon anode. The dependence on n-type dopant concentration can be used for etch stop techniques (Smith and Collins 1992). For example, a thin diaphragm of piezoresistive pressure sensor was fabricated using n^+-silicon wafer with an epitaxially grown n-silicon layer. The 10-μm-thick n-Si layer is used for the pressure diaphragms (Esashi *et al.* 1982).

In dry etching tools, silicon is etched by various kinds of gases, such as fluoride (SF_6, CF_4), chlorine (Cl_2), chloride (CCl_4), and bromide (HBr), using a plasma reactor. The fluorine-based plasma etches silicon isotropically, whereas the chlorine- and bromine-based etchings show anisotropy. The chlorine and bromine etchings need the assistance of ion bombardment rather than fluorine etching. The CMOS industries mainly use the chlorine or bromide etchings because the selectivity to silicon dioxide films is very high. Polysilicon gate electrode can be patterned on thin gate oxide layers. Instead, fluoride etchings (SF_6, CF_4) prior to this etching are required to remove the native oxide layer.

In microsystems, high aspect ratio trench etching by pulsing sulfur fluoride (SF_6) and perfluorocyclobutane (C_4F_8) gases, known as the Bosch process, is widely employed. SF_6 plasma etches silicon isotropically and C_4F_8 deposits a fluorocarbon polymer on the surfaces. The polymer protects the sidewall of the trench during fluoride plasma isotropic etching. Details are described in Chapter **1.07**.

Silicon is etched chemically by a xenon difluoride (XeF_2) gas with an isotropic feature, which can be employed for silicon sacrificial etching in the surface micromachining for dielectric structures. XeF_2 is a white granular crystal and has the equilibrium vapor pressure of about 600 Pa. The etching can be preformed simply by putting silicon and XeF_2 in a chamber. The etching has high selectivity to silicon dioxide, polymer, and metals, which is very useful in the micromachining process. However, this high selectivity to oxide film causes problems in the fabrication because native oxide layer removal prior to this etching is required.

1.01.2.5 Properties of Silicon

Silicon is a group IV semiconductor, whose energy gap is 1.12 eV. The resistivity can be controlled in a wide range by the impurities. The resistivity of the pure silicon is 198 kΩ cm at the ambient temperature. By adding impurities such as phosphorus, boron, and arsenic to 10^{21} cm^{-3}, the resistivity of silicon can be reduced to <1 mΩ cm, as shown in **Figure 2**. Though, the detailed properties as a semiconductor are omitted here, the piezoresistive properties of silicon are described later because of its application to sensor devices.

Silicon is opaque in visible light, but transparent in infrared light with a wavelength greater than 1.1 μm. The refractive index of silicon is 3.874 at a wavelength of 632.8 nm. The properties of crystalline silicon are comprehensively summarized in Hull (1999).

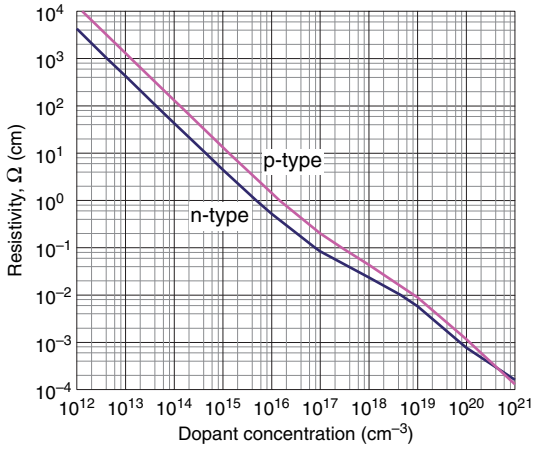


Figure 2 Resistivity of single-crystal silicon (SCS) as a function of dopant concentration.

1.01.2.6 Piezoresistive Properties

A semiconductor strain gauge using SCS has a close connection with the development of the silicon microsystems. The research on the piezoelectric effect, that is, the resistivity change by the applied strain, started in the 1950s and a large gauge factor was found in the semiconductor strain gauges when compared to metal strain gauges (Smith 1954).

The resistance change ΔR in a wire strain gauge is caused by the dimension change of the conductive wire and the piezoresistive effect

$$\frac{\Delta R}{R} = (1 + 2v + \pi_l E)\varepsilon \equiv K\varepsilon \qquad [1]$$

The gauge factor K of silicon is about 100, whereas that of a metal is about 2, which indicates that the piezoresistive effect of silicon is dominant but that of a metal is negligible.

The piezoresistive effect of semiconductor materials is caused by the change in the energy state of the conduction band by the external strain. Then the piezoresistive coefficient depends on the dopant materials and their concentrations (Kanda 1982). Since the energy band structure is anisotropic in accordance with the crystallographic structure, the piezoresistive effect is anisotropic. The first-order piezoresistive effect is described as the change in the component of the resistivity tensor ρ_{ij}

$$\Delta\rho_{ij}/\rho = \sum_{k,l} \pi_{ijkl} T_{kl} \qquad [2]$$

where π_{ijkl} is the component of the piezoresistive coefficient tensor and T_{kl} is the component of the stress tensor. In engineering notation, where the subscripts 11, 22, 33, 23, 13, and 12 are replaced by 1, 2, 3, 4, 5, and 6, respectively, eqn [2] is rewritten as

$$\Delta\rho_{\omega}/\rho = \sum_{\lambda=1}^{6} \pi_{\omega\lambda} T_{\lambda} \qquad [3]$$

The piezoresistive coefficient tensor $\boldsymbol{\Pi}$ of a cubic crystal is given by the following equation:

$$\boldsymbol{\Pi} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{12} & 0 & 0 & 0 \\ \pi_{12} & \pi_{11} & \pi_{12} & 0 & 0 & 0 \\ \pi_{12} & \pi_{12} & \pi_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \pi_{44} \end{pmatrix} \qquad [4]$$

Like elastic constants, there are three independent piezoresistive coefficients, π_{11}, π_{12}, π_{44}. The longitudinal piezoresistive coefficient of arbitral directions (l, m, n) is given by the following equation:

$$\pi_l = \pi_{11} + 2(\pi_{44} + \pi_{12} - \pi_{11})\left(l^2m^2 + m^2n^2 + l^2n^2\right) \quad [5]$$

Thus, the longitudinal piezoresistive coefficient of typical directions is

$$\pi_{<100>} = \pi_{11} \quad [6]$$

$$\pi_{<110>} = \frac{1}{2}(\pi_{11} + \pi_{12} + \pi_{44}) \quad [7]$$

$$\pi_{<111>} = \frac{1}{3}(\pi_{11} + 2\pi_{12} + 2\pi_{44}) \quad [8]$$

The typical piezoresistive coefficients of both p- and n-type silicon are given in **Table 1**. According to eqn [5], the maximum piezoresistive coefficient appears in <111> directions when π_{11} and $\pi_{44} + \pi_{12} - \pi_{11}$ are of the same sign, as is the case in p-type silicon and n-type germanium. If they are opposite in sign, π_l in <111> direction becomes maximum when $\frac{1}{3}|\pi_{44} + \pi_{12} - \pi_{11}|$ is larger than $|\pi_{11}|$. When $\frac{1}{3}|\pi_{44} + \pi_{12} - \pi_{11}|$ is smaller than $|\pi_{11}|$, π_l in <100> direction shows a maximum piezoresistive coefficient.

The mechanism of piezoresistive effect is described by the many valley model. In n-type silicon, the valley of the conduction band, i.e., the minimum energy point, is changed by the external strain. The energy level along the loading axis decreases and that perpendicular to the loading axis increases, which causes the transfer of the electron from the higher energy valleys to the lower ones. As a result, the resistivity along the tensile direction decreases and that along the shear direction increases. On the other hand, the conduction band change in p-type silicon appears in the breakage of the degeneracy at the top (highest energy point) of the valence band, which appears at the large resistivity change against the shear strain.

1.01.2.7 Elastic Properties

The relationship between the stress and the strain of a linear elastic material is described by the Hooke's law

$$\mathbf{T} = \mathbf{cS} \quad [9]$$

The stress **T** and strain **S** are the 2nd rank tensors that have six independent variables. **c** is the elasticity tensor, which is a 4th rank tensor, and is described by a 6×6 matrix. Due to the matrix symmetry, the 21 elastic coefficients are independent. In addition, considering the crystal symmetry of a cubic crystal, only three stiffness coefficients are independent. The elasticity tensor **c** is expressed by engineering notation as:

$$\mathbf{c} = \begin{pmatrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{pmatrix} \quad [10]$$

The stress–strain relationship is also described using the elastic compliance tensor **s**

$$\mathbf{S} = \mathbf{sT} \quad [11]$$

The elastic compliance tensor is described by a 6×6 matrix, which has also only three independent compliance coefficients s_{11}, s_{12}, and s_{44} that can be derived from the stiffness coefficients

$$s_{11} = \frac{c_{11} + c_{12}}{(c_{11} - c_{12})(c_{11} + 2c_{12})} \quad [12]$$

$$s_{12} = \frac{-c_{12}}{(c_{11} - c_{12})(c_{11} + 2c_{12})} \quad [13]$$

$$s_{44} = \frac{1}{c_{44}} \quad [14]$$

These stiffness and compliance coefficients for silicon are accurately measured and given in

Table 1 Piezoresistive coefficients of silicon and germanium

		π_{11}	π_{12}	π_{44}	Maximum π_l
	Resistivity (Ω cm)	($\times 10^{-12}$ Pa^{-1})			
n-Type Si	11.7	−1022	+534	−46	−1020 [100]
p-Type Si	7.8	+66	−11	+1381	+930 [111]
n-Type Ge	9.9	−47	−50	−1379	−970 [111]
p-Type Ge	1.1	−37	+32	+967	+650 [111]

Table 2 Stiffness and compliance coefficients of silicon and germanium

	s_{11}	s_{12}	s_{44}	c_{11}	c_{12}	c_{44}
	($\times 10^{-12}$ Pa^{-1})			(GPa)		
Si	7.68	−2.14	12.6	165.7	63.9	79.6
Ge	9.64	−2.60	14.9	129.2	47.9	67.0

Table 2 (Wortman and Evans 1965). Since these coefficients are related to the principal axis, the elastic constants for a particular orientation should be obtained using the tensor transformation. The elastic constants with respect to an arbitral orientation are

$$\frac{1}{E} = s'_{11} = s_{11} - 2\left(s_{11} - s_{12} - \frac{1}{2}s_{44}\right)\left(l^2m^2 + m^2n^2 + l^2n^2\right) \quad [15]$$

$$v = \frac{s'_{12}}{s'_{11}} = -\frac{s_{12} + \left(s_{11} - s_{12} - \frac{1}{2}s_{44}\right)\left(l^2l'^2 + m^2m'^2 + n^2n'^2\right)}{s_{11} - 2\left(s_{11} - s_{12} - \frac{1}{2}s_{44}\right)\left(l^2m^2 + m^2n^2 + l^2n^2\right)} \quad [16]$$

$$\frac{1}{G} = s'_{44} = s_{44} + 4\left(s_{11} - s_{12} - \frac{1}{2}s_{44}\right)\left(l^2l'^2 + m^2m'^2 + n^2n'^2\right) \quad [17]$$

where (l, m, n), (l', m', n') are the direction cosines of the loading direction and a direction perpendicular to the loading direction. Figures 3–5 show the Young's modulus E, the Poisson's ratio v, and the shear modulus G, respectively, of (100) and (110) silicon and germanium. In the (100) plane, the maximum value of 168.9 GPa appears along [011] and its equivalent directions and the minimum value of 130.2 GPa appears along [001] and its equivalent directions. The Young's modulus E, the Poisson's ratio v, and the shear modulus G of (111) silicon, 168.9 GPa, 0.262, and 66.9 GPa, respectively, do not depend on the in-plane direction, which is useful when a circular or cylindrical structure is used. In addition, the composite elastic moduli $E/(1-v)$ of (100) and (111) planes are 180.5 and 229.0 GPa, irrespective of the in-plane direction, which is very useful since it makes the deformation shape of a thin membrane structure symmetric. The calculated elastic constants of the major orientations are given in Table 3 (Brantley 1973).

The temperature coefficients of the stiffness coefficients are around −100 ppm per degree Celsius in all three independent coefficients (Hall 1967). The doping dependence of the elastic constants is expected because the elastic properties of semiconductors depend on the carrier concentration. Heavily doped (up to 5×10^{19} cm^{-3}) n-type and p-type silicon show 1–3% decrease in the stiffness coefficients and modification of the temperature dependencies (Hall 1967).

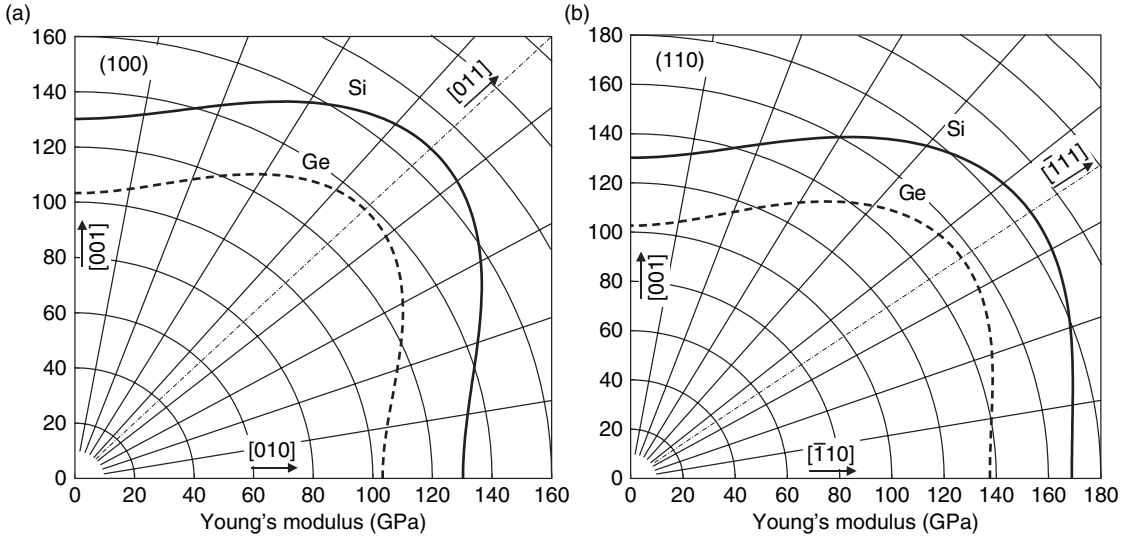


Figure 3 Young's modulus of silicon and germanium as a function of direction in (a) (100) plane and (b) (110) plane. Reprinted with permission from Wortman J J, Evans R A Young's modulus, shear modulus, and Poisson's ratio in silicon and germanium. *J. Appl. Phys*. **36** 153–6. Copyright © 1965, American Institute of Physics.

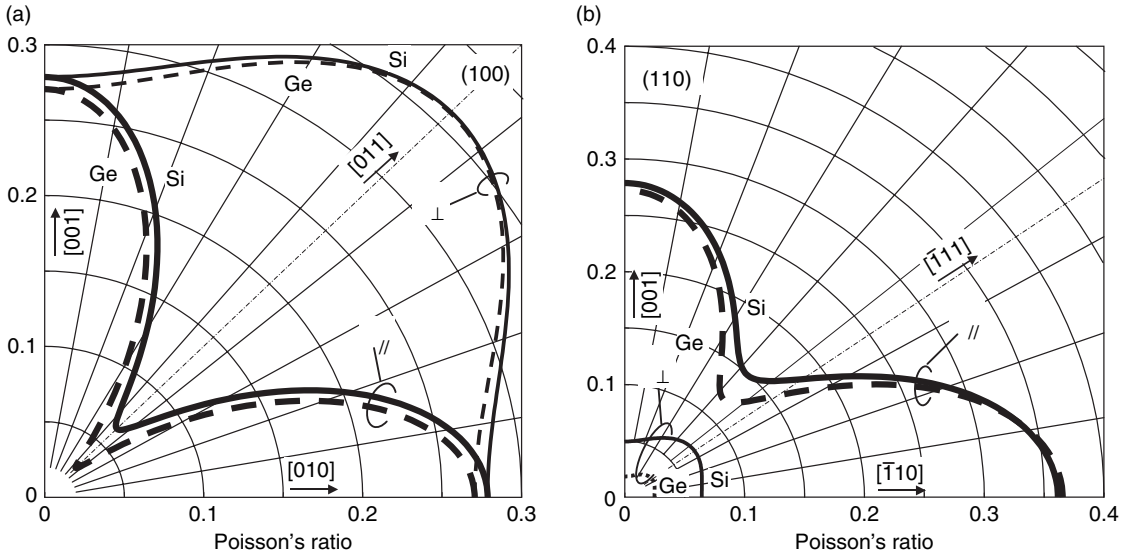


Figure 4 Poisson's ratio of silicon and germanium as a function of direction in (a) (100) plane and (b) (110) plane. $v_{\parallel}$ and $v_{\perp}$ denote the in-plane and perpendicular-to-plane Poisson's ratio, respectively.

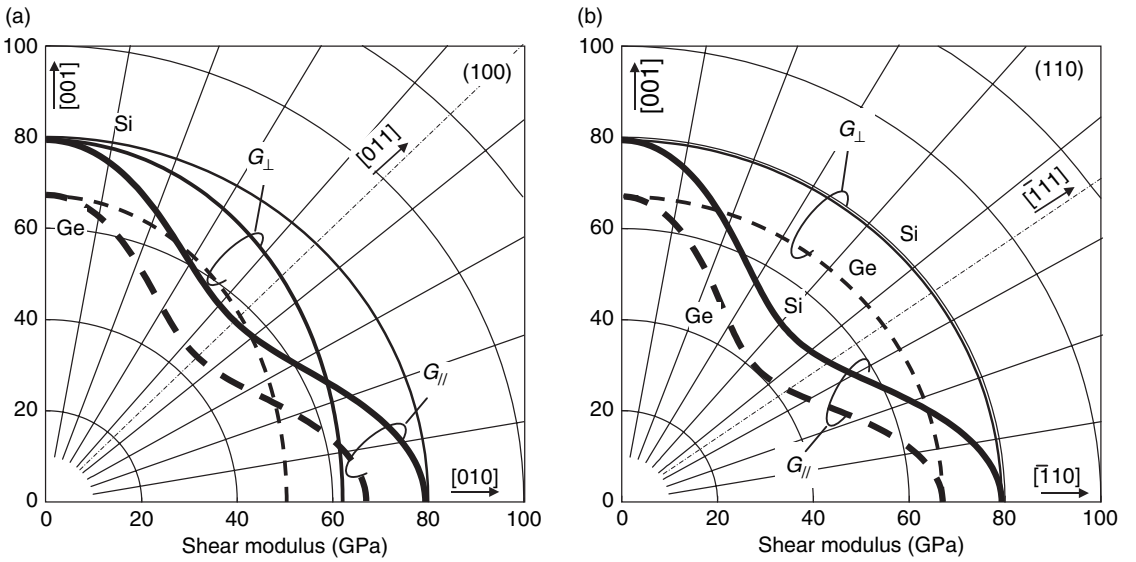


Figure 5 Shear modulus of silicon and germanium as a function of direction in (a) (100) plane and (b) (110) plane. $G_{\parallel}$ and $G_{\perp}$ denote in-plane and perpendicular-to-plane shear modulus, respectively.

Table 3 Bulk silicon elastic modulus for several crystallographic orientations

	Young's modulus		*Poisson's ratio*		
Wafer orientation	*Direction*	*GPa*	*Direction*		
(100)	<011>	168.9	<011>	<0−11>	0.064
(100)	<001>	130.2	<010>	<001>	0.279
(110)	<111>	187.5	<1−11>	<1−1−2>	0.182
(110)	<100>	130.2	<001>	<1−10>	0.279
(111)	(Any)	168.9	(Any)		0.262

In microsystems, there is no reason to negate the use of bulk elastic properties because the structures in microsystems are thick and wide enough to be considered as bulk structures. In addition, many works on the measurement of elastic properties of micrometer specimens using uniaxial tensile test have been carried out, which support the use of bulk properties. For thick (19 μm) SCS (100) films, the measured values of Young's modulus and Poisson's ratio of <110> direction are 169.0 GPa and 0.069, respectively (Isono *et al.* 2006). Though thin (~2 μm thick) SCS films show larger variation in measured values, which may be caused mainly by the variation in the thickness of the specimen, the averaged values among different measurement methods agree with the theoretical values (Tsuchiya *et al.* 2005). The Young's modulus of different crystallographic orientations and directions fabricated from SOI wafers is measured, which indicates close agreement with the bulk properties (Sato *et al.* 1998). The Young's modulus of heavily doped (about $10^{20}\,\mathrm{cm}^{-3}$) silicon microstructures sometimes showed values less by 10% or more than bulk values, which is not consistent with the above discussion. The defects introduced by the dopant may cause the reduction of the elastic coefficients.

In thin-film microstructures, the internal (residual) stress is one of the important properties of the materials. The internal stress consists of the thermal stress and the intrinsic stress. The former is generated by the mismatch of the coefficient of thermal expansion (CTE) between the film and the substrate. The latter mainly originates from the reaction of the film formation process. The internal stress of SCS is almost negligible. However, many researches showed that the doping of SCS causes internal stresses and there are several different results and different explanations for the source of the internal stress. The boron-doped membrane supported by a silicon substrate showed a small tensile stress of about 50 MPa for a membrane 2–3 μm thick. This stress is explained not only by the difference in the CTE ($1.1 \times 10^{-8}\,\mathrm{K}^{-1}$) but also by the difference in the temperature coefficient of the {100} normalized biaxial elastic modulus of the membrane ($-3.7 \times 10^{-5}\,\mathrm{K}^{-1}$) (Berry and Pritchet 1991).

A phosphorus-implanted silicon film showed tensile stress, which is attributed to the lattice dilation of phosphorus. The measured dilation constant, i.e., the ratio of the lattice strain to the dopant concentration, was $-4.5 \times 10^{-24}\,\mathrm{cm}^3$ (Lee *et al.* 1999). However, phosphorus and arsenic doping cause compressive stress in polysilicon, which is attributed to the mass and the size of the substituting ions (Orpana and Korhonen 1991).

In addition, the internal stress in the device layer of bonded SOI wafer exists near the interface to the buried oxide layer (Tan *et al.* 2003), which is caused by the oxidization process of the SCS wafer and wafer bonding process (Iida *et al.* 2000). This internal stress causes small upward deflection of the cantilever beams (Park *et al.* 2005).

1.01.2.8 Strength

The theoretical tensile strength of a perfect crystal is described in Orowan's (1948) theory, in which it is assumed that the work applied on a unit area during the fracture test is equal to the energy needed to create the unit area of two fracture surfaces

$$\sigma_{\mathrm{m}} = \sqrt{\gamma E / a_0} \qquad [18]$$

where γ is the surface energy, E is the Young's modulus, and a_0 is the interatomic distance. As for the (111) plane of SCS, the theoretical tensile strength is 32 GPa. The theoretical shear strength was evaluated by Frenkel (1926).

$$\tau_{\mathrm{m}} = (d_{\mathrm{s}}/d_{\mathrm{m}})(G/2\pi) \qquad [19]$$

where d_{s} is the atomic distance along the slipping direction, d_{m} is the distance between the row of atoms, and G is the shear modulus. As for the (111) plane of silicon, the theoretical shear strength becomes 13.7 GPa. The ratio of the shear strength to the theoretical tensile strength shows the tendency toward the fracture behavior whether it fractures by cleavage or slip. Silicon whose ratio is about 2.3 shows a cleavage fracture, because the maximum shear stress in the specimen is not less than a quarter of the tensile stress during the uniaxial tensile test.

The actual tensile strength is much smaller than the theoretical strength and has large variations, because the fabricated structure has many imperfections such as defects and surface roughness. The reported strength of SCS ranged from 0.5 to 10 GPa, which shows that there is a fracture initiating flaws whose length is >0.5 μm (Tsuchiya *et al.* 2005). In silicon microstructures, the surface geometry is considered to be the dominant parameter of the fracture strength. For example, the strength of the micromachined structure fabricated from SOI wafer using DRIE processes was dominated by the roughness of the etched surfaces (Izumi *et al.* 2005). Such

roughness can be removed or reduced by surface oxidization and by stripping away the oxide layers. The bending strength of the bulk micromachined cantilever beams improved from 6.1 to 6.6 GPa after this process (Ericson and Schweiz 1990). In addition, the loading direction should be considered. For example, the bending strength of the cantilever beams depends strongly on the loading direction. By vertical loading, the maximum stress appears at the top or at the bottom surface, which is often a polished surface (Wilson *et al.* 1996), whereas in the case of the lateral loading it appears on the sidewall surfaces, which is an etched surface (Wilson and Beck 1996).

Nowadays, the fatigue properties of SCS are being investigated widely, though it had been considered that silicon microstructure does not fracture by fatigues in the early stage of microsystems researches. In bulk micromachined devices, it was reported that the SCS cantilever beams did not fracture by fatigue after 10^7 cycles loading if it was not fractured at the first loading (Mutoh *et al.* 1990). Chen and Leipold (1985) reported that the bending strength of SCS wafer does not depend on the loading rate. These phenomena are accounted for by the fact that the defects in silicon crystal do not move at the ambient temperature.

On the other hand, the fracture by repeated stress was observed in the precracked bulk micromachined cantilever beam resonators (Connally and Brown 1992) and the loading rate effect on the bending strength of SCS cantilever beams was reported (Goto *et al.* 1996). The effect of humidity was demonstrated in both the cantilever beam (Connally and Brown 1992) and the doubly supported beam resonators (Tsuchiya *et al.* 1998a). Considering that the fracture initiates at the surface defects, the fatigue fracture is caused by the growth or widening of defects during the long-term cyclic stress application (Muhlstein *et al.* 2001a). It is well agreed that the water vapor affects or initiates the defect growth, which is discussed in detail in Section 1.01.3.

1.01.2.9 Thermal Properties

Silicon also has superior properties as a heat conducting material. At the ambient temperature, the specific heat is $713\,\mathrm{J\,kg^{-1}\,K^{-1}}$, the thermal conductivity is $156\,\mathrm{W\,m^{-1}\,K^{-1}}$, and the thermal diffusivity is $8600\,\mathrm{m^2\,s^{-1}}$. The dopant effect on the thermal conductivity was investigated for phosphor- and boron-doped silicon, both of which have a thermal conductivity of 50–70 $\mathrm{W\,m^{-1}\,K^{-1}}$ in a dopant concentration of $10^{20}\,\mathrm{cm^{-3}}$ (Slack 1964).

The CTE at the ambient temperature (300 K) is $2.616 \times 10^{-6}\,\mathrm{K^{-1}}$. It was measured in the wide temperature range and fitted to the following equation:

$$\alpha = 3.725 \times 10^{-6}\{1 - \exp[-5.88 \times 10^{-3}(T - 124)]\} + 5.548 \times 10^{-10}\,T \quad [20]$$

This equation is valid for the temperature range from 120 to 1500 K. This equation is important for the thermal stress of thin films deposited on a silicon substrate (Okada and Tokumaru 1984). The CTE is slightly changed by doping. Berry and Pritchet (1991) showed that heavy boron doping increases the CTE by $1.1 \times 10^{-8}\,\mathrm{K^{-1}}$.

1.01.3 Polycrystalline and Amorphous Silicon

1.01.3.1 Deposition

Polycrystalline silicon (polysilicon) and a-Si are mainly deposited by the pyrolysis of SiH_4 or disilane (Si_2H_6) gas in a CVD reactor. Polysilicon films deposited by the standard LPCVD process using SiH_4 gas at 630°C usually have (110) oriented columnar grain structures. Amorphous film can be deposited at 580°C or at a lower temperature. Since the deposition rate of the amorphous film by LPCVD using SiH_4 is too low, Si_2H_6 gas is often used for a-Si deposition. The amorphous film can be crystallized by annealing at 600°C or at a higher temperature. The grain size is controlled by annealing temperature; the lower the annealing temperature the larger the grain size because the nucleation rate is lower than the grain growth rate at the low annealing temperature. Crystallization at 600°C for 24 h results in polysilicon films with few micrometer grains.

Doping can be done by the same method as used for SCS, such as in ion implantation and vapor-phase diffusion. In addition, *in situ* doping during LPCVD by adding the dopant source, such as PH_4 and B_2H_6, can be used. In the surface micromachined polysilicon structures, the solid-phase diffusion from the phosphosilicate glass (PSG) films has often been employed for stress gradient control. In this case, the polysilicon film is often sandwiched by PSG layers to dope phosphor uniformly.

For thicker polysilicon films, the deposition of silicon dioxide layers using epitaxial growth tools is used. The plasma-enhanced chemical vapor

depositions (PECVDs) and sputter depositions are employed because of their lower deposition temperature mainly for sacrificial layer formation in the microsystems. For example, a PECVD a-Si film was deposited on a silicon wafer with an SiO_2 layer. Another oxide layer and vanadium oxide (VO_2) as sensing material were deposited on it and patterned to membrane structures. The sacrificial a-Si film was removed using XeF_2 gas (Chung *et al.* 2005).

1.01.3.2 Etching

Etching of polysilicon and a-Si can be done by the etching method used for SCS. Polysilicon is most widely used as structural materials for integrated sensor devices, such as pressure sensors, accelerometers, and gyroscopes. Silicate glass materials, such as nondoped silicate glass (NSG), PSG, and borophosphosilicate glass (BPSG) deposited by CVD, are used as the sacrificial layers that are removed using HF solution to release the polysilicon structures. Polysilicon films can be used as the sacrificial layers in pressure sensors and in infrared image sensors in which dielectric microstructures are employed. Polysilicon sacrificial layer is used not only for the thin cavity of capacitive pressure sensors, but for pyramidal grooves of piezoresistive pressure sensors and IR sensors. To make the pyramidal grooves in a single-crystal wafer, the sacrificial polysilicon layer is directly deposited on the wafer and the polysilicon layer is used in the undercutting layer. In sacrificial layer etching, alkaline solutions such as KOH and TMAH are widely used. But in integrated sensor devices the passivation of aluminum for electrodes is crucial. The ethylenediamine pyrocatechol (EDP) solution and silicon-dissolved TMAH solution have high selectivity against aluminum. In addition, due to the prevention of stiction, XeF_2 is examined for the sacrificial etching of a silicon layer.

1.01.3.3 Elastic Properties

The mechanical properties of polysilicon can be equated with that of SCS considering its crystal orientation and the grain boundary effect. As for the orientation, the standard polysilicon film deposited by LPCVD at 630°C has (110) orientation and the polysilicon film crystallized from the LPCVD a-Si has (111) orientation. The grain boundary has nonideality of the crystal structure, which is the source of mechanical property variations. In addition, since the impurities such as dopant and oxygen tend to segregate at the grain boundary, the effect of the grain boundary layer should be considered.

The calculated Young's moduli of randomly oriented, (110) surface-oriented, and (111) surface-oriented polysilicon are 160.3, 165.3, and 168.9 GPa, respectively. The Young's modulus was experimentally measured using tensile tests (Sharpe 2003) and bulge tests (Maier-Schneider *et al.* 1995). Most of the measured Young's modulus of polysilicon films agreed with these theoretical values. Maier-Schneider (1996) showed that the Young's modulus of LPCVD polysilicon films ranged between the theoretical values of <111> and <100> oriented polysilicon. However, sometimes the reported values of heavily doped polysilicon were lower than these calculated values because of the grain boundary effect. For example, the foundry-processed polysilicon films (multi-user MEMS processes (MuMPs), which has 2–4 mΩ cm in resistivity, shows a lower value (Sharpe *et al.* 1999). The heavily doped polysilicon films often have the oxide segregation at the grain boundary. This layer decreases the apparent Young's modulus, since silicon dioxide has a much smaller Young's modulus (70 GPa or less).

The internal stress of polysilicon is generated and its control is crucial for the surface mircromachined structure. The average internal stress and its relationship with the internal stress have been investigated since the beginning of the semiconductor fabrication process development (Adamczewska and Budzyński 1984). The standard polysilicon film and the amorphous film deposited using LPCVD have a compressive stress of about 300 MPa, because of the disorder of the crystal grain during deposition and oxygen impurities (Kamins 1990).

After the surface micromachining process has been proposed, the stress control becomes more important in order to obtain a low tensile stress and a low stress gradient film. For the low tensile stress films, high temperature anneal is the very effective method because thermal stress is not generated (Guckel *et al.* 1989, Huber *et al.* 1989). However, the CMOS backend integration limits the use of such high-temperature anneal that causes the change in the electrical properties of the transistors. For lower thermal budget in the postsurface micromachining process, a number of low-stress polysilicon processes were developed.

An as-deposited tensile stress film can be obtained at the lower deposition temperature (580–600°C) using LPCVD (Krulevitch *et al.* 1991). Though the temperature range of the tensile stress films is slightly changed by the deposition conditions, films deposited at either lower or higher temperature ranges have

compressive stress (Yang *et al.* 2000). At the temperature range, the amorphous film with compressive stress was deposited initially and was crystallized while the rest of the film was deposited. The tensile stress evolution due to crystallization was observed during postannealing of the amorphous films (Miura *et al.* 1992). The change in the stress was 1 GPa and the crystallization temperature depends on the dopant concentrations (Miura and Okamoto 1994).

Doping effects on the internal stress of polysilicon were widely investigated and were basically the same as that of SCS. The boron doping causes tensile stress and phosphorus and arsenic doping cause compressive stress (Orpana and Korhonen 1991). However, the phosphorus ion implantation at high concentration ($10^{19}\,cm^{-3}$) results in the tensile stress films. Since an as-implanted film becomes amorphous by ion bombardment, the tensile stress is induced by subsequent crystallization (Maier-Schneider 1996). Moreover, polysilicon films heavily ion implanted with both boron and phosphorus became more compressive because of the dopant segregations at the grain boundary (Orpana and Korhonen 1991). Thus, it is difficult to forecast the stress actually generated, because the stress is also changed by the crystal growth and by the densification during the doping and annealing processes.

The stress distribution along the thickness direction, which is often called the stress gradient, causes the out-of-plane deflection. For example, a cantilever beam 100 μm long and 2 μm thick shows 1-μm deflection at the free end if it has stress gradients of 30 MPa μm^{-1}. Since the stress gradient and the deflection have a linear relationship, stress gradient control often becomes a problem in the process development.

If there is no limitation in the thermal budget, which means the micromachined structure does not integrate to circuit, solid-phase doping from PSG layer is a convenient method. It is applied in the MuMPs process. Analog devices Inc. reported low-stress film in the postmicromachining process using the doping by high current ion implantation and crystallization by rapid thermal annealing (Core *et al.* 1993). Polysilicon film became amorphous by ion implantation and the rapid thermal annealing results in uniformly fine-grained polysilicon films. Lee *et al.* (1999) demonstrated the multilayered polysilicon film that is fabricated by the interleaved deposition of the low-stress film and compressive film. The low stress is obtained by averaging two different stress layers. The low-stress distribution was realized by the stress relaxation of the interface between the two different stress layers.

1.01.3.4 Strength

The strength of polysilicon films can also be equated with that of SCS. A number of experiments about polysilicon strength showed that the crack was propagated in the grains but not on the grain boundaries and the little effect of the doping. However, heavy doping may affect the surface roughness through dry etching and sacrificial etching. Heavy doping causes oxide precipitation in the grains, which acts as the mask during anisotropic dry etching, and causes residues. In sacrificial etching, HF selectively etches the grain boundaries where the dopant or oxygen precipitates. It was reported that polysilicon films fractured at the small pit on the specimen edge that should be related to the grain boundaries of the films (Tsuchiya *et al.* 1998b).

Fatigue, the decrease in strength by the cyclic stress application, was also observed using bending and tensile testing either at nonresonance or resonance (Brown *et al.* 1997). The fatigue properties were strongly affected by the humidity of test environments. A mechanism of polysilicon fatigue is proposed as reaction layer fatigue. In this theory, tiny cracks are generated and grow in the thickening native oxide layer. First, the tiny cracks are generated in the native oxide layer. Through these cracks water diffuses more easily and the oxide layer thickens. Then the crack growth and subsequent oxide thickening is repeated until they fracture the specimen (Muhlstein *et al.* 2001b, Pierron and Muhlstein 2005). The native oxide layer is so thin that no fatigue fracture is observed in bulk-scale specimens. Fatigue fracture can be controlled by surface passivation using self-assembled monolayers (Muhlstein *et al.* 2002).

1.01.3.5 Thermal Properties

The thermal conductivity of polysilicon films was measured using microstructures. The thermal conductivity of the heavily doped LPCVD polysilicon films was measured using doubly supported beams. The measured values range from 29 to 34 $W m^{-1} K^{-1}$ (Tai *et al.* 1988). An n-type heavily doped polysilicon film in a CMOS IC process has thermal conductivity of 25 $W m^{-1} K^{-1}$, which is half of that of doped SCS (von Arx *et al.* 2000, Völklein and Baltes 1992). The grain boundaries may cause phonon scattering, which reduces the thermal conductivity of polysilicon.

1.01.4 Silicon Germanium

Germanium has a cubic diamond structure like silicon but has a slightly larger lattice constant of 5.66 Å. Their alloy also has a cubic structure and the lattice constant is in between those of the two materials. Silicon germanium has higher carrier mobility for application in high-frequency devices (King *et al.* 1990) and thin-film transistors (King and Saraswat 1994). In addition, the strained silicon film is fabricated because of the mismatch of the lattice constant, which is also used for radio frequency (RF) applications (Yeo *et al.* 2000). For its application in microsystems, silicon germanium was used in a CMOS integrated device due to the low thermal budget, which is compatible with the aluminum-deposited structures (Sedkey *et al.* 2002).

1.01.4.1 Deposition

LPCVD process is widely used for the deposition of silicon germanium films. Germane (GeH_4) for Ge and SiH_4 or Si_2H_6 for Si are used as the source gas. The crystallization temperature of LPCVD polysilicon using SiH_4 is 600°C, and the films deposited below this temperature becomes amorphous and the deposition rate is too low (less than 3 nm min^{-1}) for the industries. When GeH_4 is added to the source gas, the crystallization temperature decreases and the deposition rate increases by the catalysis of Ge. A pure poly-Ge film can be obtained even at 340°C; thus, the process temperature can be significantly decreased. However, pure germanium and silicon germanium film with high germanium content has the problem of low chemical stability. They are etched by SC-1 (NH_3 and hydrogen peroxide solution) and hydrogen peroxide. The etch rates of SC-1 and hydrogen peroxide strongly depend on the germanium content. Therefore, the composition of germanium is optimized to 60–70% of the germanium content. The deposition temperature of the optimized film is about 520°C, which is compatible with aluminum-metallized CMOS devices (Sedkey *et al.* 2001).

Doping can be done by either *in situ* doping or *ex situ* doping by the ion implantation. *In situ* doping can be done by mixing PH_4 and B_2H_6 during LPCVD. p-Type poly-SiGe by B_2H_6 can be obtained more easily than n-type film by PH_4 because of the dopant activation ratio and the carrier mobility (Bang *et al.* 1995).

1.01.4.2 Fabrication

As mentioned above, pure germanium and silicon germanium with high germanium content dissolves in SC-1 and in hydrogen peroxide. The etch rate of pure germanium in SC-1 at 75°C is 30 µm min^{-1} and that of $Si_{0.4}Ge_{0.6}$ is 50 nm min^{-1}. The etch rate of $Si_{0.4}Ge_{0.6}$ in hydrogen peroxide at 90°C is very low (~0.01 nm min^{-1}), whereas the rate of pure germanium is still high (~5 µm min^{-1}). The plasma etching of silicon germanium film has been investigated. Basically, the same gas system as that used for silicon, such as fluorine, chlorine, and bromine gas, can be used.

The surface micromachined structures on the aluminum-metallized CMOS wafer was demonstrated with two different sacrificial layer and etching combinations. One is PSG as the sacrificial layer and HF as the etchant as is the case with polysilicon surface micromachining. The other is the polygermanium film as the sacrificial layer and hydrogen peroxide as the etchant of sacrificial layer.

1.01.4.3 Mechanical Properties

The stiffness and compliance coefficients of germanium are described using the same formulae as for silicon. They have three independent coefficients that can be accurately measured, and are given in **Table 2** (Wortman and Evans 1965). **Figure 3** shows the elastic properties of (100) and (110) germanium and silicon with respect to the in-plane directions. The Young's modulus and Poisson's ratio in the (111) plane are 137.5 GPa and 0.251, respectively.

There are little works on the measurement of mechanical properties of poly-SiGe films. The reported Young's modulus of poly-SiGe film is 132 GPa (Franke *et al.* 2003), which is less than the expected Young's modulus, which may be caused by the high dopant concentration and by the impurities because of the low deposition temperature.

The internal stress control is the critical issue in the surface micromachined devices. The low-stress poly-SiGe film can be obtained at a low temperature (Franke *et al.* 2003, Rusu *et al.* 2003, Sedkey *et al.* 2002). The stress gradient of the as-deposited poly-SiGe is positive, which means that cantilever beams have upward deflections. The gradient can be reduced by annealing at 500–550°C. The origin of the upward deflection is reported as the amorphous layer near the substrate interface has compressive stress (Sedkey *et al.* 2002).

The fracture strain was evaluated using on-chip testing structures, which deforms the cantilever

beams laterally. The fracture strain of poly-SiGe is 1.3%, which is comparable to that of the surface micromachined polysilicon structures (Franke *et al.* 2003). Sharpe (2003) reported that the tensile strength using the uniaxial tensile test is 1.0 GPa.

1.01.4.4 Thermal Properties

The thermal conductivity (Glassbrenner and Slack 1964), specific heat, and CTE of pure germanium crystal at the ambient temperature (300 K) are $60\,W\,m^{-1}\,K^{-1}$, $310\,J\,kg^{-1}\,K^{-1}$, and $5.8\times10^{-6}\,K^{-1}$, respectively. The thermal conductivity of SiGe alloy was measured as a function of the silicon content, which shows a significant decrease in the conductivity (Steele and Rosi 1958). For the $Si_{0.56}Ge_{0.44}$ alloy, the thermal conductivity is $10\,W\,m^{-1}\,K^{-1}$, which is one-sixth that of Ge and one-sixteenth that of Si, which is caused by the decrease in the phonon mean free pass due to the scattering by the solution strain associated with the alloying component. Such a small conductivity is useful in thermoelectric materials. Many applications in microsystems as well as in high-temperature thermoelectric generators, such as thermopiles (Van Gerwen *et al.* 1996), miniaturized thermoelectric generators (Strasser *et al.* 2002), and lateral Peltier elements (Wijngaards and Wolffenbuttel 2005), have been proposed. The measured Seebeck coefficients of n-type and p-type poly-SiGe ranged from −75 to $179\,\mu V\,K^{-1}$ and from 59 to $131\,\mu V\,K^{-1}$, respectively.

1.01.5 Silicon Nitride

Silicon nitride structure consists of SiN_4 tetrahedron and NSi_3 triangles, which form the amorphous structure, i.e., random networks with no symmetry and periodicity. Stoichiometric silicon nitride has the silicon to nitrogen atomic ratio of 0.75, which can be changed by various deposition conditions.

1.01.5.1 Deposition

Silicon nitride film is deposited by both the LPCVD and the PECVD tools using SiH_4 or dichlorosilane (SiH_2Cl_2) as the silicon source and NH_3 as the nitrogen source. The comparison of PECVD and LPCVD silicon nitride are summarized in Stoffel *et al.* (2003). Organic silane gas (tri(dimethylamino)silane) can be used in the PECVD system. These silicon nitride films are amorphous and the ratio of silicon to nitride can be controlled by the gas composition or the deposition condition (RF power).

1.01.5.2 Etching

Silicon nitride film can be etched using the hot phosphoric acid solution at 160°C. Since this solution attacks photoresist, the CVD oxide films are used for the hard mask layers, which makes wet nitride etching difficult. HF slowly etches ($\sim10\,nm\,min^{-1}$) silicon nitride, but the selectivity to silicon dioxide by the concentrated HF and buffered HF solutions is high enough (>100) to use the passivation layer in the silicon dioxide sacrificial etching.

In dry etching, plasma etching using CF_4, CHF_3, and SF_6 is widely employed for nitride etching. Fluorine atoms are adsorbed onto the surface as tetrafluorosilane (SiF_4) and nitrogen (N_2). Due to the high stability with respect to chemicals and the high mechanical strength, silicon nitride films are used as the mask layer for crystallographic anisotropic etching of silicon and as the structural and passivating materials of pressure sensors.

1.01.5.3 Mechanical Properties

The standard silicon nitride film Si_3N_4 deposited by LPCVD at 800°C has a Young's modulus of 300 GPa and a high tensile stress of about 1 GPa. The measured Young's modulus for CVD nitride films ranges from 83 to 380 GPa (Edwards *et al.* 2004), whose variation is caused by the deposition conditions mainly for the CVD tools but for the silicon and nitrogen ratio, which is discussed later. PECVD nitride films show larger variation; a lower modulus has been reported, which is caused by the hydrogen and carbon impurities from the source gas and the high porosity because of low deposition temperatures (Minkiewicz *et al.* 1994).

The high tensile stress does not allow the deposition of nitride films thicker than 300 nm and causes reliability problems in thin-film structures such as pressure sensors and in the passivation films in semiconductor devices. The thermal stress of silicon nitride film is compressive, since the CTE is larger than that of silicon. The intrinsic stress generated during the deposition process is compressive (Irene 1976).

It is well known that low-stress nitride films by LPCVD can be obtained by increasing the silicon compositions (Sekimoto *et al.* 1982), which is widely employed in microsystem applications. Many works show that there is a linear relationship between the internal stress and the refractive index. When the

silicon content is increased by increasing the ratio of the silicon source gas, the refractive index increases from 2.0 of stoichiometric films and the internal tensile stress of about 1.0 GPa decreases. The film whose refractive index is 2.2 is stress-free and the larger refractive index films show the compressive stress. The change in the refractive index originates from the silicon refractive index, which is higher than that of the stoichiometric silicon nitride film.

Nitride films deposited by the PECVD process show a similar relationship between the silicon–nitrogen ratio, the refractive index, and the internal stress (Taylor 1991). In addition, PECVD films have more variables such as residues from the source gas, RF powers, and a lower deposition temperature. The measured internal stress of the nitride films shows larger variations.

Strength is high and excellent fatigue properties are measured. The strength measured by tensile tests (Koskinen and Johnson 1989, Sharpe 2003, Yoshioka *et al.* 2000), bulge tests (Yang and Paul 2002), and on-chip structures (Fan *et al.* 1990) ranged from 4 to 8 GPa. A smaller degradation in strength by humidity was observed (Tsuchiya and Sakata 2001). The fatigue test using resonant structures has been reported, which demonstrates a small decrease in strength during 109 loading cycles (Chuang *et al.* 2005).

1.01.5.4 Thermal Properties

The thermal properties of LPCVD low-stress nitride films were measured using the two-layer composite microbridge resistors. The thermal conductivity is $3.2\,W\,m^{-1}\,K^{-1}$, the thermal diffusivity is $1.3 \times 10^{-6}\,m^2\,s^{-1}$, and the heat capacity is $0.7\,J\,kg^{-1}K^{-1}$ (Mastangero *et al.* 1990). The thermal conductivity of PECVD nitride films for CMOS passivation is $2.23\,W\,m^{-1}\,K^{-1}$ (von Arx *et al.* 2000) at the ambient temperature.

The reported CTE values of LPCVD silicon nitride films ranged from $2.5–3.85 \times 10^{-6}\,K^{-1}$ (Tokuyama *et al.* 1967). The values of $2.8 \times 10^{-6}\,K^{-1}$ (Riethmuller and Benecke 1988) and $3.0 \times 10^{-6}\,K^{-1}$ (Temple-Boyer *et al.* 1998) are often used for calculation. The CTE of PECVD nitride films shows a larger variation. Martyniuk *et al.* (2004) reported that the CTE values increase from 2.6×10^{-6} to $5.2 \times 10^{-6}\,K^{-1}$ with decreasing deposition temperature from 300°C to 50°C. Minkiewicz *et al.* (1994) showed the CTE dependence on compositions. The value increases from 1.8×10^{-6} to $2.6 \times 10^{-6}\,K^{-1}$ by increasing Si content from the stoichiometric composition.

1.01.6 Silicon Dioxide

Silicon dioxide has a glassy amorphous structure, which shows no symmetry and periodicity in the structure. The structures of glasses are described as the random network model. In silicon dioxide, the structures have 3D networks that are composed of a silica tetrahedra. The silica tetrahedra share corners, not edges or faces, and each oxygen ion on the corner is linked to not more than two cations at the center of the tetrahedra.

Silicon dioxide films are widely used in semiconductor processes, such as for device insulation and gate oxide films. Impurities in silicon dioxide occupy two different positions in the network structures. Impurities such as phosphor and boron substitutes silicon in the network known as network formers, in which boron forms triangles and phosphor forms tetrahedra but with only three of the four corners linked to other polyhedras. In the semiconductor process, phosphorous-doped and boron- and phosphor-doped silicon dioxide films are used as dielectric layers, which are called PSG and BPSG, respectively. Other impurities such as sodium, potassium, and lead occupy random positions distributed throughout the structure known as network modifiers. A sodium-containing glass is important in microsystems applications, which is used as the substrate in anodic bonding. Pyrex (sodium borosilicate glass) glass has a CTE similar to that of SCS.

Oxide film is used as a sacrificial layer in SCS/polycrystalline silicon structure, as a mask in the crystallographic anisotropic etching of silicon, and as a support membrane of gas sensors and infrared sensors because of its low thermal conductivity.

1.01.6.1 Deposition

The deposition methods of silicon dioxide film are classified into thermal oxidization and chemical and physical vapor depositions. The oxide film is grown by the thermal oxidization of silicon at 900°C or above, which is known as thermal oxide film. The thermal oxidization process involves oxygen diffusion to oxidized film and the oxidization process at the oxide/silicon interface. The thickness of the thick oxide films is proportional to square root of the oxidization time. The practical oxide thickness is limited to 2 or 3 μm.

The oxide film can be deposited by CVD using SiH_4 and nitrous oxide (N_2O) at 900°C or using tetraethylorthosilicate (TEOS, $Si(C_2H_5O)_4$) at 720°C. *In situ* doping can be done by mixing PH_4, B_2H_6,

trimethylphosphite (TMP, $(CH_3O)_3P$), and trimethylborate (TMB, $(CH_3O)_3B$). Because of the need for a low thermal budget after aluminum deposition, PECVD at a temperature below 400°C is employed widely. These doped oxide films reflow at low temperature, which is useful in planarization process for multilayer polysilicon deposition. Sputter deposition is a technique for making oxide films. It is useful because of low deposition temperature and easy composition control, especially for the bonding layer in the anodic bonding process. However, since control of the film quality is difficult and it often suffers from low breakdown strength, the use of sputtered films is limited.

1.01.6.2 Etching

Silicon dioxide film is etched by HF solution. The reaction of the etching is as given below:

$$SiO_2 + 6HF \rightarrow H_2SiF_6 + 2H_2O \qquad [I]$$

The etch rate for thermal oxide by concentrated HF solution (49%) is about 1 μm min^{-1} at the ambient temperature. However, the concentrated HF solution degrades easily with use, whereas buffered HF solution has a small decrease in the etch rate. The etch rate by 1:6 buffered hydrofluoric acid (BHF) solutions is about 0.1 μm min^{-1} at the ambient temperature. The etch rate increases with the temperature and the higher activation energy than for the concentrated HF solutions. Phosphorous-doped oxide film is etched more quickly. The etch rate of PSG film is a few micrometers per minute, about 10 times that of the thermal oxide, which is useful as a sacrificial layer for polysilicon structures (Poenar *et al.* 1994). BPSG films that can be reflown at lower temperatures have a lower etch rate when compared to PSG films with higher boron contents (French and Wolffenbuttel 1993). HF solutions do not attack silicon but polysilicon grain boundaries are attacked slowly, and the roughness of polysilicon surfaces increases by HF etching. Ammonium fluoride etches SCS very slowly; the etch pit formation based on alkali solution anisotropic etching also increases the surface roughness. Both of them are used in the stiction reduction processes. Thin (<0.2 μm) doped polysilicon films permeate the HF solution and the underlying sacrificial oxide layers can be etched. The grain boundaries are etched during the subsequent cleaning process and the polysilicon layer becomes a porous structure that has permeability.

Gas-phase etching of oxide film using anhydrous HF with water or methanol has been proposed. The etch rate is about 0.1 μm min^{-1}. Gas-phase etching is convenient for SCS structures using SOI wafer. However, it is difficult to employ for polysilicon structures where PSG sacrificial layer is used because of the residues originating from the dopant and the low selectivity to the silicon nitride film. More simply, HF vapor etching can be done by placing chips just above the concentrated HF solution, in which vaporized HF and water etches the oxide layers. This reaction is agitated by lamp heating.

Plasma etching using CF_4 and/or CHF_3 gas is employed for oxide films. CF_x radicals etch oxide films to form CO and CO_2 from oxygen and SF_4 from silicon. The source gases are also the source of the fluorocarbon polymer that forms side wall passivation, which results in an anisotropic profile.

1.01.6.3 Mechanical Properties

The Young's modulus of bulk silica glass and Pyrex grass is 72 and 68 GPa, respectively. The thermal oxide film shows similar values around 70 GPa (Petersen 1978). However, the Young's modulus of LPCVD and PECVD oxide films has wide variations, which ranges from 30 to 70 GPa. The carbon and hydrogen impurities and pore in the films are the origin of this variation.

The internal stress of thermal and CVD oxide, which is about −300 MPa, is compressive, which is mainly caused by the thermal stress generated by mismatch of the thermal expansion. The internal stress of oxide film by PECVD shows a wide range of variations, which ranges from compressive to tensile stress. Remaining organic materials and water and pore seem to be the source of variation. After annealing at 900°C or above these films have a compressive stress similar to thermal oxide films. The internal stress of the porous or doped oxide films changes with time because of the absorption of water (Chen *et al.* 2003).

The strength of oxide films is measured by both bending and tensile test. The measured strength ranges from 0.5 to 2 GPa. The strength in the air is a half of that in a vacuum, which shows that the strength decreased significantly with the presence of water vapor in the test environment. The static fatigue can be enhanced by the water in the air (Tsuchiya *et al.* 2000).

1.01.6.4 Thermal Properties

The thermal conductivity and the temperature dependency of CMOS oxide films were measured using microstructures. The thermal conductivity of various oxide films, such as thermal, LPCVD, and PECVD

(TEOS) oxide, and BPSG and PSG is almost the same, i.e., about 1.0 $W\,m^{-1}\,K^{-1}$ (von Arx *et al.* 2000).

The CTE of thermally grown oxide is $5 \times 10^{-7}\,K^{-1}$ (Blech and Cohen 1982), which is similar to that of bulk fused silica glass.

1.01.7 Summary

Comprehensive descriptions of silicon and related materials, silicon germanium, silicon oxide, and nitride, as the fundamental material in microsystems are provided. SCS has excellent electrical, mechanical, and thermal properties, which is used not only for the substrate of semiconductor and microsystem devices but also for the sensing structures in inertial sensors. Polycrystalline silicon and silicon germanium films have properties similar to those of SCS and the integration of controlling circuit to microsystem devices adds special features. The properties of silicon and silicon germanium are summarized in **Table 4**. Silicon nitride and silicon dioxide as

Table 4 Properties of single-crystal silicon, polycrystalline silicon, polycrystalline silicon germanium, and single-crystal germanium

	Si	*Poly-Si*	*Poly-$Si_{1-x}Ge_x$*	*Ge*
Electrical properties				
Energy gaps (eV)	1.12[r]		0.94 ($x = 0.44$)[a,b]	0.66[r]
Dielectric constant	11.7[r]			16.2[r]
Electron mobility ($m^2\,V^{-1}\,s^{-1}$)	0.150[r]		0.02 ($x = 0.6$)[a,c]	0.39[r]
Hole mobility ($m^2\,V^{-1}\,s^{-1}$)	0.045[r]		0.016 ($x = 0.6$)[a,c]	0.19[r]
Intrinsic resistivity (Ω cm)	2.3×105[r]			47[r]
Temperature coefficient of resistance (TCR) (K^{-1})	$<1.9 \times 10^{-3}$[d]		$2–2.4 \times 10^{-2}$ ($x = 0.3$)[e]	
Piezoresistive coefficient, π_l ($\times 10^{-12}$ Pa^{-1})	−1020 [100] (n), +930 [111] (p)[f]	15–27[g] (gauge factor)	480–760 [110] ($x = 0.1$)[a,h]	−970 [111] (n), +650 [111] (p)[f]
Optical properties				
Reflectivity ($\lambda = 590$ nm)	0.357[i]			0.523[i]
Refractive index ($\lambda = 590$ nm)	3.969[i]			5.748[i]
Mechanical properties				
Density ($10^3\,kg\,m^{-3}$)	2.33[r]			5.32[r]
Young's modulus (GPa)	168.9 <110>[j]	158[j]	132 ($x = 0.65$)[i]	137.7 <110>[j]
Tensile strength (GPa)	~7[i]	1.2–3[k]	1.0[k]	
Thermal properties				
CTE ($10^{-6}\,K^{-1}$)	2.6[r]		4.1 ($x = 0.6$)[a,l]	5.8[r]
Thermal conductivity ($W\,m^{-1}\,K^{-1}$)	156[r]	29–34[m]	10 ($x = 0.56$)[n]	60[r]
Specific heat ($J\,kg^{-1}\,K^{-1}$)	700[r]		390 ($x = 0.6$)[a,l]	310[r]
Seebeck coefficient ($\mu V\,K^{-1}$)	480 (p), −420 (n)[o]	190 (p), −120 (n)[p]	59–131 (p), −179 to −75 (n)[p]	500–1000 (p), −1000 (n)[q]
Melting point (°C)	1412[r]			937[r]

[a] Data of bulk crystal.
[b] Braunstein *et al.* (1958).
[c] Landoldt-Bornstein (1987).
[d] Hull (1999).
[e] Sedkey *et al.* (1998).
[f] Smith (1954).
[g] Seto (1976).
[h] Richter *et al.* (2005).
[i] Franke *et al.* (2003).
[j] Brantley (1973).
[k] Sharpe (2003).
[l] Schäffler (1997).
[m] Tai *et al.* (1988).
[n] Steele and Rosi (1958).
[o] Geballe and Hull (1954).
[p] Wijngaards and Wolffenbuttel (2005).
[q] Geballe and Hull (1954).
[r] Sze (1981).

Table 5 Properties of LPCVD silicon nitride and thermally grown silicon dioxide films

	SiO_2 (thermal)	SiN (LPCVD)
Melting point (°C)	~1600[g]	1880[a]
Density ($\times 10^3$ kg m^{-3})	2.2[g]	3.1[g]
Energy gap (eV)	9[g]	~5.0[g]
Refractive index	1.46[g]	2.05[g]
Dielectric constant	3.9[g]	7.5[g]
Dielectric strength (V cm^{-1})	107[g]	107[g]
DC resistivity (Ω cm)	1014–1016[g]	~1014[g]
CTE ($\times 10^{-6}$ K^{-1})	0.5[g]	2.8–3.0[b]
Thermal conductivity (W m^{-1} K^{-1})	1.4[g]	3.2[c]
Young's modulus (GPa)	70[d]	340[e]
Tensile strength (GPa)	0.5–2[f]	5–7[f]
Internal stress (MPa)	~−300[d]	~1000[g]

[a] Riley (2000).
[b] Riethmuller and Benecke (1988), Temple-Boyer *et al.* (1998).
[c] Mastangero *et al.* (1990).
[d] Pertersen (1978).
[e] Edwards *et al.* (2004).
[f] Tsuchiya and Sakata (2001).
[g] Sze 1981.

dielectric materials in semiconductor devices have wide applications in microsystems. The properties are summarized in **Table 5**.

The importance of silicon and related materials does not change although many materials are being introduced in microsystems. The description of the properties of these materials should be investigated further, especially for the mechanical properties. The mechanical reliability of silicon and related materials, which has not been well understood, needs to be investigated.

References

Adamczewska J, Budzyński T 1984 Stress in chemically vapor-deposited silicon films. *Thin Solid Films* **113**, 271–83

von Arx M, Paul O, Baltes H 2000 Process-dependent thin-film thermal conductivities for thermal CMOS MEMS. *J. MEMS* **9**, 136–45

Bang D S, Cao M, Wang A, Saraswat K C, King T-J 1995 Resistivity of boron and phosphorus doped polycrystalline $Si_{1-x}Ge_x$ films. *Appl. Phys. Lett.* **66**, 195–7

Berry B S, Pritchet W C 1991 Stress and thermal expansion of boron-doped silicon membranes on silicon substrates. *J. Vac. Sci. Technol. A* **9**, 2231–4

Blech I, Cohen U 1982 Effects of humidity on stress in thin silicon dioxide films. *J. Appl. Phys.* **53**, 4202–7

Brantley W A 1973 Calculated elastic constants for stress problems associated with semiconductor devices. *J. Appl. Phys.* **44**, 534–5

Braunstein R, Moore A R, Herman F 1958 Intrinsic optical absorption in germanium-silicon alloys. *Phys. Rev.* **109**, 695–710

Brown S, Arsdell W V, Muhlstein C L 1997 Materials reliability in MEMS devices. *Proc. Int. Conf. Solid-State Sensors and Actuators,* Chicago, IL, USA, pp. 591–3

Bruel M 1995 Silicon on insulator material technology. *Electron. Lett.* **51**, 1201–2

Celler G K, Cristoloveanu S 2003 Frontiers of silicon-on-insulator. *J. Appl. Phys.* **93**, 4955–78

Chen C P, Leipold M H 1985 Stress rate and proof-testing of silicon wafers. *J. Am. Ceram. Soc.* **68**, C54–5

Chen K-S, Zang X, Lin S-Y 2003 Intrinsic stress generation and relaxation of plasma-enhanced chemical vapor deposited oxide during deposition and subsequent thermal cycling. *Thin Solid Films* **434**, 190–202

Chuang W-H, Fettig R K, Ghodssi R 2005 An electrostatic actuator for fatigue testing of low-stress LPCVD silicon nitride thin films. *Sens. Actuators A* **121**, 557–65

Chung C-K, Tsai M-Q, Tsai P-H, Lee C 2005 Fabrication and characterization on amorphous Si films by PECVD for MEMS. *J. Micromech. Microeng.* **15**, 136–42

Connally J A, Brown S B 1992 Slow crack growth in single-crystal silicon. *Science* **256**, 1537–9

Core T A, Tsang W K, Sherman S J 1993 Fabrication technology for an integrated surface-micromachined sensor. *Solid State Technol.* **36**, 39–47

Edwards R L, Coles G, Sharpe Jr. W N 2004 Comparison of tensile and bulge tests for thin-film silicon nitride. *Exp. Mech.* **44**, 49–55

Ericson F, Schweitz J-Å 1990 Micromechanical fracture strength of silicon. *J. Appl. Phys.* **68**, 5840–4

Esashi M, Komatsu H, Matsuo T, Takahashi M, Takishima T, Imabayashi K, Ozawa H 1982 Fabrication of catheter-tip and sidewall miniature pressure sensors. *IEEE Trans. Electron Device* **ED-29**, 57–63

Fan L S, Howe R T, Muller R S 1990 Fracture toughness characterization of brittle thin films. *Sens. Actuators A* **21–3**, 872–4

Franke A E, Heck J M, King T-J, Howe R T 2003 Polycrystalline silicon-germanium films for integrated microsystems. *J. MEMS* **12**, 160–71

French P J, Wolffenbuttel R F 1993 Reflow of BPSG from sensor applications. *J. Micromech. Microeng.* **3**, 135–7

Frenkel J 1926 Zur Theorie der Elastizitätsgrenze und der Festigkeit kristallinischer Körper. *Z. Physik* **27**, 572–609

Fujitsuka N, Sakata J 2002 A new processing technique to prevent stiction using silicon selective etching for SOI-MEMS. *Sens. Actuators A* **97–8**, 716–9

Geballe T H, Hull G W 1954 Seebeck effect in germanium. *Phys. Rev.* **94**, 1134–40

Geballe T H, Hull G W 1955 Seebeck effect in silicon. *Phys. Rev.* **98**, 940–7

Glassbrenner C J, Slack G A 1964 Thermal conductivity of silicon and germanium from 3°K to the melting point. *Phys. Rev.* **134**, A1058–69

Goto H, Ikeda M, Sakata M, Imanaka K 1996 Development of miniature two-dimensional optical scanner (2). *Trans. Jpn. Soc. Mech. Engrs* **62**, 3202–9 (in Japanese)

Grüning U, Lehmann V, Engelhardt C M 1995 Two-dimensional infrared photonic band gap structure based on porous silicon. *Appl. Phys. Lett.* **66**, 3254–6

Guckel H, Sniegowski J J, Christenson T R 1989 Advances in processing techniques for silicon micromechanical devices with smooth surfaces. *Proc. IEEE Int. Workshop Micro Electromechanical Systems,* Salt Lake City, UT, USA, pp. 71–5

Hall J J 1967 Electronic effects in the elastic constants of n-type silicon. *Phys. Rev.* **161**, 756–61

Hu J Z, Merkle L D, Menoni C S, Spain I L 1986 Crystal data for high-pressure phases of silicon. *Phys. Rev. B* **34**, 4679–84

Huber W, Borionetti G, Villani C 1989 The behavior of polysilicon thin film stress and structure under rapid thermal processing conditions. *Mater. Res. Soc. Symp. Proc.* **130**, 389–93

Hull R (*ed*.) 1999 *Properties of Crystalline Silicon – EMIS Datareviews Series No. 20.* INSPEC, London

Iida T, Itoh T, Noguchi D, Takano Y 2000 Residual lattice strain in thin silicon-on-insulator bonded wafers: Thermal behavior and formation mechanisms. *J. Appl. Phys.* **87**, 675–81

Irene E A 1976 Residual-stress in silicon-nitride films. *J. Electron. Mater.* **5**, 287–98

Isono Y, Namazu T, Terayama N 2006 Development of AFM tensile test technique for evaluating mechanical properties of sub-micron thick DLC films. *J. MEMS* **15**, 169–80

Izumi K, Dolken M, Ariyoshi H 1978 CMOS devices fabricated o buried SiO_2 layers formed by oxygen implantation into silicon. *Electron. Lett.* **14**, 593–4

Izumi K, Doken M, Ariyoshi H 1980 High speed C-MOS IC using buried SiO_2 layers formed by ion implantation. *Jpn. J. Appl. Phys.* **19(Suppl. 19-1)**, 151–4

Izumi S, Ping C W, Yamaguchi M, Sakai S, Ueda Y, Suzuki A 2005 Development of specimen and test method for strength analysis of MEMS micromirror. *Eng. Fract. Mech.* **72**, 2672–85

Kamins T I 1990 Design properties of polycrystalline silicon. *Sens. Actuators A* **21–3**, 817–24

Kanda Y 1982 A graphical representation of the piezoresistance coefficients in silicon. *IEEE Trans. Electron Devices* **ED-29**, 64–70

King T-J, Saraswat K C 1994 Polycrystalline silicon-germanium thin-film transistors. *IEEE Trans. Electron Dev.* **41**, 1581–91

King T-J, Pfiester J R, Shott J D, McVittie J P, Saraswat K C 1990 A polycrystalline-Si_{1-x}-Ge_x-gate CMOS technology. *Tech. Digest Int. Electron Device Meeting*, pp. 253–6

Koskinen J, Johnson H H 1989 Silicon nitride fibers using microfabrication methods. *Mater. Res. Soc. Meet. Proc.* Boston, MA, USA **130**, 63–8

Krulevitch P, Howe R T, Johnson G C, Huang J 1991 Stress in undoped LPCVD polycrystalline silicon. *Proc. IEEE Int. Conf. Solid-State Sensors and Actuators,* San Francisco, CA, USA, pp. 949–52

Landoldt-Bornstein 1987 *Numerical Data and Functional Relationships in Science and Technology.Springer, Berlin, New Series Group III, Vol. 17a*

Lasky J B 1986 Wafer bonding for silicon-on-insulator technologies. *Appl. Phys. Lett.* **48**, 78–80

Lee C S, Lee J H, Choi C A, No K, Wee D M 1999 Effect of phosphorus on stress of multi-stacked polysilicon film and single crystalline silicon. *J. Micromech. Microeng.* **9**, 252–63

Lehmann V 1993 The physics of macropore formation in low doped n-type silicon. *J. Electrochem. Soc.* **140**, 2836–43

Maier-Schneider D, Maibach J, Obermeier E, Schneider D 1995 Variation in Young's modulus and intrinsic stress of LPCVD-polysilicon due to high-temperature annealing. *J. Micromech. Microeng.* **5**, 121–4

Maier-Schneider D, Köprülülü A, Ballhausen Holm S, Obermeier E 1996 Elastic properties and microstructure of LPCVD polysilicon flims. *J. Micromech. Microeng.* **6**, 436–46

Martyniuk M, Antoszewski J, Musca C A, Dell J M, Faraone L 2004 Stress response of low temperature PECVD silicon nitride thin films to cryogenic thermal cycling. *Proc. Conf. Optoelectronic and Microelectronic Materials and Devices,* Brisbane, Australia, pp. 381–4

Mastangero C H, Tai Y-C, Muller R S 1990 Thermophysical properties of low-residual stress, silicon-rich, LPCVD silicon nitride films. *Sens. Actuators A* **23**, 856–60

Minkiewicz V J, Stasiak J, White R L, Cornell R W, Moore J O, Eaton R, Eldridge J M, Lee F C 1994 Some mechanical and thermal properties of PECVD a-Si_xC_{1-x}:H and a-Si_xN_{1-x}:H films prepared by mixed frequency plasma processes. *Surf. Coat. Technol.* **68–69**, 229–33

Miura H, Okamoto N 1994 Crystallization-induced stress in phosphorus-doped amorphous-silicon thin-films. *J. Appl. Phys.* **75**, 4747–9

Miura H, Ohta H, Okamoto N, Kaga T 1992 Crystallization-induced stress in silicon thin films. *Appl. Phys. Lett.* **66**, 2746–8

Muhlstein C L, Brown S B, Ritchie R O 2001a High-cycle fatigue of single-crystal silicon thin films. *JMEMS* **10**, 593–600

Muhlstein C L, Brown S B, Ritchie R O 2001b High-cycle fatigue and durability of polycrystalline silicon thin films in ambient air. *Sens. Actuators A* **94**, 177–88

Muhlstein C L, Stach E A, Ritchie R O 2002 A reaction-layer mechanism for the delayed failure of micron-scale polycrystalline silicon structural films subjected to high-cycle fatigue loading. *Acta Mater.* **50**, 3579–95

Mutoh M, Iyoda M, Fujita K, Mizuno C, Kondo M, Imai M 1990 Development of integrated semiconductor-type acceleration. *Proc. IEEE Workshop on Electronic Applications in Transportation,* Dearborn, MI, USA, pp. 35–8

Okada Y, Tokumaru Y 1984 Precise determination of lattice parameter and thermal expansion coefficient of silicon between 300 and 1500 K. *J. Appl. Phys.* **56**, 314–20

Orowan E 1948 Fracture and strength of solids. *Rep. Prog. Phys.* **12**, 185–232

Orpana M, Korhonen A O 1991 Control of residual stress of polysilicon thin films by heavy doping in surface micromachining. *Proc. IEEE Int. Conf. Solid-State Sensors and Actuators,* San Francisco, CA, USA, pp. 957–60

Park S, Kwak D, Ko H, Song T, Cho D 2005 Selective silicon-on-insulator (SOI) implant: A new micromachining method without footing and residual stress. *J. Micromech. Microeng.* **15**, 1607–13

Pertersen K E 1978 Dynamic micromechanics on silicon: Techniques and devices. *IEEE Trans. Electron Devices* **ED-25**, 1241–50

Petersen K E 1982 Silicon as a mechanical materials. *Proc. IEEE* **70**, 420–57

Pierron O N, Muhlstein C L 2005 The extended range of reaction-layer fatigue susceptibility of polycrystalline silicon thin films. *Int. J. Fract.* **135**, 1–18

Plummer J, Deal M, Griffin P B 2000 *Silicon VLSI Technology: Fundamentals, Practice and Modeling*. Prentice-Hall, Upper Saddle River, NJ

Poenar D, French P J, Mallée R, Sarro P M, Wolffenbuttel R F 1994 PSG Layers for surface micromachining. *Sens. Actuators A* **41-2**, 304–9

Richter J, Hansen O, Nylandsted Larsen A, Lundsgaard Hansen J, Eriksen G F, Thomsen E V 2005 Piezoresistance of silicon and strained $Si_{0.9}Ge_{0.1}$. *Sens. Actuators A* **123–4**, 388–96

Riethmuller W, Benecke W 1988 Thermally excited silicon microactuators. *IEEE Trans. Electron Devices* **35**, 758–63

Riley F L 2000 Silicon nitride and related materials. *J. Am. Ceram. Soc.* **83**, 245–65

Rusu C, Sedky S, Parmentier B, Verbist A, Richard O, Brijs B, Geenen L, Witvrouw A, Lärmer F, Fischer F, Kronmüller S, Leca V, Otter V 2003 New low-stress PECVD poly-SiGe layers for MEMS. *JMEMS* **12**, 816–25

Sato K, Yoshioka T, Ando T, Shikida M, Kawabata T 1998 Tensile testing of silicon film having different crystallographic orientations carried out on a silicon chip. *Sens. Actuators A* **70**, 148–52

Schäffler F 1997 High-mobility Si and Ge structures. *Semicond. Sci. Technol.* **12**, 1515–49

Schwarz B, Robbins H 1976 Chemical etching of silicon – IV. Etching technology. *J. Electrochem. Soc.* **123**, 1903–9

Sedkey S, Fiorini P, Caymax M, Verbist A, Baert C 1998 IR bolometers mad of polycrystalline silicon germanium. *Sens. Actuators A* **66**, 193–9

Sedky S, Witvrouw A, Bender H, Baert K 2001 Experimental determination of the maximum post-process annealing temperature for standard CMOS wafers. *IEEE Trans. Electron Devices* **48**, 377–85

Sedkey S, Witvrouw A, Baert K 2002 Poly SiGe, a promising material for MEMS monolithic integration with the driving electronics. *Sens. Actuators A* **97–8**, 503–11

Seidel H, Csepregi L, Heuberger A, Baumgärtel H 1990 Anisotropic etching of crystalline silicon in alkaline solutions II. Influence of dopants. *J. Electrochem. Soc.* **137**, 3626–32

Sekimoto M, Yoshihara H, Ohkubo T 1982 Silicon nitride single-layer X-ray mask. *J. Vac. Sci. Technol.* **21**, 1017–21

Seto J Y W 1976 Piezoresistive properties of polycrystalline silicon. *J. Appl. Phys.* **47**, 4780–3

Sharpe Jr. W N 2003 Tensile testing at the micrometer scale: Opportunities in experimental mechanics. *Exp. Mech.* **43**, 228–37

Sharpe Jr. W N, Turner K T, Edwards R L 1999 Tensile testing of polysilicon. *Exp. Mech.* **39**, 162–70

Slack G A 1964 Thermal conductivity of pure and impure silicon, silicon carbide, and diamond. *J. Appl. Phys.* **35**, 3460–6

Smith C S 1954 Piezoresistance effect in germanium and silicon. *Phys. Rev.* **94**, 42–9

Smith R L, Collins S D 1992 Porous silicon formation mechanisms. *J. Appl. Phys.* **71**, R1–22

Steele M C, Rosi F D 1958 Thermal conductivity and thermoelectric power of germanium-silicon alloys. *J. Appl. Phys.* **29**, 1517–20

Stoffel A, Kovács A, Kronast W, Müller B 2003 LPCVD against PECVD for micromechanical applications. *J. Micromech. Microeng.* **6**, 1–13

Strasser M, Aigner R, Franosch M, Wachutka G 2002 Miniaturized thermoelectric generators based on poly-Si and poly-SiGe surface micromachining. *Sens. Actuators A* **97–8**, 535–42

Sze S M 1981 *Physics of Semiconductor Devices*. 2nd edn. John Wiley and Sons, New York

Tai Y-C, Mastangero C H, Muller R S 1988 Thermal conductivity of heavily doped low-pressure chemical vapor deposited polycrystalline silicon films. *J. Appl. Phys.* **63**, 1442–7

Tan C M, Gan Z, Gao X 2003 Temperature and stress distribution in the SOI structure during fabrication. *IEEE Trans. Semicond. Manufact.* **16**, 314–8

Taylor J A 1991 The mechanical properties and microstructure of plasma enhanced chemical vapor deposited silicon nitride thin films. *J. Vac. Sci. Technol. A* **9**, 2464–8

Temple-Boyer P, Rossi C, Saint-Etienne E, Scheid E 1998 Residual stress in low pressure chemical vapor deposition SiN_x films deposited from silane and ammonia. *J. Vac. Sci. Technol. A* **16**, 2003–7

Tokuyama T, Fujii Y, Sugita Y, Kishino S 1967 Thermal expansion coefficient of a pyrolytically deposited silicon nitride film. *Jpn. J. Appl. Phys.* **6**, 1252–3

Tsuchiya T, Sakata J 2001 Tensile testing of thin films using electrostatic force grip. *ASTM STP 1413: Mechanical Properties of Structural Films*. American Society of Testing and Materials, Orlando, FL, pp. 214–28

Tsuchiya T, Inoue A, Sakata J, Hashimoto M, Yokoyama A, Sugimoto M 1998a Fatigue test of single crystal silicon resonator. *Tech. Digest 16th Sensor Symposium*, Kawasaki, Japan, pp. 277–80

Tsuchiya T, Tabata O, Sakata J, Taga Y 1998b Specimen size effect on tensile strength of surface micromachined polycrystalline silicon thin films. *JMEMS* **7**, 106–13

Tsuchiya T, Inoue A, Sakata J 2000 Tensile testing of insulating thin films; humidity effect on tensile strength of SiO_2 films. *Sens. Actuators A* **82**, 286–90

Tsuchiya T, Hirata M, Chiba N, Udo R, Yoshitomi Y, Ando T, Sato K, Takashima K, Higo Y, Saotome Y, Ogawa H, Ozaki K 2005 Cross comparison of thin film tensile-testing methods examined with single-crystal silicon, polysilicon, nickel, and titanium films. *JMEMS* **14**, 1178–86

Van Gerwen P, Slater T, Chévrier J B, Baert K, Mertens R 1996 Thin-film boron-doped polycrystalline silicon 70%–germanium 30% for thermopiles. *Sens. Actuators A* **53**, 325–9

Völklein F, Baltes H 1992 A microstructure for measurement of thermal conductivity of polysilicon thin films. *JMEMS* **1**, 193–6

Wijngaards D D L, Wolffenbuttel R F 2005 Thermo-electric characterization of APCVD poly$Si_{0.7}Ge_{0.3}$ for IC-compatible fabrication of integrated lateral Peltier elements. *IEEE Trans. Electron Devices* **ED-52**, 1014–25

Williams J S, Chen Y, Wong-Leung J, Kerr A, Swain M V 1999 Ultra-micro-indentation of silicon and compound semiconductors with spherical indenters. *J. Mater. Res.* **14**, 2338–43

Wilson C J, Beck P A 1996 Fracture testing of bulk silicon microcantilever beams subjected to a side load. *JMEMS* **5**, 142–50

Wilson C J, Ormeggi A, Narbutovskih M 1996 Fracture testing of silicon microcantilever beams. *J. Appl. Phys.* **79**, 2386–93

Wortman J J, Evans R A 1965 Young's modulus, shear modulus, and Poisson's ratio in silicon and germanium. *J. Appl. Phys.* **36**, 153–6

Yang J, Paul O 2002 Fracture properties of LPCVD silicon nitride thin films from the load–deflection of long membranes. *Sens. Actuators A* **97–8**, 520–6

Yang J, Kahn H, He A-Q, Phillips S M, Heuer A H 2000 A new technique for producing large-area as-deposited zero-stress LPCVD polysilicon films: The multipoly process. *JMEMS* **9**, 485–93

Yeo Y-C, Liu Q, King T-J, Hu C, Kawashima T, Oishi M, Mashiro S, Sakai K 2000 Enhanced performance in sub-100 nm CMOSFETs using strained epitaxial silicon-germanium. *Tech. Digest IEEE Int. Electron Device Meeting,* San Francisco, USA, pp. 753–6

Yonehara T, Sakaguchi K 2001 ELTRAN®; novel SOI wafer technology. *JSAP Int.* **4**, 11–16

Yonehara T, Sakaguchi K, Sato N 1994 Epitaxial layer transfer by bond and etch back of porous Si. *Appl. Phys. Lett.* **64**, 2108–10

Yoshioka T, Ando T, Shikida M, Sato K 2000 Tensile testing of SiO_2 and Si_3N_4 films carried out on a silicon chip. *Sens. Actuators A* **82**, 291–6

Zhang X G, Collins S D, Smith R L 1989 Porous silicon formation and electropolishing of silicon by anodic polarization in HF solution. *J. Electrochem. Soc.* **136**, 1561–5

Biography

Toshiyuki Tsuchiya received his MS degree from the University of Tokyo and his Ph.D. degree from Nagoya University in 1993 and 2004, respectively. He worked with Toyota Central Research and Development Laboratories from 1993 to 2004. In 2004, he became an associate professor of the Department of Mechanical Engineering, Kyoto University, and since 2005 he has been an associate professor of the Department of Micro Engineering, Kyoto University. He is currently engaged in the research of silicon surface micromachining, its sensor application, and thin-film mechanical property evaluation. He is a member of IEEE, Material Research Society, the Japan Society of Mechanical Engineers, and the Institute of Electrical Engineers of Japan.

1.02 Compound Semiconductors

Nethaji Dharmarasu and Hartmut Hillmer, Institute of Nanostructure Technologies and Analytics (INA) and Center for Interdisciplinary Nanostructure Science and Technology (CINSaT), University of Kassel, Germany

Glossary

DBR Distributed Bragg Reflector
DWDM Dense Wavelength Division Multiplexing
LED Light-emitting Diode
MEMS Microelectromechanical system
MOEMS Microoptoelectromechanical system
VCSEL Vertical Cavity Surface-Emitting Laser

1.02.1 Compound Semiconductors for Microsystems

Compound semiconductor microsystems have become an emerging technology because of their unique intrinsic properties, which offer a number of materials-based advantages for optoelectronic and sensing applications. The intrinsic properties of these materials such as direct or indirect band gap, variation of band gap (from zero via narrow to wide), piezoelectricity, piezoresistivity, variable thermal conductivity, higher saturation velocity of electrons, operation at high temperature and high frequencies make them superior to the well-developed silicon technology. The zincblende structure of most common compound semiconductors allows piezoelectricity, which leads to interesting sensing applications. The compound semiconductor materials allow monolithic integration of optoelectronic devices such as lasers, light-emitting diode (LEDs), and photodiodes with micromechanical structures. This enables

light generation, transmission, modulation, and detection on a single chip. Furthermore, epitaxially grown single-crystalline compound semiconductor layers give rise to well-controlled mechanical characteristics. The epitaxial advantage also facilitates the growth of monolithic multilayers with ternary and quaternary alloys, enabling precise lattice matching. This introduces a wide variety of heterostructure-based physical effects for microsystems. The monolithic multilayers can be selectively etched to release the mechanical structures, as the compound semiconductors are very rich in chemistry.

Micromachining on compound semiconductors has been successfully applied to fabricate cantilevers, membranes, waveguide switches, radio frequency (RF) inductors, microbolometers, micromotors, wavelength-tunable microcavity devices such as optical filters, optical amplifiers, vertical cavity surface-emitting lasers (VCSELs), LEDs, and photodiodes.

Diverse nature of compound semiconductors provides numerous possibilities to design various material compositions. However, compounds from group III–V (gallium arsenide (GaAs)-, indium phosphide (InP)-, and nitride-based materials) and compounds from group IV (SiGe and SiC materials) have been used for microsystems due to their material properties, which is required for the present technological advances. The structural and mechanical properties of these materials are listed in **Table 1**. Dielectric material-based micromechanical structures with group II–VI compound semiconductors have also been investigated for (hybrid) microsystems by Keating *et al.* (2006).

GaAs-based materials are most extensively studied for optoelectromechanical devices in compound semiconductors as is silicon in semiconductors. The GaAs-based materials are promising for the fabrication of integrated sensors because of their direct band gap, high piezoelectricity, piezoresistivity, and thermoelectricity properties, which are used as a light source and detectors, dynamic and static pressure and temperature sensors, respectively (Hjort *et al.* 1994b). These materials also take advantage of GaAs/aluminum gallium arsenide (AlGaAs) material systems, which can be monolithically grown with lattice matching, and either one of them can be selectively etched to create microstructures. The possibility of semi-insulating GaAs material facilitates easy separation of the individual devices of a circuit, which enables high-speed performance of field-effect transistors (FETs) and MEMS for the development of monolithic microwave-integrated circuit (IC) on a single chip. Thermal conductivity of GaAs is three times lower than that of silicon and thus has a better prospect of meeting the requirement for maximum thermal resistance. In GaAs, the physical mechanisms that change the resistance due to an applied stress are different from that of silicon. AlGaAs is a very interesting semiconductor with piezoelectric properties for force and resonant sensors, with high Seebeck coefficients for infrared thermopile sensors, and with a wide energy gap, when compared to silicon and GaAs, for high-temperature sensors (Dehé *et al.* 1995). Thus, excellent bolometric infrared sensors have been realized using suspended AlGaAs membranes. Recently, strain-driven, self-assembled 3D micro/nanostructures have also been fabricated on GaAs-based materials (Prinz *et al.* 2001). Because of all these features, GaAs-based materials meet the special requirements of integrated microoptoelectromechanical systems (MOEMS) for various applications.

Micromachining on InP closely resembles the techniques used for GaAs. Many of the properties of InP are similar to GaAs in terms of crystal structure and mechanical properties (Greek *et al.* 1999, Pruessner *et al.* 2003). However, the band gap of InP-based materials make them the promising candidate for light source of optical fiber communication systems because of low loss and low dispersion at 1.3 μm and 1.5 μm wavelengths. Micromachined InP/air gap distributed Bragg reflector (DBR)-based tunable vertical cavity devices such as filters, VCSELs, and photodiodes are attracting much interest because of their unique features of a wide continuous wavelength tuning and 2D array integration, thus making it possible to realize dense wavelength division multiplexing (DWDM) systems on InP.

Recently, some efforts have been directed toward employing wider band gap compound semiconductor materials such as gallium nitride (GaN) (Davies *et al.* 2004, Strittmatter *et al.* 2001, Yang *et al.* 2006) and SiC (Mehregany and Zorman 2004) for microsystems owing to their excellent electrical and mechanical properties with regard to chemical inertness and their high-temperature stability, which make them a suitable choice for MEMS applications in harsh environments.

Polysilicon is widely used for surface-micromachined MEMS structures with silicon IC compatibility, which needs high-temperature processing (>800°C). An alternative low-temperature processing material is poly-SiGe. Micromachined bolometers and thermopiles with superior thermoelectric properties have been fabricated using poly-SiGe by Dong *et al.* (2003). Freestanding SiGe micro/nanostructures such as tubes, helical coils, and bridges have also been fabricated on SiGe (Prinz *et al.* 2001).

Table 1 Structural and mechanical properties of the compound semiconductors

	GaAs	*InP*	*GaN*	*$Si_{1-x}Ge_x$*	*3C-SiC*
Crystal structure	Zincblende	Zincblende	Zincblende/wurtzite	Diamond (random alloy)	Zincblende
Lattice constant (Å)	5.6533	5.8687	4.52/a-3.189 c-5.18	$(5.431 + 0.20x + 0.027x^2)$	4.3596
Density ($g\,m^{-3}$)	5.317	4.81	6.15	$(2.329 + 3.493x - 0.499x^2)$	3.21
Melting point (°C)	1240	1060	2500	Solidus $1412 - 738x + 263x^2$	2830
Specific heat ($J\,g^{-1}\,°C^{-1}$)	0.33	0.31	0.49	0.4075 at $x = 0.75$ 0.505 at $x = 0.50$ 0.6025 at $x = 0.25$	0.69
Thermal conductivity ($W\,cm^{-1}\,°C^{-1}$)	0.55	0.68	1.3	$(0.046 + 0.084x)$	3.6
Thermal expansion coefficient ($10^{-6}\,°C^{-1}$)	5.73	4.6	$\alpha_{\parallel} = \alpha_a = 5.59/\alpha_{ort} = \alpha_c = 3.17$	$2.6 + 2.55x$ ($x < 0.85$) $7.53 - 0.89x$ ($x > 0.85$)	3.8
Debye temperature (K)	360	425	600	$640 - 226x$	1200
Elastic constants (10^{11} dyn cm^{-2})					
C_{11}	11.90	10.11	29.3/39.0	$16.58 - 3.73x$	35.23
C_{12}	5.38	5.61	15.9/14.5	$6.39 - 1.56x$	14.04
C_{44}	5.96	4.56	15.5/10.5	$7.96 - 1.28x$	23.29
[100] Young's modulus (Y_o) 10^{11} dyn cm^{-2}	8.59	6.11	18.1	$13.02 - 28.1x$	74.8
Piezoelectric constant ($C\,m^{-2}$)	$e_{14} = -0.16$	$e_{14} = -0.035$	$e_{14} = 0.4/e_{15} = -0.30/e_{31} = -0.33/e_{33} = 0.65$	$e_{14} = -0.0114$ at $x = 0.2$	$e_{15} = 0.08$ $e_{33} = 0.20$

Microsystems have been successfully implemented on compound semiconductors; however, there is a skepticism with regard to the mechanical strength of some of the III–V compound semiconductors such as GaAs- and InP-based materials for microstructures. However, measured fracture properties of GaAs are shown to be sufficiently enough, with an average fracture strength of 2.7 GPa, which is at least three times as high as that of most construction steels (Hjort *et al.* 1994a).

1.02.2 Micromachining Aspects

The basis of micromachining is similar for compound semiconductors and silicon; however, there are some key factors that influence the micromachining on compound semiconductors such as epilayer growth and etching process.

1.02.2.1 Epilayers

Advanced crystal growth techniques enable the growth of most of the compound semiconductors in well-controlled single-crystalline layers. The crystal growth techniques such as molecular beam epitaxy (MBE) and metal–organic chemical vapor deposition (MOCVD) allow epilayers to be grown on a substrate, monolayer by monolayer, with great precision. These epilayers of compound semiconductor heterostructures provide more flexibility and precision in micromachining, which results in very sharp interfaces. Due to different composition, these layers can easily be etched by wet or dry etching techniques with an excellent compositional selectivity. A precise control in thickness and composition of compound semiconductor-based micromechanical structures such as cantilevers, membranes, or bridges can be achieved directly via the thickness of the grown materials over an etch stop layer. Also, stress of epitaxial films is much more accurately controllable, by control of the lattice mismatch, than that of polycrystalline materials-based microstructures.

1.02.2.2 Etching Process

Etching plays a major role in micromachining. In general, wet and dry etching have been employed for the micromachining of compound semiconductors. The zincblende structure of the most common compound semiconductors has different etching when compared to a diamond-like silicon semiconductor. As a result, the underetching rates and profiles are controlled by the crystallographic directions and thus the sacrificial layer etching is determined by the choice of patterning orientation. Hjort *et al.* (1994a, b, 1996) have extensively investigated the basis of micromachining on III–V compound semiconductors. Various sacrificial etch systems for compound semiconductors have been documented in literature (Hjort 1996) and are listed in **Table 2**, including sacrificial AlGaAs, InGaAs, and InGaN (indium gallium nitride) layers for GaAs, InP, and GaN structures, respectively. Ion implantation-induced material modifications have also been used to form a etch stop layer in GaAs materials for micromachining (Miao and Hartnagel 2004); however, this method gives less freedom to fabricate complicated microsystems and is not cost effective.

Figure 1 shows three basic surface micromachining process steps that are as follows:

1. The device multilayer (layers B and C) structure grown by epitaxy or other deposition methods on a substrate material A (**Figure 1(a)**).
2. The device structure is photolithographically masked and then the uncovered areas are subsequently etched (vertical) to define the mesa structures. Depending on the process, slightly sloped sidewalls, vertical sidewalls, or slightly undercut sidewalls can be obtained (**Figure 1(b)**).
3. Etching sacrificial layers by selective underetching (horizontal etching) to enable freestanding mechanical structures (**Figure 1(c)**).

Selective etching is an important step in micromachining. Vertical and horizontal etching are done by wet or dry etching, which is dependent on the etch chemistry of the elemental composition of the layer for horizontal underetching; a highly selective wet chemical etching is used to underetch the sacrificial layer. To underetch a thin layer over its entire area, the selectivity must be extremely high. The critical reaction product in semiconductor etching is dissolved in hydrogen gas, which is unable to diffuse out through such a narrow gap. This gas has to be diffused out by proper rotation of etching solvents without giving pressure to the releasing microstructures.

The etching of semiconductors proceeds by the successive dissolution of layers of the materials. An additional factor arises for compound semiconductors, since they are polar semiconductors and have different surface activity, e.g., in GaAs, the (111) Ga and (111) As faces have different etching rates. Because the As face has two dangling bonds per

Table 2 Sacrificial layer wet etching of compound semiconductors by addition of new materials

Etch stop layer	*Sacrificial layer*	*Etchant*	*Etch rates (μm min^{-1}), selectivity*
GaAs	$Al_xGa_{1-x}As$, $x \geq 0.5$	$HF:H_2O$	High rate, cs > 10^6
	$Al_{0.5}In_{0.5}P$	$HCl:H_2O$	High rate, cs > 10^6
	$In_{0.5}Ga_{0.5}P$	$HCl:H_2O$	High rate, cs > 10^6
	$Al_{0.5}Ga_{0.5}P$	$HF:H_2O$	High rate, cs > 10^6
$Al_xGa_{1-x}As$, $x \geq 0.4$	GaAs	$NH_4OH:H_2O_2$	5, >100
$Al_{0.3}Ga_{0.7}As$	GaAs	Succinic acid:NH_4OH	0.2, >10^6
	GaAs	$C_6H_8O_7:H_2O_2:H_2O$	0.3, 116
	$In_{0.2}Ga_{0.8}As$	$C_6H_8O_7:H_2O_2:H_2O$	0.3, 121
AlAs	$In_{0.53}Ga_{0.47}As$	Succinic acid:NH_4OH	0.1, >1100
	$In_{0.53}Al_{0.47}As$	d:o	0.06, >550
InP	$In_{0.53}Ga_{0.47}As$	$FeCl_3:H_2O$	0.7, cs > 10^6
		$C_6H_8O_7:H_2O_2:H_2O$	0.2, 473
	$In_{0.53}Al_{0.47}As$	$C_6H_8O_7:H_2O_2:H_2O$	0.02, 102
	GaAs	$C_6H_8O_7:H_2O_2:H_2O$	0.3, 960
	$Al_{0.3}Ga_{0.7}As$	$C_6H_8O_7:H_2O_2:H_2O$	0.2, 486
	$In_{0.53}Ga_{0.47}As$	$HF:H_2OO_2:H_2O$	Low rate, cs > 10^6
$In_{0.53}Ga_{0.47}As$	$In_{0.53}Ga_{0.13}Al_{0.34}As$	$HCl:H_2O$	0.1, >329
	$In_{0.53}Al_{0.47}As$	$HCl:H_2O$	0.6, >1944
	InP	$HCl:H_2O$	High rate, cs > 10^6
InAs	Alsb	HF	High rate, cs > 10^6
$Al_{0.5}Ga_{0.5}Sb$	InAs	$C_6H_8O_7:H_2O_2:H_2O$	0.1, >3850
	GaAs	$C_6H_8O_7:H_2O_2:H_2O$	0.3, >13,650
	$GaAs_{0.85}Sb_{0.15}$	$C_6H_8O_7:H_2O_2:H_2O$	0.1, >3789
GaN	InGaN	KOH:H_2O (PEC etching)	30
SiGe	Si	$NH_4OH:H_2O$ at 75(C	1800
	SiO_2	HF	0.1
SiC	Si[a]	KOH	30
	SiO_2[a]	HF	0.1

[a]Micromolding technique.
Source: Hjort (1996), reproduced with permission from IOP Publishing Limited.cs, complete selectivity; PEC, photoelectrochemical etching.

atom whereas Ga face has no dangling bonds, the etches are much slower when compared to that for the As face.

Micromachining on III–V nitrides such as GaN, InGaN, and AlN suffers from low chemical reactivity. Therefore, III–V nitrides have been confined to dry etching processes such as reactive ion etching and plasma etching. Nevertheless, considerable progress has been made on the fabrication of micromechanical structures by photoelectrochemical (PEC) etching, which is rapidly developing as an efficient technique for micromachining on III–V nitride material systems (Strittmatter *et al.* 2001). The mechanism of PEC etching involves the creation of electron–hole pairs, the subsequent oxidative dissociation of the semiconductor into its component elements (a reaction that consumes the photogenerated holes), and the reduction of the oxidizing agent in the solution by reaction with the photogenerated electrons. The dissolution rates of a semiconductor can be altered in acid or base solutions by illumination with light ($E > E_{gap}$). Generally, n-type material is readily etched under these conditions, while p-type material is not due to the requirements for confining photogenerated holes at the semiconductor-electrolyte interface. The PEC etching mechanisms rely heavily on the absorption of incident light, and it can be defect selective, dopant selective, and band gap selective. Using InGaN as a sacrificial layer, GaN micromechanical structures have been fabricated by band gap-selective PEC etching with an illumination power of 3.5 mW cm^{-2}. Using dopant-selective PEC etching with an opaque mask, a n-GaN layer was selectively etched below p-GaN layer by illumination with light ($E > E_{gap}$) while immersed in an aqueous KOH solution with a lateral etching rate of 30 μm min^{-1} (Strittmatter *et al.* 2001).

Silicon carbide (SiC) is a very hard material and it is not etched by any acids at room temperature in its single-crystalline condition. However, KOH at molten temperature (>600°C) and some high-intensity plasma are used to etch SiC. A micromolding technique has been used to fabricate mechanical

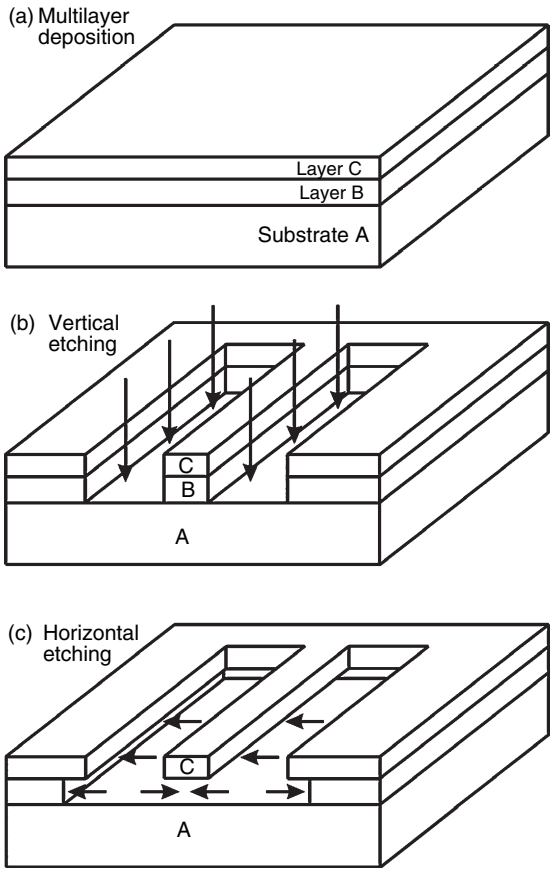


Figure 1 Three basic surface micromachining process steps: (a) multilayers structure deposition of layers B and C on substrate A; (b) vertical etching; and (c) horizontal etching (sacrificial layer etching).

structures on SiC (Yasseen *et al.* 2000). The PEC etching technique (Shor *et al.* 1992) has also been employed for SiC, which allows for selective removal of n-SiC from an underlying p-SiC layer.

In order to provide an overview on micromachining to the reader, postetching process is briefly explained here. After etching the sacrificial layer, freeze drying or critical point drying (CPD) process is used to avoid the sticking, due to the capillary pull, of the released freestanding microstructures. In freeze drying, the rinse solution is replaced through a gradual series of dilutions with *t*-butyl alcohol, which freezes at 25.6°C and then sublimated. In the CPD, the rinse solution is gradually replaced by liquid CO_2 at elevated pressures inside a high-pressure chamber and then the sample is taken to the critical point of CO_2, where the interface between the liquid and gas does not exist. This technique is highly successful with nearly 100% yield.

1.02.3 Compound Semiconductor Microsystems

1.02.3.1 Micromachined Sensors

1.02.3.1.1 Cantilevers, suspensions (beams), and membranes

Cantilevers are most commonly used in micromechanical structures in the field of MEMS as an actuator, resonator, and transducer and also as an element to evaluate the basic mechanical properties of a material. **Figure 2** shows an InP cantilever structure as an example of a compound semiconductor cantilever. This cantilever is fabricated on a (100) InP substrate of thickness 1.7 μm, typical optical waveguide thickness, using $In_{0.53}Ga_{0.47}As$ layer as a sacrificial layer (Pruessner *et al.* 2003).

Due to a higher thermal resistivity and operation at high temperatures, GaAs-based materials are able to perform electrothermal conversion with higher conversion efficiency than silicon. GaAs-based micromachined thermal converter sensors such as RF and microwave power sensors (Dehé 1996, Dehé *et al.* 1995, Lalinský *et al.* 1995, 1999, 2005), infrared thermal sensors (infrared bolometers) (Dehé *et al.* 1995, Leclercq *et al.* 1998), pressure sensors (Dehé *et al.* 1995), and thermally actuated microactuators (Lalinský *et al.* 2000a, b) have already been fabricated. Recently, piezoelectric sensing at cryogenic temperatures has been explored in FETs fabricated from GaAs/AlGaAs heterostructures (Beck *et al.* 1996), GaAs/AlGaAs cantilevers with integrated FETs (Beck *et al.* 1998), AlGaAs suspensions coupled to

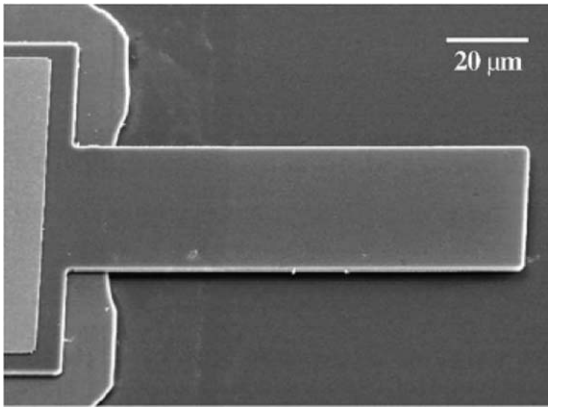


Figure 2 Micromachined InP cantilever. (Source: Reprinted from *Sens. Actuators A*, **105**, Pruessner M *et al.*, Mechanical Property measurement of InP-based MEMS for optical communications, 190–200, Copyright (2003), with permission from Elsevier.)

single-electron transistors (Knobel *et al.* 2002), and nanomechanical displacement sensing using quantum point contacts (Cleland *et al.* 2002).

For MOEMS application, GaAs/AlGaAs-based piezoelectric multimorph actuators have been proposed and designed by Ongkodjojo *et al.* (2005). This multimorph structure is designed to enhance the piezoelectric effect. More recently, Li *et al.* (2006) have demonstrated microscale piezoelectric bimorph cantilever actuators on AlGaAs (**Figure 3**). They used different Si-doping concentrations in AlGaAs to realize bottom and top electrodes as well as the piezoelectric layer in a single-lattice-matched heterostructure, virtually eliminating residual stress-induced curvature and offering the potential to reduce damping compared to traditional devices with polycrystalline piezoelectric layers and amorphous metal electrodes.

Yamaguchi *et al.* (2004) have been reporting on the magnetopiezoresistance of micromechanical cantilevers fabricated from 2D InAs/AlGaSb heterostructures, which show large enhancement of piezoresistance by quantum effects. **Figure 4** shows a suspended cantilever with a width of 4 μm and a length of 10 μm. InAs-based cantilevers systems have an additional advantage of structure-size reduction to a nanometer scale. This is a gain because the native electron accumulation in the near-surface region of InAs allows fabrication of conductive cantilevers of nanometer scale.

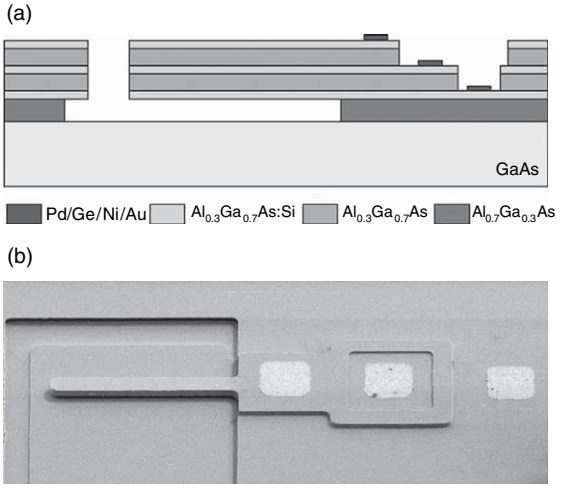


Figure 3 AlGaAs bimorph cantilever: (a) schematic cross-sectional view and (b) SEM image of a 100-μm-long cantilever. (Source: Li L, Kumar P, Kanakaraju S, DeVoe D L 2006 Piezoelectric AlGaAs bimorph microactuators. *J. Micromech. Microeng.* **16**, 1062–6; reproduced with permission from IOP Publishing Ltd.)

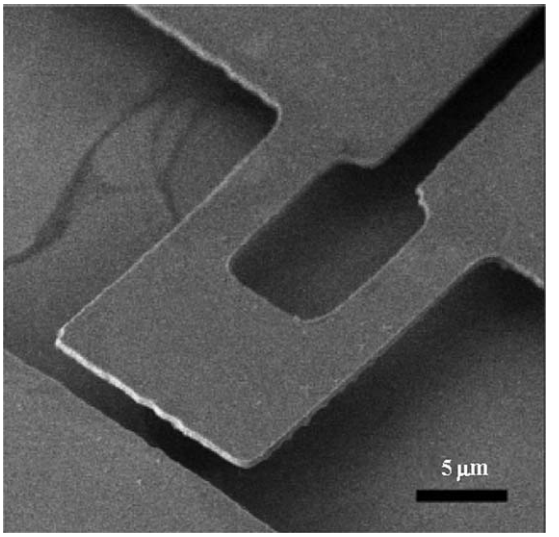


Figure 4 An InAs/AlGaSb piezoresistive cantilever. (Source: Reprinted from Yamaguchi H, Kanisawa K, Miyashita S, Hirayama Y 2004 InAs/GaAs(111)A heteroepitaxial systems. *Physica E* **23**, 285–92 Copyright (2004), with permission from Elsevier.)

Micromachining on GaAs has also been applied to enhance microwave and millimeter-wave passive devices and circuits beside the microsensors and microactuators (Han *et al.* 2006, Leclercq *et al.* 1998, Pantazis *et al.* 2005). GaAs-based micromachined membranes have been used for millimeter-wave passive circuit elements such as inductors, capacitors, transmission lines, and filters antennae. These elements need high resistive membranes to reduce high-frequency losses and parasitic capacitance. This has been implemented on semi-insulating GaAs substrates by bulk micromachining techniques. A fully monolithically integrated millimeter-wave receiver module, having the antenna as well as the detecting Schottky diodes on the same GaAs membrane, has been reported by Neculoiu *et al.* (2006).

Recently, attention has been focused to explore microsystems on III–V nitride-based materials. The III-V nitrides are emerging materials for shorter wavelengths, high temperature, and high frequency applications. The piezoelectric constants of nitrides are approximately an order of magnitude higher than that of more conventional III-V compound semiconductors such as GaAs and InP (Bykhovski *et al.* 1996, Gaska *et al.* 1997, Strittmatter *et al.* 2001). Micromachined GaN sensors such as GaN Schottky diodes and GaN MIS diodes for piezoelectric strain sensing and high electron mobility transistor (HEMT) capacitance pressure sensor have already

been fabricated (Strittmatter *et al.* 2003a, b). However, there is a hurdle to develop III–V nitride-based microsystems due to the lack of suitable underetching processes because the nitrides are chemically stable and insoluble in most common etchants at room temperature as discussed in Section 1.02.2.2. Moreover, GaN epilayers have been grown on sapphire or SiC substrates due to the lack of bulk GaN substrate. The sapphire and SiC substrates are expensive. However, recently, GaN-based materials have also been grown on silicon, which is motivated by the potential of using silicon as a less expensive substrate material. Chen and coworkers (Yang *et al.* 2005, 2006) and Davies *et al.* (2004) have fabricated GaN membrane structures on silicon substrates. Chen and coworkers have used GaN on patterned silicon (GPS) technique and fabricated suspended cantilevers array released by isotropic wet etching and a microdisc released by anisotropic wet etching. The suspended GaN microstructure, cantilevers array, is shown in **Figure 5**. They have also integrated HEMTs on the cantilevers, which can sense the change in the stress induced by deflections applied to the cantilevers. Another interesting piezoelectric III–V nitride material is AlN which is light, stiff, and piezoelectrically active that can be epitaxially grown on silicon (Cleland *et al.* 2001, Olivares *et al.* 2005). AlN is also used to integrate semiconducting electronic and surface acoustic devices (Cleland *et al.* 2001). Iborra *et al.* (2004) have demonstrated the microsystem feasibility on AlN by fabricating suspended bridge structures consisting of an AlN/polysilicon piezoelectric bimorph structures, **Figure 6**. The integration of III–V nitride microstructures on silicon takes advantage of III–V nitride-based piezoelectricity and of piezoresistivity properties of transducer elements such as cantilevers or membranes in silicon MEMS devices.

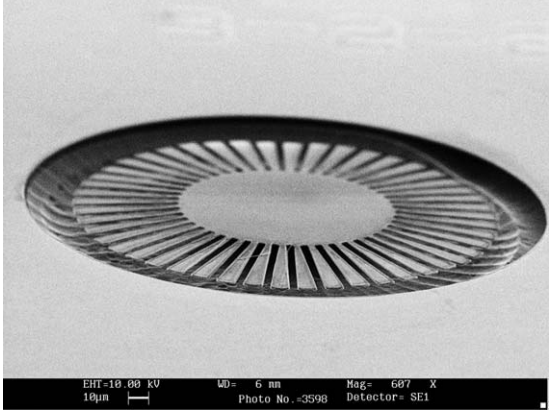


Figure 5 Suspended GaN cantilevers array released by isotropic wet etching. (Source: Yang Z, Wang R, Wang D, Zhang B, Chen K J, Lau K M 2005 GaN on patterned silicon (GPS) technique for GaN-based integrated microsensors. *IEEE, IEDM Technical Digest*, pp. 298–301; reproduced with permission from IEEE.)

1.02.3.1.2 Micromotors

Though SiC has been established as a microelectronic material for use in high-temperature and high-power applications, attention has also been focused on developing microsystems due to the outstanding chemical and mechanical properties along with its piezoresistance. SiC exists in many different polytypes; of these, the cubic 3C–SiC and the hexagonal 6H–SiC and 4H–SiC have been studied for device applications. The 3C–SiC is the only polytype that can be grown in single-crystalline form on silicon. This enables the fabrication of microsystem on 3C–SiC by taking advantage of silicon micromachining process. Mehregany and coworkers have extensively studied SiC materials for microsystems (Mehregany and Zorman 1999). Various microsystems such as resonators, piezoresistive pressure sensors, accelerometers, and micromotors have already been fabricated on SiC (Atwell *et al.* 2003, Eickhoff *et al.* 1999, Jiang *et al.* 2006, Mehregany and Zorman 1999, Roy *et al.* 2002, Wiser *et al.* 2005).

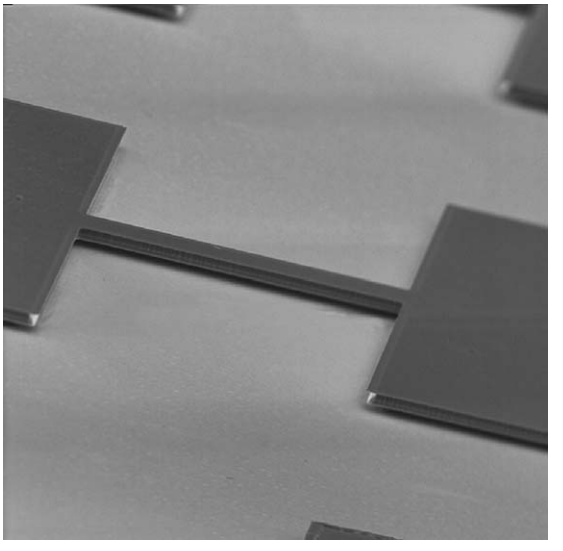


Figure 6 Microbridge structure consisting of a bimorph of AlN/polysilicon, the length and the width of the bridge are 100 and 20 μm, respectively. (Source: Reprinted from *Sens. Actuators A*, **115**, Iborra E *et al.*, Piezoelectric properties and residual stress of sputtered AlN thin films for MEMS applications, Copyright (2004), with permission from Elsevier.)

Although SiC properties are attractive for microsystems, unfortunately micromachining has been challenging because of the chemical inertness and high hardness. Due to physical process limitations, PEC and silicon bulk micromachining techniques have not been used to fabricate complex 3D bulk-micromachined structures on SiC. An alternative approach has been developed by Rajan *et al.* (1999) to create 3D bulk-micromachined structures on SiC. This approach uses silicon molds fabricated by deep reactive ion etching combined with poly-SiC deposition and silicon etching to produce fuel atomizers. Yasseen *et al.* (2000) have developed a micromolding technique that uses polysilicon molds in conjunction with poly-SiC film deposition and mechanical polishing to pattern poly-SiC films. This process has been successfully applied to fabricate micromotors as shown in **Figure 7**.

1.02.3.1.3 Microbolometers

Microbolometers are devices with a temperature-dependent resistor and an IR absorber for measuring incident electromagnetic radiation. To reach high bolometer performance, temperature coefficient of resistance (TRC) has to be high enough. Though a wide variety of materials (Richards 1994) have been used for microbolometers, IC-compatible materials have an advantage of integrating the electronic devices. GaAs and SiGe materials have TRC >4% per Kelvin, which is higher than that for silicon. Infrared microbolometers have been fabricated on AlGaAs/GaAs material systems by Dehé *et al.* (1995). Their simulated results indicate that the maximum responsivity and detectivity of $Al_xGa_{1-x}As$ is for the aluminum concentration of $x = 0.3$–0.45. One of the advantages of these bolometers is that they do not need cooling as in the case of quantum well (QW) diode detectors; however, cooling increases the performance by an order of magnitude. SiGe has high TRC and is IC compatible with silicon technology. Recently, Dong *et al.* (2003) have demonstrated an uncooled microbolometer infrared detector based on polycrystalline SiGe. The SEM image of the poly-SiGe microbolometers with two and fourfold suspended microbridge is shown in **Figure 8**. The devices were fabricated using silicon anisotropic etching, bulk micromachining, and ICP etching. The bolometers have maximum detectivity of $8.3 \times 10^8\,\mathrm{cm\,Hz^{1/2}\ W^{-1}}$ at 15 Hz with a thermal response time of 16.6 ms.

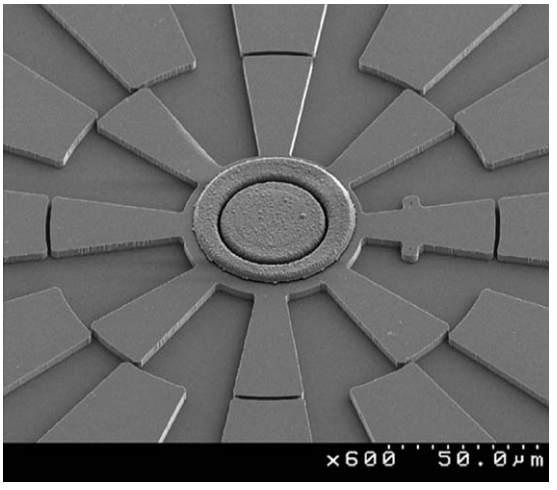


Figure 7 SiC micromotor fabricated using a multilayer surface micromachining process based on micromolding. (Source: Reprinted from *Thin Solid Films*, **355–6**, Mehregany M, Zorman C A, SiC MEMS: Opportunities and challenges for applications in harsh environments, 518–24, Copyright (1999), with permission from Elsevier.)

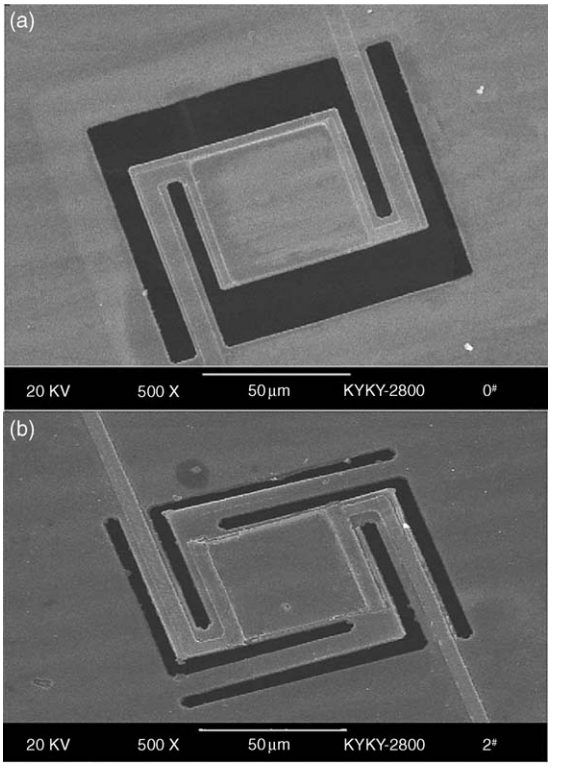


Figure 8 SiGe microbolometers with (a) two suspensions and (b) four suspensions. (Source: Reprinted from *Sens. Actuators A*, **105**, Dong L *et al.*, An uncooled microbolometer infrared detector based on poly-SiGe thermistor, 286–92, Copyright (2003), with permission from Elsevier.)

1.02.3.2 Wavelength-tunable Devices

Wide and continuous wavelength-tunable devices such as filters, lasers, and photodiodes with high spectral purity and fast tuning have a high demand for many applications such as DWDM-based optical communications, sensorics, analytics, spectroscopy, process control, metrology as well as medical diagnostics. Of these, DWDM has revolutionized telecom and datacom by increasing the signal capacity of embedded fiber, i.e., the incoming optical signals are assigned to specific wavelengths within a designated frequency band and then multiplexed onto one fiber. This process allows multiple data channels to be transmitted over one fiber while maintaining system performance and enhancing transport systems. The DWDM provides potentially unlimited transmission capacity. During the last decade, such devices have been implemented in micromachined MOEMS. Micromachining takes advantage of the vertical cavity-based tunable filters, lasers, and photodiodes to achieve wide and continuous tuning by a single tuning parameter, the actuation voltage, that varies the resonator length.

1.02.3.2.1 Micromachined tunable filters

Tunable Fabry–Pérot (FP) filters are key devices for wavelength-tunable applications. The tuning principle of electrostatically tunable micromachined vertical cavity FP optical filter is shown in Figure 9. The filter structure consists of an air cavity (length L) defined between p-doped top DBR and n-doped bottom DBR, which create a p–i–n junction. When the p–i–n junction is reversed biased, an electrostatic force acts between these two DBRs, and cavity length L decreases by increasing the reverse bias as shown in Figure 9, and the reverse bias increases from left to right. Depending on L, the filter is transparent for only one wavelength while blocking the others. Thus, the cavity length can be continuously tuned to select a specific resonance wavelength. In an ideal case, the effective cavity length is just a multiple of $\lambda/2$ of the individual wavelength.

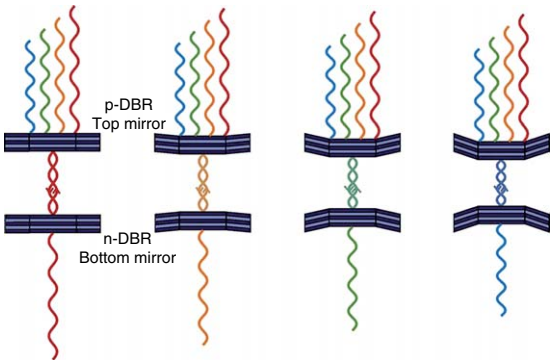


Figure 9 Tuning principle of electrostatically tunable vertical cavity Fabry–Pérot optical filters: from left to right reverse bias between the top and the bottom DBRs increases by decreasing the cavity length L.

Micromachined compound semiconductor tunable filters have been implemented on GaAs-based materials for a short wavelength of <1 μm (Amano *et al.* 2003, Lin 2004, Tayebati *et al.* 1998a, Vail *et al.* 1995) and on InP-based materials for a long wavelength of >1 μm (Dantec *et al.* 1999, Hillmer *et al.* 2002, Irmer *et al.* 2003, Strassner *et al.* 2005, Streubel *et al.* 1996). Several materials such as Si_3N_4/SiO_2, Si/SiO_2, SiO_2/TiO_2, $Al_xGa_{1-x}As/GaAs$, InP/air gap, and $AlO_x/GaAs$ have been investigated to fabricate top movable DBR tunable filters with cantilevers and four suspension membrane geometries. However, among these materials, the $AlO_x/GaAs$ and InP/air gap are promising for GaAs-based and InP-based materials, respectively, because they have high refractive index contrast, which provides a spectrally broad reflectivity spectrum. These DBR structures can also be grown monolithically, enabling a monolithic fabrication process rather than postgrown dielectric DBR structures.

InP-based tunable filters play a major role in telecom DWDM systems. An InP/air gap-based DBR structure has a very high refractive index contrast of $\Delta n = 2.167$ at 1.55 μm, enabling the achievement of a very high reflectivity with only a very few numbers of DBR periods. Figure 10 shows SEM image of a micromachined InP/air gap-tunable filter; Figure 10(a) shows a filter membrane in the middle with four suspensions connecting to supporting posts and Figure 10(b) shows a magnified view of the top and bottom DBRs (InP membranes/air gaps) separated by an air cavity. The filter structure consists of a 2.5-period p-doped top DBR, an intrinsic λ air cavity ($\lambda = 1.55$ μm), and a 3.5-period n-doped bottom DBR. Each DBR period consists of $3\lambda/4$-thick InP membrane and a $\lambda/4$-thick air gap. The air gaps and air cavity have been formed in the filter structure by selectively etching the GaInAs sacrificial layers using $FeCl_3$ etchant.

The tuning range and the required voltage for tuning depend on many parameters such as suspension lengths, number of suspensions, suspension widths, membrane diameter, membrane and suspension thickness as well as cavity length. Filters having

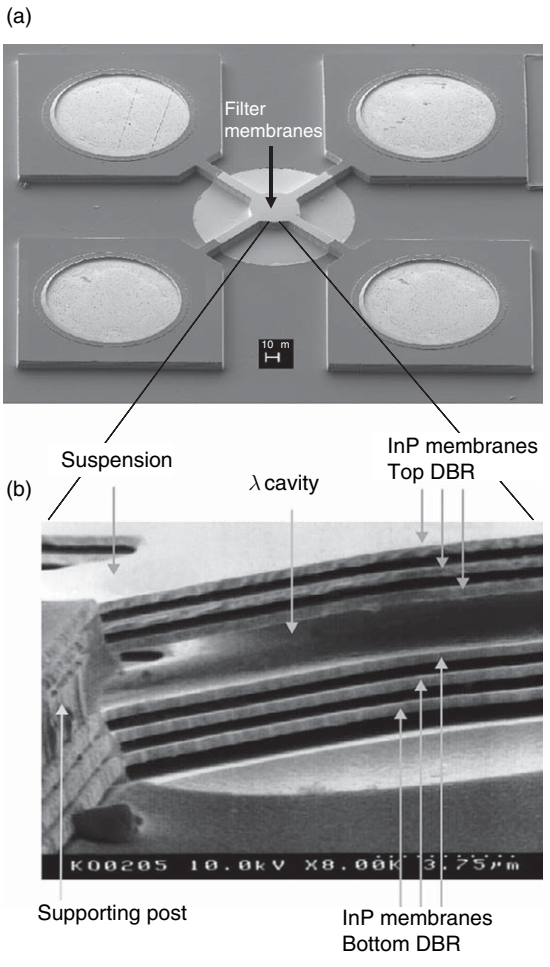


Figure 10 Micromachined multiple InP membranes/air gaps filter, (a) filter membrane in the middle with four suspensions connected to the supporting post and (b) resonator details.

four suspensions of 40 μm length, 40 μm membrane diameter, three $3\lambda/4$ InP membranes, separated by $\lambda/4$ air gaps, a $\lambda/2$ cavity and an air gap 637.5 nm to the InP substrate show a tuning range of 142 nm with a low tuning voltage of 3.2 V (Hillmer *et al.* 2002). The transmission dip of the device is located at $\lambda = 1.599\mu m$ for nonactuated membranes ($V = 0$) and at $\lambda = 1.457\mu$m. Thus, a continuous tuning of >9% of the absolute wavelength is obtained. This tuning range already exceeds the spectral width of the unchirped QW semiconductor laser gain profiles, typically in the range of 110 nm. Due to miniaturization, the efficiency of the electrostatic force increases considerably (Hillmer *et al.* 2002). Another filter with four suspensions of 30 μm length, 20 μm membrane diameter, three $3\lambda/4$ InP membranes, separated by $\lambda/4$ air gaps, a λ cavity and a spacing of 465 nm to the InP substrate, and even a larger tuning range has been obtained. The filter is weakly strained and thus at $V = 0$ has curved membranes. Using an electrostatic tuning voltage of 28 V they obtained 221 nm tuning (**Figure 11**), and the inset shows the corresponding reflectance spectra (Hasse *et al.* 2006). These filters have a fast tuning, low actuation voltages, and extremely wide tuning ranges; they are batch process-compatible and thus have a low-cost production. Therefore, these filters have a potential application in DWDM systems.

One of the issues in the InP/air gap DBR is to realize a strain-free InP membrane due to the arsenic carryover arsenic memory effect during the growth of the InP/GaInAs epilayers structure. Special care has to be taken to ensure optimum lattice matching in the epitaxial growth process since compressive strain and vertical strain gradients are very critical, leading to undesired bends in the underetched membranes. The arsenic incorporation in the InP layer creates a compressive strain in the InP membranes and causes slight bending in the membranes. As a consequence of the membrane bending a remarkable influence exists in the filter properties such as spectral tuning range, filter line transmission, line width, and lateral mode structure. Arsenic carryover effects on InP

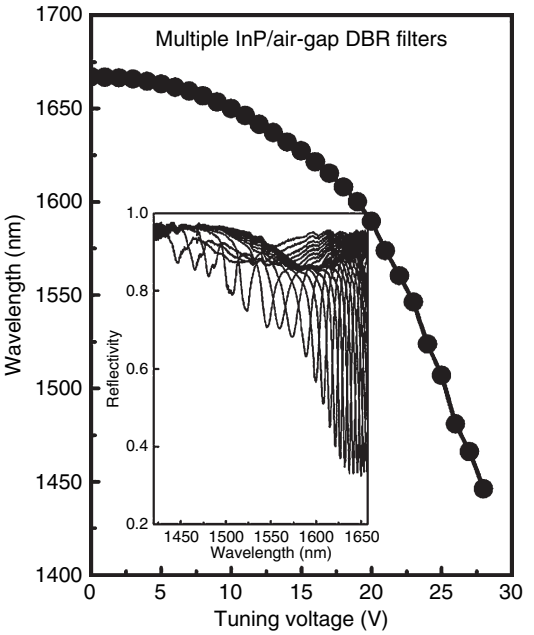


Figure 11 Measured tuning characteristics and the corresponding reflectance spectra (*inset*) of the InP/air gap filter under electrostatic actuation.

membranes have been investigated in detail by Strassner *et al.* (2000). They proposed growth interruption and compensation mechanism to compensate the arsenic carryover-induced compressive strain in the InP membranes. Recently, Kussorow *et al.*, (2006) have carried out a systematic investigation on the compensation layer thickness as a function of compressive strain in the InP membrane.

An alternative configuration for GaAs-based micromachined tunable cantilever filter structure with thermal tuning was proposed and implemented by Amano *et al.* (2003) for short wavelength applications. This filter structure is controlled by a thermal strain force and has a tuning range of 18.1 nm for the tuning voltage of 4.9 V (0.79 nm K^{-1}). The dynamic response of this kind of filters is lower than that of electrostatically actuated filters.

Table 3 compares the micromachined electrostatically tunable filters of compound semiconductors with multiple-period DBRs and shows the development progress over a decade. One of the important properties of a tunable filter is the full width at half maximum (FWHM), which influences the finesse of the wavelength tunability. Especially for DWDM applications a small FWHM of 0.8 nm is necessary. A more stable filter membrane geometry and flat membrane (without bending) will give higher finesse. A detailed comparison of dielectric-based DBR structure tunable filters from the DWDM application point of view has been discussed by Hillmer *et al.* (2006).

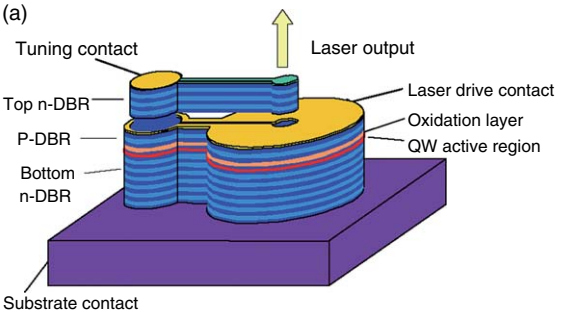
1.02.3.2.2 High-performance tunable VCSELs

The most exciting application of micromachining in direct band gap compound semiconductors is the monolithically fabricated wavelength-tunable light sources. The tunable lasers enable dynamic networks with wavelength reconfiguration, networking flexibility, and cost reduction. Micromachining takes advantage of the vertical cavity geometry of VCSEL and allows mechanically tunable laser diodes with a single tuning parameter (actuation voltage). The top movable DBRs have been fabricated by micromachining the active region of the VCSELs, which can be electrostatically actuated to tune the wavelength by changing the cavity length. This device gives a circular beam profile and enables 2D integration and batch processing.

Several research groups have been reporting various approaches to implement mechanically tunable lasers. Of these, cantilever VCSEL (c-VCSEL) by Chang-Hasnain (2000) and membrane VCSEL (m-VCSEL) by Harris (2000) are investigated in detail on GaAs-based materials, which are based on a monolithic approach. **Figure 12(a)** and **12(b)** shows the schematic diagram of a top emitting tunable c-VCSEL with a cantilever, consisting of a movable top DBR and a fixed bottom DBR, a cavity layer with an active region, and a SEM image of the fabricated device, respectively (Chang-Hasnain 2001). The top mirror consists of three parts: a p-DBR, an air gap, and a top n-DBR that is freely suspended above the laser cavity and supported by the cantilever structure. **Figure 13** shows the continuous tuning spectra of the c-VCSEL with a single tuning control across 20 channels in C-band at a spacing of 100 GHz with side-mode suppression ratio >45 dB. The advantages of c-VCSEL are that the complete epilayer structure can be grown in one single step and electrically pumped monolithic wavelength-tunable device. The disadvantage is polarization dependence. Tuning is achieved with a reverse-biased junction, which is embedded in the top DBR.

Table 3 Comparison and progress of compound semiconductors based micromachined electrostatically tunable filters

DBR material system	*Air gaps*	*Starting wavelength, λ (nm)*	*Tuning range, Δλ (nm)*	*Actuation voltage (V)*	*FWHM (nm)*	*References*
AlAs/GaAs	1	937	32	14	3	Larson *et al*. (1995)
AlAs/GaAs	1	972	70	5	8	Vail *et al*. (1995)
AlAs/GaAs	1	1633	103	35	1.2	Peerlings *et al*. (1997)
InP/air	6	1550	62	14	0.6–3	Dantec *et al*. (1999)
InP/air	4	1470	65	21	4.3	Strassner *et al*. (1999)
AlAs/GaAs	1	1600	100	23	1	Mateus *et al*. (2002)
InP/air	6	1600	142	3.2	1–3	Irmer *et al*. (2003)
InP/air	6	1528	48	8	0.4	Strassner *et al*. (2005)
InP/air	6	–	160	2.3	–	Strassner *et al*. (2005)
InP/air	6	1660	221	28	4–15	Hasse *et al*. (2006)



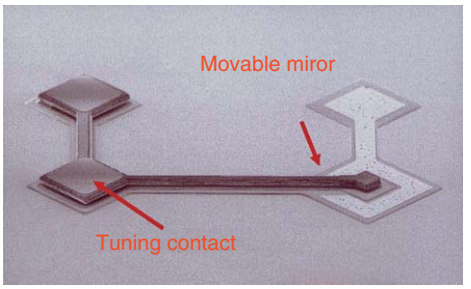


Figure 12 (a) A schematic diagram of a top emitting tunable c-VCSEL consisting of top DBR (cantilever) and fixed bottom DBR, and (b) an SEM image of the fabricated device. (Source: Chang-Hasnain J C 2001 Tunable VCSELs: Enabling wavelength-on-demand in metro networks. *Compound Semiconductors Magazine*; reproduced with permission from IOP Publishing Ltd.)

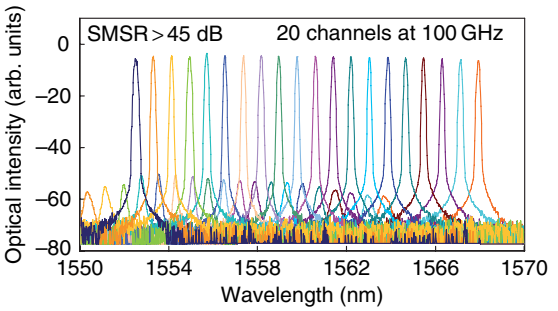


Figure 13 Tuning spectra of a c-VCSEL. (Source: Chang-Hasnain J C 2001 Tunable VCSELs: Enabling wavelength-on-demand in metro networks. *Compound semiconductors Magazine*; reproduced with permission from IOP Publishing Ltd.)

Harris and coworkers have proposed a membrane-based movable top DBR and implemented tunable VCSELs (Harris 2000, Larson and Harris 1996, Larson *et al.* 1995). A schematic diagram of the tunable m-VCSEL and lasing wavelength tuned near 960 nm are shown in **Figures 14** and **15**, respectively. The basic tuning mechanism of c-VCSEL and m-VCSEL is the same, i.e., the cavity length is decreased as a

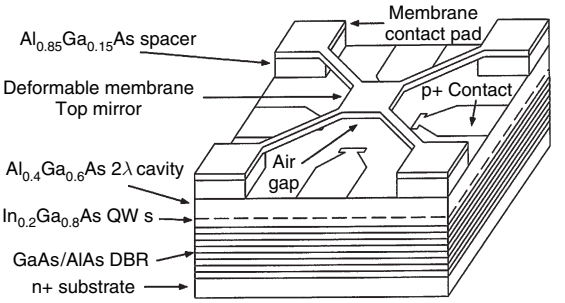


Figure 14 Schematic diagram of the tunable m-VCSEL. (Source: Reprinted with permission from Larson M C, Harris J S Jr. 1996 Wide and continuous wavelength tuning in a vertical-cavity surface emitting laser using a micromachined deformable-membrane mirror. *Appl. Phys. Lett*. **68**, 891–3. Copyright (1996), American Institute of Physics.)

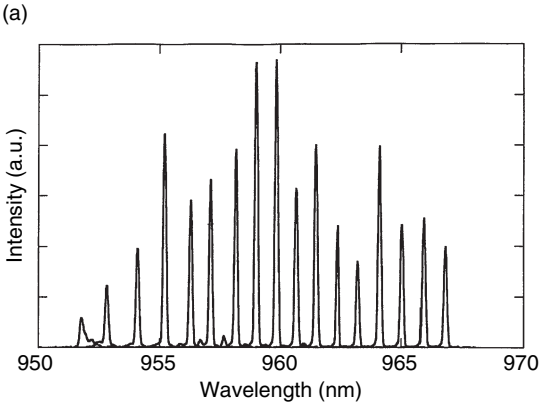


Figure 15 Tuning spectra of an m-VCSEL. (Source: Reprinted with permission from Larson M C, Harris J S Jr. 1996 Wide and continuous wavelength tuning in a vertical-cavity surface emitting laser using a micromachined deformable-membrane mirror. *Appl. Phys. Lett*. **68**, 891–3, Copyright (1996), American Institute of Physics.)

function of tuning voltage. However, the micromachined top mirror (membrane) geometry and materials are different from c-VCSEL. The top mirror consists of four suspensions extending from four posts at a diagonal position to support the central mirror. Two different material systems have been used for the top DBR mirrors, one is a semitransparent Au reflector/electrode on top of a SiN_xH_y phase-matching layer and GaAs cap, and the other is Al_2O_3/GaAs-based. The Al_2O_3/GaAs system has low loss, high refractive index contrast, and high mechanical strength when compared to other materials.

Tayebati *et al.* (1998b) have reported a new class of tunable VCSELs for short wavelength application in which the spatial mode of the laser oscillation is

controlled by the curvature of the actuable top DBR. The device tuned from 863 to 833 nm under a 13-V voltage swing.

Recently, micromachined tunable vertical cavity semiconductor optical amplifiers (VCSOA) have also been reported (Cole *et al.* 2005). The difference between VCSOA and VCSEL is the reduced mirror reflectivity used in the resonant cavity structure of VCSOA. The VCSOA is an alternative to erbium-doped fiber amplifiers and conventional in-plane SOAs for use in fiber-optic communication systems, and it is also of low cost.

1.02.3.2.3 Micromachined tunable hybrid dielectric DBR structures

The InP-based tunable long-wavelength VCSELs are highly desirable for telecommunication applications; however, few devices based on monolithic approach operating at 1.55 μm have been reported (Chang-Hasnain 2000, Sun *et al.* 2004, Vakhshoori *et al.* 1999). Some of the major design issues that have to be overcome are the low index contrast and poor thermal conductivity of lattice-matched DBRs on InP materials. Therefore, hybrid integrated two-chip concept-based dielectric top DBR tunable VCSELs have been proposed and demonstrated for long-wavelength applications (Kögel *et al.* 2006, Maute *et al.* 2004, Riemenschneider *et al.* 2003, 2004, Tarraf *et al.* 2004, Tayebati *et al.* 1998b). The two-chip concept devices are based on an active region and an air gap cavity embedded between two DBRs of different materials. This has the advantage that both parts can be optimized separately and then be integrated.

Tarraf *et al.* (2004) have carried out a detailed investigation on the stress in the dielectric (Si_3N_4/SiO_2) top DBR mirrors for the two-chip hybrid integrated VCSELs. They have implemented micromachined continuously tunable 1.55-μm VCSELs by precisely curved dielectric top DBRs using the two-chip concept. The curved top DBR structure is designed for a precise lateral mode control. The dielectric DBR period comprises a Si_3N_4 and a SiO_2 layer, each with an optical thickness of $\lambda/4$ at 1550 nm. The air gap length L and the membrane curvature were adjusted by precisely tailoring the intrinsic stress of the dielectric layers across the DBR by varying the duty cycle plasma excitation frequency intervals during the deposition in the plasma-enhanced chemical vapor deposition (PECVD) system. **Figure 16** shows a schematic cross section of the device. The correlation between the design of the membrane and the vertical distribution of the stress across the DBR forces the membrane to

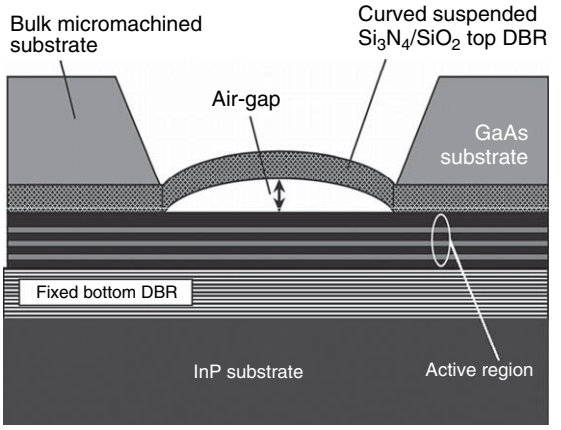


Figure 16 Cross-sectional view of the two-chip concept-based tunable VCSEL.

bend toward the GaAs substrate after substrate removal. Thus, a radius of curvature of 4.5 mm and an air gap length L of 16 μm are obtained. The circular membrane has a diameter of 300 μm and is fixed to the substrate by four suspensions (600 μm length and 70 μm width). The bottom DBR is designed for a central wavelength of 1575 nm and consists of 48 InP/InGaAsP pairs yielding 99.8 % nominal reflectivity. The semiconductor part of the resonator has an optical thickness of 4.5λ and consists of a periodic gain active region optimized for optical pumping. The total cavity (length L) consists of a $3\lambda/2$ long active region, embedded by two InP confinement layers and the air gap. The spectral tuning is achieved through micromechanical actuation of the membrane. The deflection is achieved electrothermally by injecting a small heating current through a thin metallic layer on top of the flexible suspensions. The metallic layer acts as a heating resistor, converting the dissipated power into thermal heat, which increases the length of the suspensions slightly. The VCSEL is optically pumped and shows a continuous tuning range of 26 nm. The device has a free spectral range (FSR) of 47 nm and a side-mode suppression ratio of 57 dBm. The laser peak has a line width (FWHM) of 0.1 nm, resulting in a finesse of approximately 470, which is the ratio of FSR to FWHM. The relation between wavelength tuning and heating power due to thermal actuation is exactly linear with a sensitivity of $7\,\text{nm}\,\text{mW}^{-1}$ (Tarraf *et al.* 2004).

Maute *et al.* (2004) and Riemenschneider *et al.* (2004) have reported electrically pumped two-chip concept-based long-wavelength-tunable VCSELs. The VCSELs consist of a single air gap, a GaAs/AlGaAs top mirror with 24.5 pairs, and a dielectric CaF_2/a-Si DBR bottom mirror (2.5 pairs) with an additional gold

coating, incorporating a tunnel junction for improved current injection, and reduced optical loss (**Figure 17**). The top mirror has a concave bending with a radius of curvature of 5.5 mm, resulting in an initial deflection of 6.5 μm after removing the substrate. By injecting a small heating current through the doped membrane material the flexible suspension warms up and slightly expands and thus the clamped membrane deflects and the VCSEL resonator increases. This tunable VCSEL shows a continuous tuning over 40 nm at long wavelengths. The gold mirror, in combination with a thin dielectric DBR, leads to a high reflectivity and to a low thermal resistance for transferring the heat out of the active region. This device appears to be a promising candidate of its kind for tuning at long wavelengths.

In the following, different FP filter designs are compared with respect to finesse, FSR (corresponding to the usable tuning range), and FWHM (corresponding to the linewidth of the filter dip). In a rough approximation, both FSR and FWHM are proportional to the inverse cavity length L. Since the finesse is the FSR divided by the FWHM, the finesse does not depend on L in this theoretical model. Larger cavities (e.g., two-chip concepts) theoretically show not only smaller FWHM (which is desirable) but also smaller FSR (which is not desirable). For small cavity lengths (e.g., monolithic single-chip concepts) the situation is reversed. Therefore, a counteracting situation reveals that small cavities are beneficial for larger FSR and larger tunability, and larger cavities are beneficial for smaller FWHM. In practice the theoretically predicted FSR values are also found experimentally. On the contrary, in practice it is much more difficult to obtain the theoretically predicted low FWHM values. Due to membrane bending, stress, interface and surface roughness, and individual layer thickness fluctuations, the FWHM is often much higher than theoretically predicted. This in practice causes a dependence of the finesse on cavity length. In summary, larger cavities (e.g., two-chip concepts) show in practice smaller finesse than do smaller cavities (monolithic single-chip concepts). When membrane and cantilever-based FP filters are compared, the membrane-based devices show larger finesse and much lower polarization dependence.

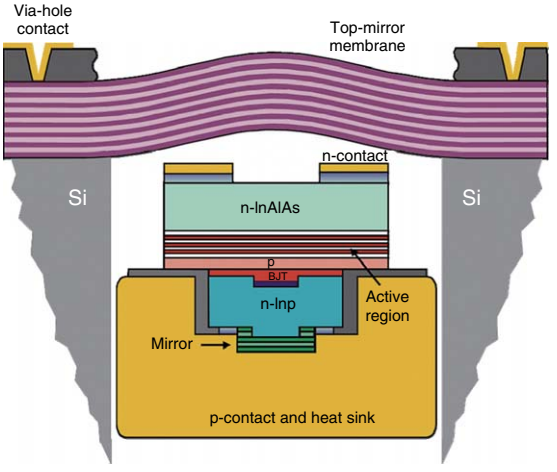


Figure 17 Schematic cross-sectional view of two-chip tunable VCSEL. (Source: Maute M, Böhm G, Amann MC, Riemenschneider F, Halbritter H, Meissner P 2004a Micromechanically and widely tunable long-wavelength VCSELs. *IEEE 19th International Semiconductor Laser Conference*, 119–20, © 2004, reproduced with permission from IEEE.)

Wavelength tuning was also achieved by refractive index modulation by changing the temperature and current injection. Temperature tuning is limited to low tuning range and slow speed. Current tuning is complex and continuous tuning is difficult. Therefore, electrostatic mechanical tuning has greater tuning range than that achieved by refractive index modulation.

1.02.3.2.4 Tunable photodiodes

Optically transmitted signals are finally converted into electronic signals by a photodiode. A highly sensitive high-speed photodiode is necessary for optical communication and sensing application. Especially for optical communication, a highly selective and wide continuously tunable photodiode is necessary for wavelength tracking in the DWDM system, which enables ultrahigh capacity transmission. Various kinds of photodiodes such as p–i–n photodiodes (PIN-PDs) and metal–semiconductor–metal photodiodes have been reported in the literature. The PIN-PD allows tunable photodiodes because of the geometry, which can be integrated with micromachined movable top mirrors. There are also reports on tunable photodiodes using quantum confined Stark-effect; however, this is limited in the tuning range.

Micromachined tunable photodiodes design basically can have two different choices, both related to the position of the absorption layer: resonant cavity-enhanced (RCE) photodiode or FP filter photodiode. In the RCE, an absorption layer is placed inside the cavity and the standing wave intensity increases sharply at resonance and so does the photodiode response. In the FP filter photodiode two separate devices can be considered: filter and photodiode where the filter is tuned to a desired wavelength to pass to the photodiode.

Resonant cavity-enhanced QW tunable photodiodes for long wavelength have been proposed and implemented by Leclercq *et al.* (2000). In these devices the absorption layer is placed between the top dielectric DBR mirror and the bottom InP/air gap DBR mirror. The preliminary results of this device show 9 nm selectivity, 18% quantum efficiency with a dark current of 10^{-6} A cm^{-2} at 2 V.

Wu *et al.* (1996) have developed a tunable micromechanical vertical cavity photodiode using a cantilever as a top DBR mirror, which has a geometry similar to their c-VCSEL geometry. The device consists of, starting from the substrate, 13-period n-doped AlAs/GaAs fixed bottom DBR, an absorbing region, which is sandwiched between the top and the bottom DBRs, formed from three 8-nm $In_{0.16}Ga_{0.84}As$ strained QWs in a λ cavity, a fixed portion of the top 2-period p+-doped $Al_{0.6}Ga_{0.4}As/Al_{0.1}Ga_{0.9}As$ DBR, and a movable top mirror with 18-periods n-doped $Al_{0.6}Ga_{0.4}As/Al_{0.1}Ga_{0.9}As$ DBRs. By applying a tuning voltage to create electrostatic attraction, the cantilever is attracted toward the substrate, thereby changing the air gap spacer thickness and consequently the resonant FP wavelength. They have achieved 30-nm continuous tuning for the resonant cavity detector with a 7-V tuning voltage.

Hybrid integrated long-wavelength-resonant vertical cavity photodiode have also been fabricated by Christenson *et al.* (1997). The device consists of a GaAs-based DBR, i.e., wafer bonded to InP-based active layers, with a micromachined tunable top dielectric DBR mirror to produce the wavelength shift. A tuning range of 75 nm has been achieved for a voltage of 50 V, which is relatively high. Another hybrid integrated two-chip concept-based long wavelength-tunable photodiode by placing the PIN-PD outside the resonant cavity has been reported by Peerlings *et al.* (1999). The device consists of a PIN detector part and a micromachined actuator part. The membrane mirror of the actuator, supported by two beams, is realized by selective wet chemical etching of the bulk InP substrate. The device geometry allows the use of dielectric DBR mirrors as it is a two-chip concept. Micromachining has also been used to fabricate resonant cavity-enhanced untuned photodiodes by fabricating InP/air gap DBRs (Huang *et al.* 2004). The main advantage of this device is a high detection efficiency at resonance.

1.02.3.2.5 Optical waveguide switches

All-optical technology uses photons to carry the signal, which escalates the limitation imposed by the present electronic technology. In all-optical technology, optical fibers transmit information over long distances and micromachined optical switches manipulate the optical signals directly rather than converting them into the electrical domain for manipulation, thereby increasing the network speed together with greater channel density. The advantages include reduced crosstalk, noise, and power dissipation at high frequencies when compared to electrical interconnects. All-optical technology also enables optical ICs on a single chip where MEMS play a key role because optics relies on the precision motion that MEMS can give. In all-optical systems, electronics control the optical signals from outside the optical signal path. This enables system implementers to optimize electronics and optics independently to achieve the all-optical integration. As in electronic ICs, more function can be added into the all-optical IC in a single chip.

Optical waveguides have been utilized to tight optical confinement for dense integration. Such integrated optics enables the monolithic integration of many optical components in a compact device. Micromachining enables movable and actuated waveguides. Optical waveguide MEMS switches play a vital role in light propagations, network reconfigurations, and optical cross connections. This device provides light propagation in the wafer plane, allowing dense integration, monolithic fabrication, small size and low cost, and increased manufacturability due to decreased assembly and packaging requirements. There have been numerous optical waveguide MEMS devices implemented in silicon; these devices are limited to passive optics due to the indirect band gap in silicon. Recently, optical waveguide MEMS switches have been implemented on GaAs and InP compound semiconductors.

GaAs-based waveguide MEMS switches have been demonstrated by Spahn and coworkers (Bakke *et al.* 2002). They have demonstrated 1×2 switches based on lateral deflection of a planar waveguide (**Figure 18**). The waveguide consist of 4-μm-thick GaAs/AlGaAs layers while the released layer is composed of 2 μm of AlGaAs. The switch characteristics show low actuation voltage (3–20 V), short switching times (32–200 μs), and low crosstalk (30 dB).

Ghodssi and coworkers have demonstrated micromachined InP waveguide switches, which have potential application in optical communications as well as in sensors by taking advantage of the direct band gap. They have developed suspended waveguides supported by lateral tethers (Kelly *et al.* 2004), optical waveguide MEMS switches with evanescent coupling mechanism (Pruessner *et al.* 2005), and electrostatically actuated end-coupled optical waveguide devices on

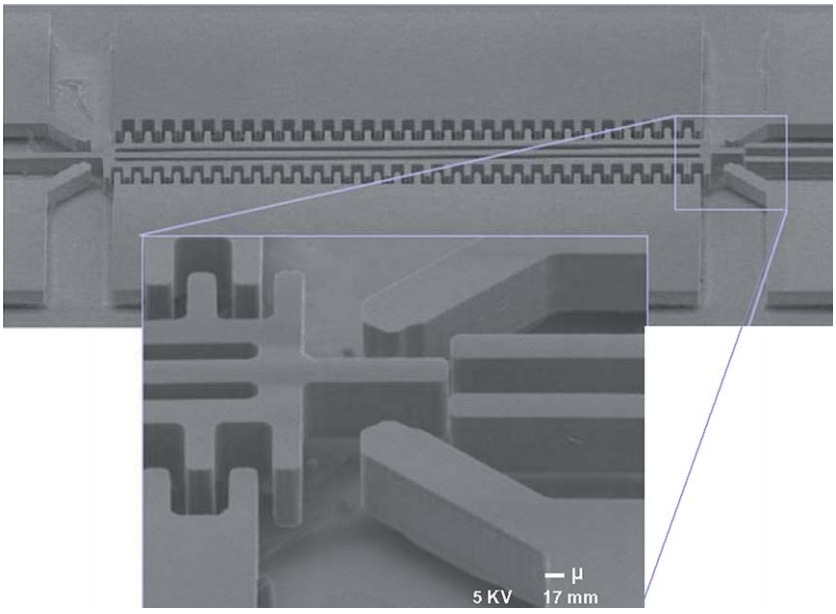


Figure 18 GaAs optical waveguide switch. (Source: Bakke T, Tigges C P, Lean J J, Sullivan C T, Spahn O B 2002 Planar microoptomechanical waveguide switches. *IEEE J. Select. Top. Quantum Electron*. **8**, 64–72, © 2002, reproduced with permission from IEEE.)

InP. The optical switch with evanescent coupling mechanism has two closely spaced waveguides, and the coupling is controlled by changing the gap and the coupling length between the two waveguides via electrostatic pull-in. This enables both optical switching and variable optical coupling at voltages below 10 V (Pruessner *et al.* 2006). This device has a channel isolation as high as 47 dB and coupling efficiencies of up to 66% with switching losses of <0.5 dB. **Figure 19** shows a 1×2 end-coupled optical waveguide switch, which has a comb-drive actuator and serpentine suspension for low-voltage switching of <7 V. They have also developed resonant sensors with integrated optical readout by modifying the optical waveguide switch (Pruessner *et al.* 2005). Using this approach, they have also demonstrated electrostatically actuated in-plane (horizontal) tunable FP filters with integrated waveguides revealing a tuning range of ~400 nm, which covers the entire CWDM spectral range (Datta *et al.* 2004). This filter geometry allows monolithic integration of InP-based devices such as edge-emitting lasers, waveguides, and photodiodes on single chips.

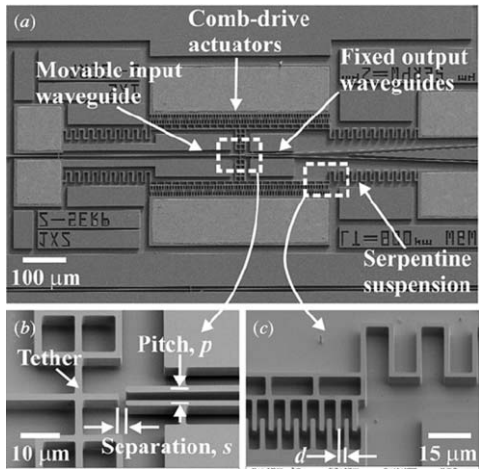


Figure 19 (a) End-coupled 1×2 switch, (b) details of movable input and fixed output waveguides, and (c) details of comb-drive and serpentine suspension. (Source: Pruessner M W, Siwak N, Amarnath K, Kanakaraju S, Chuang W H, Ghodssi R 2006 End-coupled optical waveguide MEMS devices in the indium phosphide material system. *J. Micromech. Microeng*. **16**, 832–42, © 2006, reproduced with permission from IOP Publishing Ltd.)

1.02.3.3 Hybrid Integrated Microsystems

In this chapter, compound semiconductor-based microsystems have been discussed; however, it is worthy to mention hybrid integrated microsystems, which give great opportunity for developing on-chip mechanical components that can be integrated with optical devices for both optical communication and optical sensing systems. Though monolithic integration is desired, as discussed in other sections of this

chapter, in reality, highly functional optical systems need micromachined hybrid (different individual devices of different main materials) integrated microsystems such as optical scanners (Sasaki *et al.* 2003), laser-to-fiber coupling (Ishikawa *et al.* 2003), external cavity continuous tunable lasers (Liu *et al.* 2003, 2004), and a free-space micro-optical bench (Lin *et al.* 1997). More complicated micromachined hybrid integrated systems such as Smart Dust has been conceptually proposed and partially implemented by Kahn *et al.* (2000). The Smart Dust consists of microfabricated sensors, optical receivers, passive and active optical transmitters, signal processing, control circuitry, and power on a single chip.

The hybrid integrated microsystems mostly take MEMS devices from silicon and optical devices from direct band gap compound semiconductors. The silicon and compound semiconductor each possess desired properties that the other does not; there have been many efforts to grow III–V material directly on silicon substrates in order to take advantage of each material's capabilities on a single chip. Unfortunately, due to large differences in the lattice constants, and the thermal expansion coefficients of these materials, defect densities in the heteroepilayer are high and device performance results are poor. To circumvent this issue, hybrid integration has been adopted to integrate compound semiconductor optical devices onto silicon MEMS. Hybrid integration needs a special bonding technology such as wafer bonding, epilayer transfer/grafting, or flip chip bonding. The choice of the bonding method depends on the initial substrate and the final applications. A heteromicromachining concept is also implemented by Fricke *et al.* (1999), which integrates compound semiconductors with silicon by using etching selectivity of these two materials.

A multidisciplinary team of end users and suppliers has collaborated for developing a novel broadly enabling process for the design and the fabrication of hybrid integrated light emitters and detectors with silicon on insulator (SOI)-based MOEMS (Kubby *et al.* 2001). This system has crystalline silicon optical components (mirrors and lenses), which can be rotated out of plane using hinges defined in polysilicon layers. With these mechanical devices, laser diodes and control as well as detection electronics have also been integrated using flip chip bonding.

Sasaki *et al.* (2003) have demonstrated an optical scanner on a 3D micro-optical bench by 3D photolithography using resist spray-coating techniques on silicon wafers. This system consists of an edge-emitting laser chip mounted on the bulk silicon terrace facing the scanning mirror. The laser emission is reflected to the out-of-plane direction by the skew mirror. An optical scan angle of more than 30° and a cutoff frequency of 100 Hz have been reported. This is close to the specification required for barcode readers.

A laser-to-fiber coupling is a critical task in the optical fiber network. This has been successfully achieved by a hybrid integrated micromechanical mirror with an optical laser diode (**Figure 20**) (Ishikawa *et al.* 2003). The advantages of this system are active alignment, batch assembling, and compactness.

Continuous tunable lasers have also been developed by hybrid integration using silicon-based comb-drive actuators. Uenishi (1996) have reported micromachined nickel micromirrors to compact coupled-cavity edge-emitting lasers. The micromirror has a height of 19 μm and a width of 50 μm, and the facet reflectivity was estimated to be 63%. Recently, Liu *et al.* (2003, 2004) have developed an external cavity continuous tunable lasers using micromachined blazed grating, circular/flat mirror integrated with fiber onto a single silicon chip (**Figure 21**). The device with blazed grating has a tuning range of 30.3 nm with a resolution of 0.03 nm V^{-2}, a maximum power of 0.4 dBm, and a side-mode suppression ratio of 26 dB. The circular mirror works in multilongitudinal modes. More recently, Li *et al.* (2005) have achieved hybrid integrated photonic subsystems for optical network application by integrating tunable lasers and optical switches on a single silicon chip with a dimension of 3.5 mm × 3 mm × 0.6 mm. The hybrid integrated microsystems are, generally, limited to miniaturization due to manual assembling of the optoelectromechanical compounds.

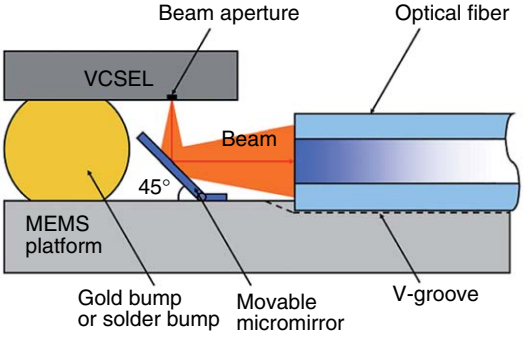


Figure 20 A schematic cross sectional view of VCSEL-to-fiber coupling by hybrid integrated microsystems. (Source: Reprinted from *Sens. Actuators A*, **103**, Ishikawa K, *et al.*, An integrated micro-optical system for VCSEL-to-fiber active alignment, 109–15, Copyright (2003), with permission from Elsevier.)

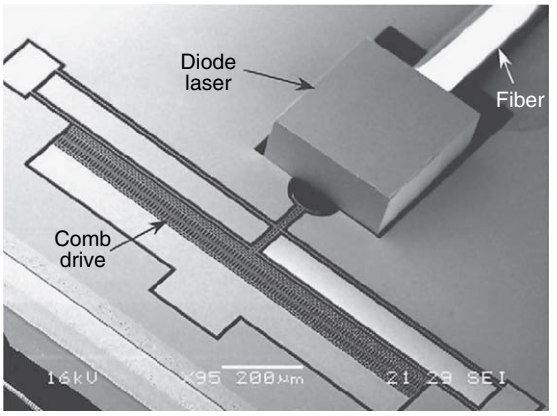


Figure 21 Hybrid integrated silicon MEMS (comb-drive) with laser diode and optical fiber. (Source: Reprinted from *Sens. Actuators A*, **108**, Liu A Q *et al*., Single-/multi-mode tunable lasers using MEMS mirror and grating, 49–54, Copyright (2003), with permission from Elsevier.)

1.02.4 Self-Assembled Micro/Nanostructures

Micromachined self-assembled structure on compound semiconductor materials is an exciting and less-explored technique. First, Prinz *et al.* (1996) demonstrated that strained bilayers can be used to form micro/nanotubes and coils. The structures were formed by lattice-mismatched epitaxial layers, for example, GaAs/InGaAs, that rolled up when released from the substrate due to the lattice mismatch-induced built-in strain. The strain-induced deformation is determined by the lattice mismatch, elastic properties, and the thickness of the bilayers. By taking advantage of this technique, various micro/nanostructures have been reported on InGaAs/GaAs and SiGe/Si material systems by Prinz *et al.* (2001), Prinz (2004), and Schmidt *et al.* (2001, 2002).

1.02.4.1 Self-Assembly Mechanism

The basic principle of a self-assembled technique is as follows: when two single-crystalline epitaxial layers of materials A and B with different lattice constant a ($a_A < a_B$) are grown pseudomorphically on a substrate of material A (e.g., GaAs/InGaAs on GaAs or Si/SiGe on Si), the layer B is biaxially compressed in the plane of the substrate. The strained bilayer, consisting of the layers A and B, is released from the substrate by selective etching of an additional sacrificial layer, the freestanding bilayer bends with a curvature radius owing to relaxation of mechanical stress. The interatomic forces in the bilayer act to increase the interatomic distance in the compressed B layer and to decrease it in the tensile-stressed A layer (**Figure 22**). The elastic forces F_1 and F_2 are oppositely directed, and they give rise to a nonzero moment of forces M, which tend to bend the bilayer. Under the action of this moment, the initially planar bilayer rolls up in a scroll. The rolled-up layers stick together, thus forming a perfectly bonded tube wall (**Figure 23**). The

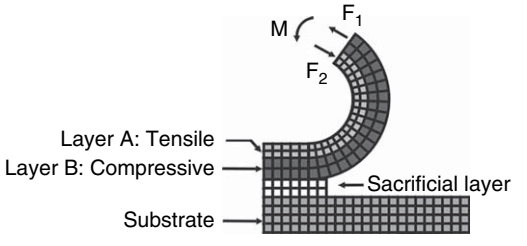


Figure 22 Schematic diagram of the strain-induced bending of an InAs/GaAs bilayer after releasing it from substrate. (Source: Reprinted from *Surf. Sci.*, **532–535**, Prinz A V, Prinz V Y, Application of semiconductor micro- and nanotubes in biology, 911–15, Copyright (2003), with permission from Elsevier.)

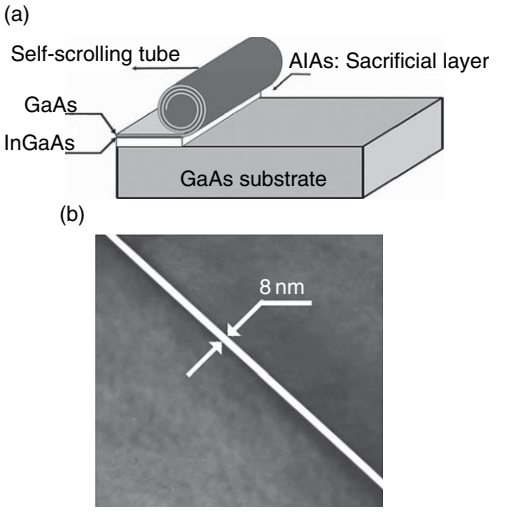


Figure 23 Self-rolling nanotube: (a) schematic diagram of a single tube of an InGaAs/GaAs bilayer rolling with length L and (b) an SEM image of the tube. (Source: Reprinted from *Physica E*, **23**, Prinz Y V, Chekhovskiy A V, Preobrazhenskii V V, Semyagin B R, Gutakovsky A K, 2002 A technique for fabricating InGaAs/GaAs nanotobes of precisely controlled lengths. *Nanotechnology* **13**, 231–3, Copyright (2003), with permission from Elsevier.)

radius of curvature of the tube structure can be determined using the following equation:

$$R = \frac{d\{[3(1+m)^2] + [1+(mn)][m^2 + \frac{1}{mn}]\}}{6\varepsilon(1+m)^2} \qquad [1]$$

where n is the ratio of Young's modulus of the bilayers, m is the thickness ratio of the bilayers, and ε is the strain due to lattice mismatch. Recently, the rolled-up single-crystalline nanotube structures have been probed using X-ray microdiffraction by Krause *et al.* (2006).

1.02.4.1.1 3D micro/nanostructures

A new class of precise micro/nanostructures such as tubes, scrolls, ring, and helices have been fabricated by self-rolling of strained compound semiconductor bilayers (Prinz 2004, Prinz *et al.* 2001). By using this technique, ultrathin epitaxial layers, down to few monolayers, can be controllably detached from the substrates by selective etching of the underlying sacrificial layers and rolled up, under the action of internal stresses, into various cylindrical micro/nanostructures. The diameter of the tubes can be scaled by changing the overall layer thickness, the built-in strain, and the thickness ratio of the top and lower layers. Schmidt and coworkers have also reported various nanoobjects on InGaAs/GaAs and SiGe/Si material systems (Deneke and Schmidt 2004, Schmidt *et al.* 2001, 2002). They have measured the tube diameter as a function of bilayer thicknesses, which agrees with theory (**Figure 24**) (Deneke and Schmidt 2004). Furthermore, the directional rolling concept was introduced in order to direct the free-standing micro/nanotubes to a particular orientation or to fabricate arrays of structures (Vorob'ev and Prinz 2002). This technique enables novel nanostructuring architecture for nanotechnology to fabricate atomically precise 3D nanostructures.

Novel physical phenomena have also been explored from the strained bilayer nanostructures. The numerical simulations of the quantum confinement in nanocorrugated freestanding InAs layers and quantum properties of InAs/GaAs nanotubes have been reported by Osadchii and Prinz (2000, 2006), respectively. Kleiner (2003) has shown that any curved electron gas-containing layers placed in a magnetic field should generate spin-polarized currents. This effect opens up new possibilities in using spin currents in integrated solid-state spintronic devices. The magnetotransport properties of 2D electron gas in rolled-up QWs have been reported by Vorob'ev *et al.* (2004). A difference between the magnetoresistances of planar and rolled QWs has been observed, i.e., the electrical resistance of the active layer increases after the detachment of the layer from the substrate due to the depletion region formed in the active layer at its rear-free surface.

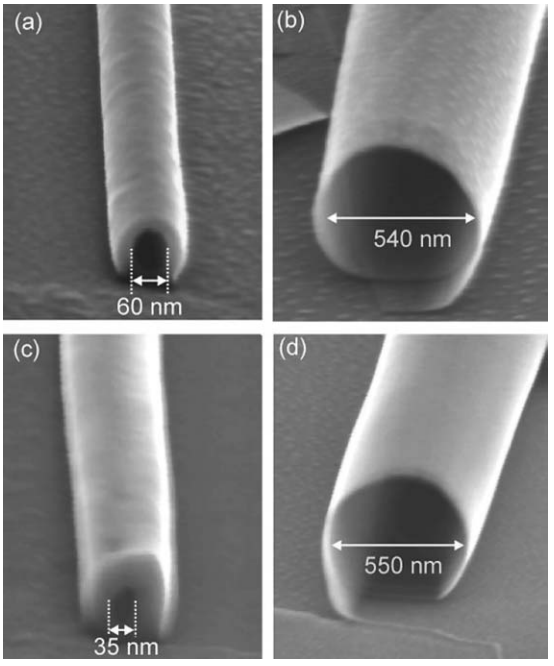


Figure 24 Four typical tube openings consisting of the following as-grown bilayers: (a) 1.4ML $In_{0.33}Ga_{0.67}As$/6.4ML GaAs, (b) 14.1ML $In_{0.33}Ga_{0.67}As$/19.1ML GaAs, (c) 1.4ML InAs/6.4 ML GaAs, (d) 1.4ML InAs/20.4 ML GaAs. (Source: Deneke *et al.* (2002), © 2003, reproduced with permission from IOP Publishing Limited.)

1.02.4.2 Micro-Origami Technique

In the strain-driven bilayers, a hinge concept was introduced and self-assembled 3-D microstructures have been fabricated by Vaccaro *et al.* (2001). The hinges connect released structures with the substrate. This technique is called micro-origami technique, which is in a way similar to the origami technique of paper folding. They applied this technique to fabricate freestanding 3D microstructures such as micromirrors, retroreflectors, and microstages on compound semiconductors. These 3D structures can be electrostatically actuated.

In addition to the strained bilayer, the hinge plays a role in determining the standing position of the released microstructures. The simple hinge structure is formed by a thin flexible layer, with no sliding parts that could get stuck or be degraded by friction.

Furthermore, top (scale)-down to submicron (nano)-order mechanical systems are feasible because the simple structure of the hinge does not impose limitations in the lithographic process. The main advantage of this technique is that 3D microstructures are self-assembled by the strain force built in the bilayers, whereas conventional techniques need manual assembling process. This enables monolithic integration of optoelectronic devices with micromechanical structures.

A schematic cross-sectional view of a micromirror produced by micro-origami technique is shown in **Figure 25**. This micromirror epilayer structure from the top consists of a compensation layer (InGaAs), a component layer (GaAs), an etch stop layer (AlGaAs), a strained bilayer (GaAs/InGaAs), and a sacrificial layer (AlGaAs/AlAs) above the substrate. In the micromirror structure, the component layer can be used to have passive optical elements such as DBR structures. The sacrificial layer thickness is kept constant whereas the bilayer and compensation thicknesses can be varied to position the micromirror at a desired angle. The SEM image of a retroreflector fabricated by micro-origami technique is shown in **Figure 26**; the mirrors are released from the substrate and stand vertically (Aida *et al.* 2002).

The hinged 3D microstructures have been fabricated by photolithography and the wet etching process. The first step is to fabricate hinges, the second step defines the shape of the microstructure (e.g., micromirrors) by vertical etching and then the micromirrors are released by etching (horizontal) the sacrificial layer (see **Figure 1**). Finally, the postetching process is followed as discussed in Section 1.02.2.2. When the sacrificial layers are etched away by selectively etching, the strain in the bilayer results

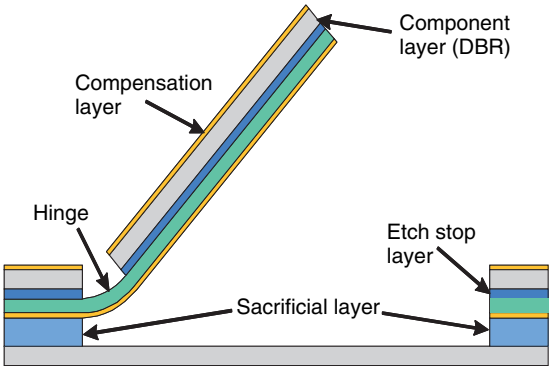


Figure 25 Schematic cross-sectional view of an self-assembled micromirror by micro-origami technique.

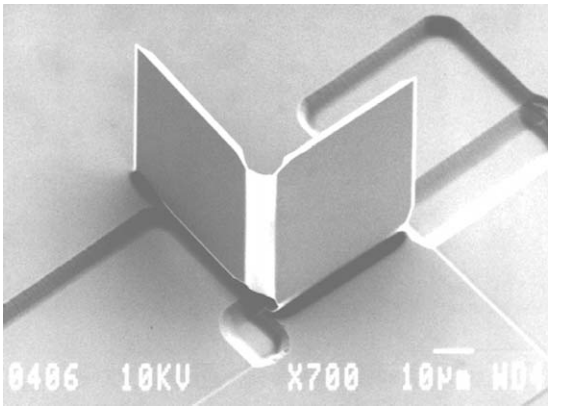


Figure 26 Self-assembled retroreflector micromirrors. (Source: Aida T, Vaccaro P O, Kubota K 2002 On the design and fabrication precision of micro-origami devices. In: IEEE/LEOS (eds.) *Optical MEMS 2002*, Lugano, Switzerland, pp. 43–4, © 2002, reproduced with permission from IEEE.)

in a deformation, which lifts the micromirror out of the substrate and is self-positioned. The micromirror standing position depends on the bilayer thickness, composition, and the width of the hinge. In order to maintain the micromirror flatness, a compensation layer is to be added above the strained bilayer. A flat or curved mirror surface, depending on the application, can be obtained by controlling the thickness of the compensation layer.

Kubota *et al.* (2003) proposed and implemented a complex self-assembled microstage in the InGaAs/GaAs materials system using two types of hinges that bend toward opposite directions, called valley fold and mountain fold. After releasing the multilayered structure from the substrate by the selective etching of a sacrificial layer, the microstage moves into its final position powered by the strain release in the InGaAs layers, where the stage is kept parallel to the substrate at a height of 50 µm owing to the combination of the two types of hinges (**Figure 27**). By taking advantage of strain-driven self-assembling mechanism, Dharmarasu *et al.* (2004) have designed micromirrors with a self-locking mechanism, which has side plates that lock the micromirror at the desired angle. The elevation angles of micromirrors and side plates are independently controlled by varying the hinge width for a given bilayer thickness. The geometry of these side plates assists in locking the micromirror precisely with the substrate at the designed value and keeps the micromirror flat. The undercut effect on the micromirror is suppressed and the stiction of the micromirrors with the substrate is eliminated by the self-locking mechanism.

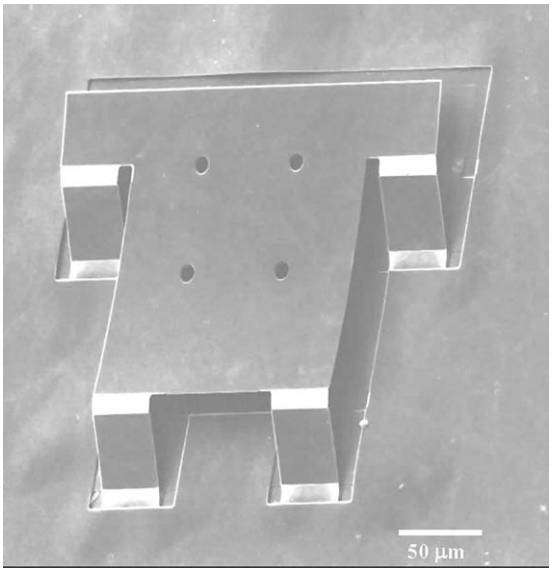


Figure 27 A microstage that elevated 50 μm above the substrate using valley-fold and mountain-fold hinges. (Source: Kubota K, Fleischmann T, Saravanan S, Vaccaro P O, Aida T 2003 Self-assembly of micro-stage using micro-origami technique. *Jpn. J. Appl. Phys*. **42**, 4079–83, © 2003, reproduced with permission from Japanese Journal of Applied Physics.)

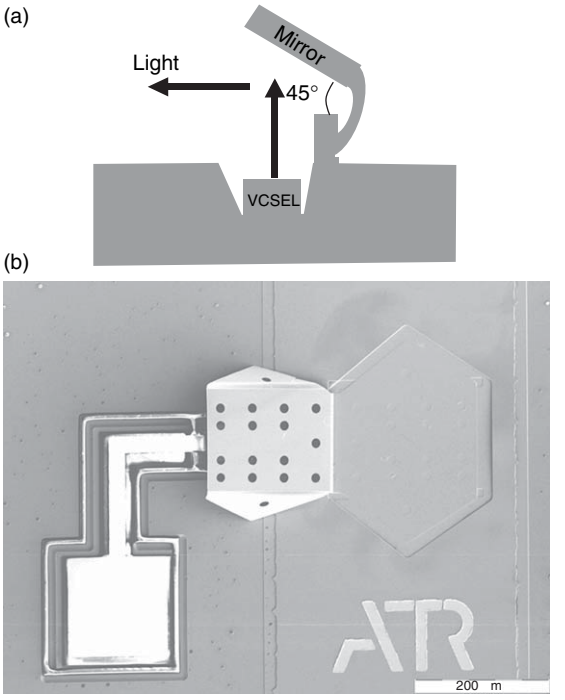


Figure 28 Monolithic integration of self-assembled micromirror with VCSEL: (a) a schematic illustration and (b) an SEM image of the integrated device.

1.02.4.3 Monolithically Integrated Microsystems

Monolithic integration of GaAs-based cantilever, laser diode, and photodiode on lateral geometry has been demonstrated by Ukita *et al.* (1993). Recently, GaAs-based self-assembled micromirrors with LEDs and VCSELs have been integrated monolithically on vertical geometry by Dharmarasu *et al.* (2005). The schematic cross section and an SEM image of the integrated device are shown in **Figure 28(a)** and **28(b)**, respectively. The device epilayer structure was grown in a single step. The micromirrors were self-assembled by the strain-driven mechanism at an angle of 135°. The micromirror is positioned to deflect the vertically emitted light parallel to the substrate and therefore enables the optical elements to be cascaded and integrated on the same substrate. This result shows the feasibility of monolithic integration of compound semiconductor-based micromachined structures (mirrors and lenses) with active optoelectronic devices (LEDs, VCSELs, and photodiodes), which enables light generation, transmission, and detection on a single chip for optical ICs.

1.02.5 Summary and Future Prospects

A comprehensive review of microsystems on compound semiconductors is presented. Material properties, micromachining aspects, various kinds of micromachined sensors, and microsystems have been discussed in detail. They include cantilevers, membranes, waveguide switches, RF inductors, microbolometers, micromotors, and wavelength-tunable microcavity devices such as filters, VCSELs, LEDs, and photodiodes. Wavelength-tunable microsystems, both monolithic and hybrid, approaches have been accomplished by various research groups. These systems have potential application in wavelength division multiplexing system for optical communication systems. Optical waveguide switches, which manipulate the optical signal directly rather than converting them into the electrical domain for manipulation, have also been demonstrated on compound semiconductors. The performance of these microsystems can be enhanced further by the introduction of photonic crystal structures (photonic band gap), which confine and control the light such as polarization and spectral control.

Furthermore, micromachined self-assembled micro/nanostructures have also been explored. This might introduce a new architecture and phenomena on compound semiconductor micro/nanosystems, enabling monolithic integration of optoelectronic devices.

When compared with silicon, many of the compound semiconductors have superior intrinsic properties such as optoelectronic properties, high piezoelectric and piezoresistive constants, and wide band gap enabling high-temperature and high-frequency applications. On the other hand, some of the compound semiconductors (group III–V) have lower mechanical strength; however, they are still fully satisfactory as reliable microsystems.

The cost has always been a concern with the compound semiconductors technology; however, the maturing technology, batch processing with high-volume manufacturing, and high yield will, hopefully, drive down the cost. There is already a need for non-classical CMOS development, especially on III–V compound semiconductors, for improving the capability of CMOS where compound semiconductor microsystems can be integrated as in silicon technology.

The III–V nitride and SiC materials with much potential, as discussed in this chapter, are attractive options in optoelectronic as well as sensing applications. However, much development is foreseen to improve the epilayers of these materials on cheap substrates as well as to improve the sacrificial etching processes. By the introduction and maturity of material deposition techniques such as MBE and MOCVD it is possible to achieve the goal.

Furthermore, fundamentally new architectures for the challenging future requirements can be achieved through optical technology where compound semiconductor microsystems can play a key role. The advancing microsystem technology on compound semiconductors has already begun to integrate optoelectromechanical devices into a single chip, which would play a key role in optical technology, and also to mimic one or more of our five human sensors.

References

Aida T, Vaccaro P O, Kubota K 2002 On the design and fabrication precision of micro-origami devices. In: IEEE/LEOS (eds.) *Optical MEMS 2002,* Lugano, Switzerland, pp. 43–4

Amano T, Koyama F, Hino T, Arai M, Matsutani A 2003 Design and fabrication of GaAs–GaAlAs micromachined tunable filter with thermal strain control. *J. Lightwave Technol.* **21**, 596–601

Atwell A R, Okojie R S, Kornegay K T, Roberson S L, Beliveau A 2003 Simulation, fabrication and testing of bulk micromachined 6H–SiC high-g piezoresistive accelerometers. *Sens. Actuators A Phys.* **104**, 11–18

Bakke T, Tigges C P, Lean J J, Sullivan C T, Spahn O B 2002 Planar microoptomechanical waveguide switches. *IEEE J. Select. Top. Quantum Electron.* **8**, 64–72

Beck R G, Eriksson M A, Westervelt R M, Campman K L, Gossard A C 1996 Strain-sensing cryogenic field-effect transistor for integrated strain detection in GaAs/AlGaAs micromechanical systems. *Appl. Phys. Lett.* **26**, 3763–5

Beck R G, Eriksson M A, Topinka M A, Westervelt R M, Maranowski K D, Gossard A C 1998 GaAs/AlGaAs self-sensing cantilevers for low-temperature scanning probe microscopy. *Appl. Phys. Lett.* **73**, 1149–51

Bykhovski A, Kaminski V, Shur M, Chen Q, Khan M 1996 Piezoresistive effect in wurtzite n-type GaN. *Appl. Phys. Lett.* **68**, 818–20

Chang-Hasnain J C 2000 Tunable VCSEL. *IEEE J. Select. Top. Quantum Electron.* **6**, 978–86

Chang-Hasnain J C 2001 Tunable VCSELs: Enabling wavelength-on-demand in metro networks. *Compd. Semicond. Mag*.

Christenson G L, Tran A T T D, Zhu Z H, Lo Y H, Hong M, Mannaerts J P, Bhat R 1997 Long-wavelength resonant vertical-cavity LED/photodetector with a 75-nm tuning range. *IEEE Photon. Technol. Lett.* **9**, 725–7

Cleland A N, Pophristic M, Ferguson I 2001 Single-crystal aluminum nitride nanomechanical resonators. *Appl. Phys. Lett.* **79**, 13–24

Cleland A N, Aldridge J S, Driscoll D C, Gossard A C 2002 Nanomechanical displacement sensing using a quantum point contact. *Appl. Phys. Lett.* **81**, 1699–701

Cole G D, Björlin E S, Chen Q, Chan C Y, Wu S, Wang C S, MacDonald N C, Bowers J E 2005 MEMS-tunable vertical-cavity SOAs. *IEEE J. Quantum Electron.* **41**, 391–407

Dantec R L, Benyattou T, Guillot G, Spisser A, Seassal C, Leclercq J L, Viktorovitch P, Rondi D, Blondeau R 1999 Tunable microcavity based on InP–air Bragg mirrors. *IEEE J. Select. Top. Quantum Electron.* **5**, 111–14

Datta M, Pruessner M W, Kelly D P, Ghodssi R 2004 Design of MEMS-tunable novel monolithic optical filters in InP with horizontal Bragg mirrors. *Solid-State Electron.* **48**, 1959–63

Davies S, Huang T S, Gass M H, Papworth A J, Joyce T B, Chalker P R 2004 Fabrication of GaN cantilevers on silicon substrates for microelectromechanical devices. *Appl. Phys. Lett.* **84**, 2566–8

Dehé A, Fricke K, Hartnagel H L 1995 Infrared thermopile sensor based on AlGaAs–GaAs micromachining. *Sens. Actuators A* **46–47**, 432–6

Dehé A, Krozer V, Chen B, Hartnagel H L 1996 High-sensitivity microwave power sensor for GaAs-MMIC implementation. *Electron. Lett.* **32**, 2149–50

Deneke C, Schmidt O G 2004 Lithographic positioning areal density increase and fluid transport in rolled-up nanotubes. *Physica E* **23**, 269–73

Deneke C, Müller C, Jin-Phillipp N Y, Schmidt O G 2002 Diameter scalability of rolled-up In(Ga)As/GaAs nanotubes. *Semicond. Sci. Technol.* **17**, 1278–81

Dharmarasu N, Kubota K, Saravanan S, Vaccaro P O, Saito N 2004 Monolithic integration of GaAs-based micro-origami mirrors with light emitting diodes. In: IEEE/LEOS (eds.) *Optical* MEMS 2004, Takamatsu, Japan, pp. 210–11

Dharmarasu N, Kubota K, Saravanan S, Vaccaro P O, Saito N 2005 Self-assembled GaAs micromirrors monolithically integrated with LEDs. *SPIE Proc.* **5650**, 135–40

Dong L, Yue R, Liu L 2003 An uncooled microbolometer infrared detector based on poly-SiGe thermistor. *Sens. Actuators A* **105**, 286–92

Eickhoff M, Möller H, Kroetz G, Berg J V, Ziermann R 1999 A high temperature pressure sensor prepared by selective deposition of cubic silicon carbide on SOI substrates. *Sens. Actuators* **74**, 56–9

Fricke K, Peiner E, Chahoud M, Schlachetzki A 1999 Fractures properties of InP microcantilevers by hetero-micromachining. *Sens. Actuators A* **76**, 395–402

Gaska R, Yang J, Bykhovski A, Shur M, Kaminskii V, Soloviov S 1997 Piezoresistive effect in GaN–AlN–GaN structures. *Appl. Phys. Lett.* **71**, 3817–19

Greek S, Gupta R, Hjort K 1999 Mechanical considerations in the design of a micromechanical tuneable InP-based WDM filter. *J. Microelectromech. Syst.* **8**, 328–34

Han L, Huang Q A, Liao X 2006 A novel micromachined microwave power sensor using GaAs MMIC process. *J. Phys.* **34**, 546–51

Harris J S Jr. 2000 Tunable long-wavelength vertical-cavity lasers: The engine of next generation optical networks. *IEEE J. Select. Top. Quantum Electron.* **6**, 1145–55

Hasse A, Irmer S, Daleiden J, Dharmarasu N, Hansmann S, Hillmer H 2006 Wide continuous tuning range of 221 nm by InP/air-gap vertical-cavity filters. *Electron. Lett.* **42**, 974–5

Hillmer H, Daleiden J, Prott C, Römer F, Irmer S, Rangelov V, Tarraf A, Schüler S, Strassner M 2002 Potential for micromachined actuation of ultra-wide continuously tunable optoelectronic devices. *Appl. Phys. B* **75**, 3–13

Hillmer H, Irmer S, Römer F, Hasse A, Strassner M, Daleiden J (2006) MOEMS vertical-air-cavity filters are widely and continuously tunable around 1.55 microns. *SPIE Newsroom Article*, 10.1117/2.1200601.0005

Hjort K 1996 Sacrificial etching of III–V compounds for micromechanical devices. *J. Micromech. Microeng.* **6**, 370–5

Hjort K, Ericson F, Schweitz J A 1994a Micromechanical fracture strength of semi-insulating GaAs. *Sens. Mater.* **6**, 359–67

Hjort K, Söderkvist J, Schweitz J Å 1994b Gallium arsenide as a mechanical material. *J. Micromech. Microeng.* **4**, 1–13

Huang H, Huang Y, Wang X, Wang Q, Ren X 2004 Long wavelength resonant cavity photodetector based on InP/air-gap Bragg reflectors. *IEEE Photon. Technol. Lett.* **16**, 245–7

Iborra E, Olivares J, Clement M, Vergara L, Sanz-Hervás A, Sangrador J 2004 Piezoelectric properties and residual stress of sputtered AlN thin films for MEMS applications. *Sens. Actuators A* **115**, 501–7

Irmer S, Daleiden J, Rangelov V, Prott C, Römer F, Strassner M, Tarraf A, Hillmer H 2003 Ultra low biased widely continuously tunable Fabry–Pérot filter. *IEEE Photon. Technol. Lett.* **15**, 434–6

Ishikawa K, Zhang J, Tuantranont A, Bright V M, Lee Y C 2003 An integrated micro-optical system for VCSEL-to-fiber active alignment. *Sens. Actuators A* **103**, 109–15

Jiang L, Cheunga R, Hedley J, Hassan M, Harris A J, Burdess J S, Mehregany M, Zorman C A 2006 SiC cantilever resonators with electrothermal actuation. *Sens. Actuators A* **128**, 376–86

Kahn J, Katz R H, Pister K 2000 Emerging challenges: mobile networking for smart dust. *J. Commun. Netw.* **2(3)**, 188–96

Keating A J, Silva K K M B D, Dell J M, Musca C A, Faraone L 2006 Optical characterization of Fabry–Pérot MEMS filters integrated on tunable short-wave IR detectors. *IEEE Photon. Technol. Lett.* **18**, 1079–81

Kelly D P, Pruessner M W, Amarnath K, Datta M, Kanakaraju S, Calhoun L C, Ghodssi R 2004 Monolithic suspended optical waveguides for InP MEMS. *IEEE Photon. Technol. Lett.* **16**, 1298–300

Kleiner A 2003 Chiral spin currents and quantum Hall effect in nanotubes. *Phys. Rev.* **B 67**, 155311.1–155311.5

Knobel R, Cleland A N 2002 Piezoelectric displacement sensing with a single electron transistor. *Appl. Phys. Lett.* **81**, 2258–60

Kögel B, Maute M, Halbritter H, Riemenschneider F, Böhm G, Amann M C, Meissner P 2006 Long-wavelength MEMS tunable vertical-cavity surface-emitting lasers with high side mode suppression. *J. Opt. A Pure Appl. Opt.* **8**, S370–6

Krause B, Mocuta C, Metzger T H, Deneke C, Schmidt O G 2006 Local structure of a rolled-up single crystal: An X-ray microdiffraction study of individual semiconductor nanotubes. *Phys. Rev. Lett.* **96**, 165502.1–165502.4

Kubby J A 2001 On behalf of the MOEMS manufacturing consortium, combining light with motion: Hybrid integration of light emitters and detectors with SOI-based micro-opto-electro-mechanical systems. *SPIE Proc.* **4293**, 32–45

Kubota K, Fleischmann T, Saravanan S, Vaccaro P O, Aida T 2003 Self-assembly of micro-stage using micro-origami technique. *Jpn. J. Appl. Phys.* **42**, 4079–83

Kusserow T, Dharmarasu N, Hillmer H, Nakamura T, Hayakawa T, Vengatesan B 2006 Tailored stress in InP/GaInAs layers for InP/air-gap DBR-filters with photonic crystals. In: IEEE/LEOS (eds.) *Optical MEMS 2006,* Big Sky, MT, USA, pp. 88–9

Lalinský T, Kuzmík J, Porges M, Haščík S, Mozolová Ž, Grno L 1995 Monolithic GaAs MESFET power sensor microsystem. *Electron. Lett.* **31**, 1914–15

Lalinský T, Haščík S, Mozolová Ž, Burian E, Držík M 1999 The improved performance of GaAs micromachined power sensor microsystem. *Sens. Actuators A* **76**, 241–6

Lalinský T, Burian E, Držík M, Haščík S, Mozolová Ž, Kuzmík J, Hatzopoulos Z 2000a Performance of GaAs micromachined microactuator. *Sens. Actuators A* **85**, 365–70

Lalinský T, Burian E, Držík M, Haščík S, Mozolová Ž, Kuzmík J 2000b Thermal actuation of a GaAs cantilever beam. *J. Micromech. Microeng.* **10**, 293–8

Lalinský T, Držík M, Chlpík J, Krnáč M, Haščík S, Mozolová Ž, Kostič I 2005 Thermo-mechanical characterization of micromachined GaAs-based thermal converter using contactless optical methods. *Sens. Actuators A* **123–124**, 99–105

Larson M C, Harris J S Jr. 1996 Wide and continuous wavelength tuning in a vertical-cavity surface emitting laser using a micromachined deformable-membrane mirror. *Appl. Phys. Lett.* **68**, 891–3

Larson M C, Pezeshki B, Harris J S Jr. 1995 Vertical coupled-cavity microinterferometer on GaAs with deformable-membrane top mirror. *IEEE Photon. Technol. Lett.* **7**, 382–4

Leclercq J L, Ribas R P, Karam J M, Viktorovitch P 1998 III–V micromachined devices for microsystems. *J. Microelectron.* **29**, 613–19

Leclercq J L, Garrigues M, Letartre X, Seassal C, Viktorovitch P 2000 InP-based MOEMS and related topics. *J. Micromech. Microeng.* **10**, 287–92

Li J, Zhang X M, Liu A Q, Lu C, Hao J Z 2005 Monolithically integrated photonic MEMS subsystem for optical network applications. *Opt. Commun.* **249**, 579–86

Li L, Kumar P, Kanakaraju S, DeVoe D L 2006 Piezoelectric AlGaAs bimorph microactuators. *J. Micromech. Microeng.* **16**, 1062–6

Lin L Y, Shen J L, Lee S S, Wu M C 1997 Surface-micromachined micro-*xyz* stages for free-space microoptical bench. *IEEE Photon. Technol. Lett.* **9**, 345–7

Lin C C, Fu J, Harris J S Jr. 2004 Widely tunable Al_2O_3/GaAs DBR filters with variable tuning characteristics. *IEEE J. Select. Top. Quantum Electron* **10**, 614–21

Liu A Q, Zhang X M, Li J, Lu C 2003 Single-/multi-mode tunable lasers using MEMS mirror and grating. *Sens. Actuators A* **108**, 49–54

Liu A Q, Zhang X M, Tang D Y, Lu C 2004 Tunable laser using micromachined grating with continuous wavelength tuning. *Appl. Phys. Lett.* **85**, 17–25

Mateus C F R, Chang C H, Chrostowski L, Yang S, Sun D, Pathak R, Chang-Hasnain C J 2002 Widely tunable torsional optical filter. *IEEE Photon. Technol. Lett.* **14**, 819–21

Maute M, Riemenschneider F, Böhm G, Halbritter H, Ortsiefer M, Shau R, Meissner P, Amann M C 2004b Micromechanically tunable long wavelength VCSEL with buried tunnel junction. *Electron. Lett.* **40**, 430–1

Maute M, Böhm G, Amann M C, Riemenschneider F, Halbritter H, Meissner P 2004a Micro-mechanically and widely tunable long-wavelength VCSELs. *IEEE, 19th International Semiconductor Laser Conference,* 119–20

Maute M, Kögel B, Böhm G R, Meissner P, Amann M C 2006 MEMS-tunable 1.55-mm VCSEL with extended tuning range incorporating a buried tunnel junction. *IEEE Photon. Technol. Lett.* **18**, 688–90

Mehregany M, Zorman C A 1999 SiC MEMS: Opportunities and challenges for applications in harsh environments. *Thin Solid Films* **355–356**, 518–24

Mehregany M, Zorman C A 2004 Silicon carbide micro-and nanoelectromechanical systems. *SPIE Proc.* **5342**, 9–15

Miao J, Hartnagel L H 2004 High-energy ion implantation: An alternative technology for micromachining three-dimensional GaAs structures. *Sens. Actuators A* **114**, 505–9

Neculoiu D, Muller A, Konstantinidis G 2006 Electromagnetic modeling of GaAs membrane supported mm-wave receivers. *J. Phys. Conf. Ser.* **34**, 28–33

Olivares J, Iborra E, Clement M, Vergara L, Sangrador J, Sanz-Hervás A 2005 Piezoelectric actuation of microbridges using AlN. *Sens. Actuators A* **123–124**, 590–5

Ongkodjojo A, Tay F E H, Akkipeddi R 2005 Micromachined III–V multimorph actuators for MOEMS applications – Concept, design and model. *J. Microelectromech. Syst.* **14**, 610–18

Osadchii V M, Prinz V Y 2000 Charge-carrier separation in rolled heterostructures. *JETP Lett.* **72**, 312–15

Osadchii V M, Prinz V Y 2005 Quantum confinement in nanocorrugated semiconductor films. *Phys. Rev. B* **72**, 033313.1–033313.4

Pantazis A, Neculoiu D, Hatzopoulos Z, Vasilache D, Lagadas M, Dragoman M, Buiculescu C, Petrini I, Müller A A, Konstantinidis G, Müller A 2005 Millimeter-wave passive circuit elements based on GaAs micromachining. *J. Micromech. Microeng.* **15**, S53–9

Peerlings J, Dehé A, Vogt A, Tilsch M, Hebeler C, Langenhan F, Meissner P, Hartnagel H L 1997 Long resonator micromachined tunable GaAs–AlAs Fabry–Perot filter. *IEEE Photon. Technol. Lett.* **9**, 1235–7

Peerlings J, Riemenschneider R, Kumar N V, Strassner M, Pfeiffer J, Scheuer V, Daleiden J, Mutamba K, Herbst S, Hartnagel H L, Meissner P 1999 Two-chip InGaAs–InP Fabry–Pérot p–i–n receiver for WDM systems. *IEEE Photon. Technol. Lett.* **11**, 260–2

Prinz Y V 2004 Precise semiconductor nanotube and nanoshells fabricated on (110) and (111) Si and GaAs. *Physica E* **23**, 260–8

Prinz A V, Prinz V Y 2003 Application of semiconductor micro- and nanotubes in biology. *Surf. Sci.* **532–535**, 911–15

Prinz Y V, Seleznev V A, Samoylov V A, Gutakovsky A K 1996 Nanoscale engineering using controllable formation of ultra-thin cracks in heterostructures. *Microelectron. Eng.* **30**, 439–42

Prinz Y V, Grützmacher D, Beyer A, David C, Ketterer B, Deckardt E 2001 A new technique for fabricating three-dimensional micro- and nanostructures of various shapes. *Nanotechnology* **12**, S1–4

Prinz V Y, Chekhovskiy A V, Preobrazhenskii V V, Semyagin B R, Gutakovsky A K 2002 A technique for fabricating InGaAs/GaAs nanotubes of precisely controlled lengths. *Nanotechnology* **13**, 231–3

Pruessner M, King T T, Kelly D P, Grover R, Calhoun L C, Ghodssi R 2003 Mechanical property measurement of InP-based MEMS for optical communications. *Sens. Actuators A* **105**, 190–200

Pruessner M W, Amarnath K, Datta M, Kelly D, Kanakaraju S, Ho P T, Ghodssi R 2005 InP-based optical waveguide MEMS switches with evanescent coupling mechanism. *J. Microelectromech. Syst.* **14**, 1070–81

Pruessner M W, Siwak N, Amarnath K, Kanakaraju S, Chuang W H, Ghodssi R 2006 End-coupled optical waveguide MEMS devices in the indium phosphide material system. *J. Micromech. Microeng.* **16**, 832–42

Rajan N, Mehregany M, Zorman C A, Stefanescu S, Kicher T P 1999 Fabrication and testing of micromachined silicon carbide and nickel fuel atomizers for gas turbine engines. *J. Microelectromech. Syst.* **8**, 251–7

Richards P L 1994 Bolometers for infrared and millimeter waves. *J. Appl. Phys.* **76**, 1–24

Riemenschneider F, Sagnes I, Böhm G, Halbritter H, Maute M, Symonds C, Amann M-C, Meissner P 2003 A new concept for tunable long wavelength VCSEL. *Opt. Commun.* **222**, 341–50

Riemenschneider F, Maute M, Halbritter H, Boehm G, Amann M C, Meissner P 2004 Continuously tunable long-wavelength MEMS–VCSEL with over 40-nm tuning range. *IEEE Photon. Technol. Lett.* **16**, 2212–14

Roy S, DeAnna R G, Christian A Zorman, Mehregany M 2002 Fabrication and characterization of polycrystalline SiC resonators. *IEEE Trans. Electron Devices* **49**, 2323–32

Sasaki M, Yamaguchi T, Song J H, Hane K, Hara M, Hori K 2003 Optical scanner on a three-dimensional microoptical bench. *J. Lightwave Technol.* **21**, 602–8

Schmidt O G, Eberl K 2001 Thin solid films roll up into nanotube. *Nature* **412**, 42

Schmidt O G, Deneke C, Manz Y M, Müller C 2002 Semiconductor tubes, rods and rings of nanometer and micrometer dimension. *Physica E* **13**, 969–73

Shor J S, Zhang X G, Osgood R M 1992 Laser assisted photoelectrochemical etching of n-type-SiC. *J. Electrochem. Soc.* **139**, 1213–16

Strassner M, Daleiden J, Chitica N, Keiper D, Stalnacke B, Greek D, Hjort K 2000 III–V semiconductor material for tunable Fabry–Perot filters for coarse and dense WDM systems. *Sens. Actuators A* **85**, 249–55

Strassner M, Esnault J C, Leroy L, Leclercq J L, Garrigues M, Sagnes I 2005 Fabrication of ultrathin and highly flexible InP-based membranes for microoptoelectro-mechanical systems at 1.55 μm. *IEEE Photon. Technol. Lett.* **17**, 804–6

Streubel K, Rapp S, André J, Chitica N 1996 1.26-μm Vertical cavity laser with two InP/air-gap reflectors. *Electron. Lett.* **32**, 1369–70

Strittmatter P R, Beach R A, McGill T C 2001 Fabrication of GaN suspended microstructures. *Appl. Phys. Lett.* **78**, 3226

Strittmatter R P, Beach R A, Brooke J, Preisler E J, Picus G S, McGill T C 2003a GaN Schottky diodes for piezoelectric strain sensing. *J. Appl. Phys.* **93**, 5675–81

Strittmatter R P, Beach R A, Picus G S, McGill T C 2003b Piezoelectrically enhanced capacitive strain sensors using GaN metal–insulator–semiconductor diodes. *J. Appl. Phys.* **94**, 5958–63

Sun D, Fan W, Kner P, Boucart J, Kageyama T, Zhang D, Pathak R, Nabiev R F, Yuen W 2004 Long wavelength-tunable

VCSELs with optimized MEMS bridge tuning structure. *IEEE Photon. Technol. Lett.* **16**, 714–16

Tarraf A, Riemenschneider F, Strassner M, Daleiden J, Irmer S, Halbritter H, Hillmer H, Meissner P 2004 Continuously tunable 1.55 μm VCSEL implemented by precisely curved dielectric top DBR involving tailored stress. *IEEE Photon. Technol. Lett.* **16**, 720–2

Tayebati P, Wang P D, Vakhshoori D, Sacks R N 1998a Widely tunable Fabry–Pérot filter using Ga(Al)As–AlO deformable mirrors. *IEEE Photon. Technol. Lett.* **10**, 394–6

Tayebati P, Wang P, Vakhshoori D, Lu C, Azimi M, Sacks R 1998b Half-symmetric cavity tunable micro electromechanical VCSEL with single spatial mode. *IEEE Photon. Technol. Lett.* **10**, 1679–1681

Uenishi Y 1996 Coupled-cavity laser diode with micromachined external mirror. *The 1996 IEEE/LEOS Summer Topical Meeting*, Keystone, CO, USA, pp. 33–4

Ukita H, Uenishi Y, Tanaka H 1993 A photomicrodynamic system with a mechanical resonator monolithically integrated with laser diodes on gallium arsenide. *Science* **260**, 786–9

Vaccaro O P, Kubota K, Aida T 2001 Strain-driven self-positioning of micro machined structures. *Appl. Phys. Lett.* **78**, 2852–4

Vail E C, Wu M S, Li G S, Eng L, Chang-Hasnain C J 1995 GaAs micromachined widely tunable Fabry–Perot filters. *Electron. Lett.* **31**, 228–9

Vakhshoori D, Tayebati P, Lu C C, Azimi M, Wang P, Zhou J H, Canoglu E 1999 2 mW CW single-mode operation of a tunable 1550 nm vertical cavity surface emitting laser with 50 nm tuning range. *Electron. Lett.* **35**, 900–1

Vorob'ev B A, Prinz V Y 2002 Directional rolling of strained heterofilms. *Semicond. Sci. Technol.* **17**, 614–16

Vorob'ev A B, Prinz V Y, Yukecheva Y S, Toropov A I 2004 Magnetotransport properties of two-dimensional electron gas on cylindrical surface. *Physica E* **23**, 171–6

Wiser R F, Tabib-Azar M, Mehregany M, Zorman C A 2005 Polycrystalline silicon-carbide surface-micromachined vertical resonators-Part II: Electrical testing and material property extraction. *J. Microelectromech. Syst.* **14**, 579–89

Wu M S, Vail E C, Li G S, Yuen W, Chang-Hasnain C J 1996 Widely and continuously tunable micromachined resonant cavity detector with wavelength tracking. *IEEE Photon. Technol. Lett.* **8**, 98–100

Yamaguchi H, Kanisawa K, Miyashita S, Hirayama Y 2004 InAs/GaAs(111)A heteroepitaxial systems. *Physics E* **23**, 285–92

Yang Z, Wang R, Wang D, Zhang B, Chen K J, Lau K M 2005 GaN on patterned silicon (GPS) technique for GaN-based integrated microsensors. *IEEE, IEDM Technical Digest*, Washington, DC, USA, pp. 298–301

Yang Z, Wang R, Jia S, Wang D, Zhang B S, Lau K M, Chen K J 2006 Mechanical characterization of suspended GaN microstructures fabricated by GaN-on-patterned-silicon technique. *Appl. Phys. Lett.* **88**, 041913.1–041913.3

Yasseen A A, Wu C H, Zorman C A, Mehregany M 2000 Fabrication and testing of surface micromachined polycrystalline SiC micromotors. *IEEE Electron. Device Lett.* **21**, 164–6

Biographies

Dharmarasu Nethaji received his Ph.D. degree in Physics (compound semiconductors) from the Anna University, India, in 1999 and part of the Ph.D. research was carried out at Friedrich Schiller University, Germany. He was awarded DAAD, Germany, fellowship in 1998. He has been working in the field of compound semiconductors for the past 12 years. From 1999 to 2002 he worked at Toyota Technological Institute, Nagoya, Japan, on III–V compound semiconductor multijunction space solar cells. From 2002 to 2005 he worked at the Advanced Telecommunications Research Institute International, Kyoto, Japan, on III–V compound semiconductor-based optical and mechanical devices. He has integrated optical devices (LED and VCSEL) with micromirrors and also achieved high-density LED arrays (2400 dpi). Since 2006 he has been with the Institute of Nanostructure Technologies and Analytics (INA) at the University of Kassel, Germany, where he is heading the optoelectromechanical devices group. He has published more than 50 papers, holds 4 patents, and is a member of the Indian Association for Crystal Growth, India.

Hartmut Hillmer holds a diploma (1985) and a doctoral thesis (1989, summa cum laude) in Physics on optically detected electronic transport in Si and III/V QW structures, both in physics from Stuttgart University. From 1989 to 1998 he was with German Telekom Research Center, Darmstadt, and in 1991 with NTT Optoelectronics Labs, Atsugi-shi, Japan. His habilitation thesis (1996) was on chirped DFB gratings by bent waveguides, ultrafast lasers, and MBE growth of AlGaInAs strain-compensated QW laser structures (Darmstadt University). Since 1999 he has been a full professor in the electrical engineering faculty of Kassel University, Germany, and holds the chair of Technological Electronics. His current research interests are technological implementation, characterization and theoretical model calculations of optoelectronic III/V semiconductor devices: spectral tuning of semiconductor lasers and filters, microsystem technology, and nanotechnology. From 2000 to 2002 he was the dean of the electrical engineering faculty, starting a course in applied computer engineering, and is one of the founding members of the Center for Interdisciplinary Nanostructure Science and Technology (CINSaT). He coordinates the participation of Kassel University with the NanoNetwork Hessen. He was awarded the GINO Innovation award in 2005 and the European Grand Prix of Innovation in 2006. He has published more than 190 papers, holds 14 patents, and is a member of IEEE and the German Physical Society (DPG).

1.03 Metals and Alloys

Kazuki Takashima[1] **and Akira Ishida**[2], [1]Kumamoto University, Kumamoto, Japan, [2]National Institute for Materials Science, Ibaraki, Japan

Glossary

EMAR Electromagnetic Acoustic Resonance
FIB Focused Ion Beam
IEC International Electronics
ISO International Standard Organization
LIGA Acronym for the German words *Lithographie, Galvanoformung und Abforming*
MEMS Microelectromechanical Systems
RF Radio Frequency
RIE Reactive Ion Etching
RUS Resonance Ultrasound Spectroscopy
SMA Shape-Memory Alloy
SME Shape-Memory Effect

1.03.1 Introduction

Metals and alloys are promising materials to realize microelectromechanical systems (MEMS) with higher performances, since they have wide variety of mechanical properties (elastic modulus, strength, fracture toughness, and fatigue properties) and functional properties (electric and magnetic properties). This chapter is devoted to describing the mechanical properties of metallic thin films, which usually differ from those of counter bulk materials. Section 1.03.2 is focused on the elastic and tensile properties of thin films, while the fatigue and fracture properties are discussed in Section 1.03.3. The mechanical properties data of various materials including amorphous, metallic glasses, and intermetallics, measured by various methods, are presented and their dispersions are discussed in terms of film structure and length-scale effects. In addition, new strengthening methods for thin films are introduced and recent activity in Japan for standardization of micromechanical tests is reported.

In the last part, special attention is paid to shape-memory alloy (SMA) thin films. These intermetallic films have begun to attract considerable interest as powerful microactuators for MEMS. Their mechanical properties and shape-memory behavior are discussed in connection with their microstructures first and then some promising applications of thin-film SMA microactuators are introduced along with the description of their microfabrication techniques.

1.03.2 Mechanical Characterization

1.03.2.1 Introduction

The continued development of MEMS devices is dependent on continued innovation in materials development for application on the micrometer scale. To date, single and polycrystalline silicon have been used in MEMS devices because of their compatibility with micromachine processing. In order to develop high-performance MEMS, however, the application of materials other than silicon is required. Metals and alloys have wide variety of mechanical properties (elastic modulus, strength, fracture toughness, and fatigue properties) and functional properties (electric and magnetic properties). This indicates that the use of metallic materials is promising to realize MEMS with higher performances.

For designing MEMS, materials selection is important to enable durable MEMS, and it is necessary to know the mechanical properties of free-standing metallic thin films on the micrometer scale. There are several length-scale effects on the properties of such microsized elements. Length-scale effects include several factors and may be classified into two categories (Spearing 2000). One is intrinsic effect, and this includes cube–square scaling law and mechanism-dependent scaling effects. The mechanism-dependent scaling effects have been investigated extensively for metallic materials particularly based on the plastic strain gradient mechanism (Hutchinson 2000). The other is extrinsic effect, and this includes microstructure, surface roughness, and micro/nanosized defects, which are usually process-dependent properties, because the microstructure of thin films strongly depends upon their preparation method and defects that may be introduced during processing. Both these types of length-scale effects are considered to affect the properties of microsized elements, and this suggests that we are not able to use the data of counter bulk materials for designing MEMS.

In these materials properties, mechanical properties including elastic constants, yield stress, tensile strength, fracture strength, fatigue properties, and fracture toughness are essential for designing and developing reliable and durable MEMS. These mechanical properties are also important as database for simulation-based design methodology of MEMS.

In Section 1.03.2, mechanical properties of metals and alloys films, which have been and will be used in MEMS devices, are described. Particularly, elastic and tensile properties are emphasized in this section. The mechanical properties including fatigue and fracture of conventional metals and alloys for MEMS materials are also well reviewed in literatures (Allameh 2003, Connolley *et al.* 2005, Spearing 2000). Therefore, new metallic MEMS materials including intermetallics and amorphous and metallic glasses are mentioned in this section.

1.03.2.2 Characterization Method

1.03.2.2.1 Elastic Properties

Elastic properties of metallic thin films are one of the fundamental parameters for the design of MEMS including spring elements, which are used as mechanical sensing and actuating devices. Particularly, the elastic constants determine the

sensitivity of pressure sensor and accelerometer, and the required force during actuation of radio frequency (RF) switch and micromirrors. The measurement methods for elastic constants can be divided into two groups (Ogi *et al.* 2005). One is static or quasistatic method and the other is dynamic one. Static methods include microtensile test and nanoindentation, and dynamic methods include flexural vibration method, ultrasonic method such as electromagnetic acoustic resonance (EMAR), resonant ultrasound spectroscopy (RUS)/laser, and picosecond-laser ultrasound methods.

Microtensile testing method is a simple technique to obtain Young's modulus. Young's modulus can be determined from the slope of the linear elastic portion in the stress versus strain (or force versus elongation) relation, same as bulk materials. In microtensile method, microsized tensile specimens should be prepared from thin-film layers using suitable micromachining technique such as lithography or focused ion beam (FIB) machining. If the gripping of the specimen is not completed, bending and torsional stresses arise during loading, which may affect significantly the stress–strain relationship. In addition, dimensional errors and strain measurement error may also affect the measurement values. Care should be taken with this measurement.

In nanoindentation technique, it is not necessary to remove substrate and to prepare special-type specimen, and elastic modulus in nanometer region can be measured. This technique would therefore be the most suitable to measure elastic constants for in-depth (out-of-plane) direction. The measurement of elastic modulus for in-plane direction is, however, difficult to obtain by this technique. The accuracy of the measurement value depends on those of the indenter shape, and its calibration is essential.

In the flexural vibration method, a cantilever-type specimen is prepared from thin-film layer, and the specimen is vibrated by application of external force. Elastic constants are determined from the resonance frequency and resonance mode. In this measurement method, the gripping (fixed) condition of specimen should be paid attention because a complete fixed (rigid) boundary is assumed in the calculation and gripping condition may affect the measurement accuracy.

The principal advantage of ultrasonic measurement methods is the high accuracy and resolution. Particularly, in the RUS method, the accuracy of the resonance frequency is usually 10^{-6} to 10^{-7}, so the elastic constants can be measured with high resolution. In addition, elastic constants can be measured without removing the substrate in the RUS method, and this is suitable to measure elastic constants of thin films on substrate.

1.03.2.2.2 Tensile Properties

Uniaxial tensile properties are the most important since these properties are essential for designing actual MEMS. Many investigations, therefore, have been carried out for measuring tensile properties of metallic thin films (Connolley *et al.* 2005). The specimen size is, however, usually very small, and there are several difficulties in measuring accurate mechanical properties. Recent progress of the tensile testing method of MEMS materials is described by Sharpe, Jr., *et al.* (2003).

1.03.2.3 Mechanical Properties

1.03.2.3.1 LIGA Ni

The typical pure metal for MEMS application is nickel prepared by LIGA (an acronym for the German words *Lithographie, Galvanoformung und Abforming*) process (Hruby 2001). The LIGA process consists of X-ray or ultraviolet (UV) lithography, electroplating, and molding to produce microscale components. X-ray lithography is used to produce a polymer mold (typically poly(methyl methacrylate) (PMMA)), and then the metallic components are formed by electroplating into the mold. This process can create high aspect ratio structure using a deep molding (more than 100 μm deep). The LIGA process is considered to be one of the most suitable techniques to fabricate metallic microcomponents, as this process uses an electroplating process. Pure Ni is the most common material for this process, but this process basically can be applied for any metals and alloys that can be deposited by electroplating.

LIGA Ni is expected to be applied to micromechanical components of MEMS including accelerometer, RF switch, and connector, and the mechanical properties of LIGA Ni films have therefore been investigated intensively. Young's modulus values for LIGA Ni have been measured (Cho *et al.* 2003, Hemker and Last 2001, Lee *et al.* 2005, Mazza *et al.* 1996, Sharpe and McAleavey 1998) by tensile testing. Mazza *et al.* (1996) obtained Young's modulus of LIGA Ni, which is consistent with that of bulk value (207 GPa, *Metals Handbook*, 10th ed., vol. 2, ASTM International). Other researchers, however, reported that the Young's modulus values were between 150 and 180 GPa (Cho *et al.* 2003, Hemker

and Last 2001, Lee *et al.* 2005, Sharpe and McAleavey 1998), and these values were significantly smaller than that of bulk Ni. These Young's modulus values were measured using microtensile specimens, the loading axis of which is parallel to in-plane direction of the film. In as-deposited LIGA Ni films, the grain structures tend to be columnar and tend to have a <001> crystallographic out-of-plane texture (Last *et al.* 2000). Hemker and Last (2001) modeled in-plane Young's modulus of an out-of-plane textured thin film and calculated the in-plane Young's modulus value to be 171–177 GPa. These values are consistent with measured values, and this suggests that the decrease in Young's modulus values of LIGA Ni film is due to the preferred crystallographic orientations. In contrast, Young's modulus values for out-of-plane direction were obtained by dynamic method (Majjad *et al.* 1999) and nanoindentation (Stauss *et al.* 2003). The measured Young's modulus values were 195 and 200 GPa. This indicates that LIGA Ni thin films have anisotropic mechanical properties.

In addition to elastic modulus, tensile properties of LIGA Ni were also investigated widely. Yield stress (usually 0.2% proof stress) obtained was between 300 and 400 MPa, and tensile strength was between 550 and 800 MPa for as-deposited condition (Allameh *et al.* 2003, Buchheit *et al.* 2003, Cho *et al.* 2003, Hemker and Last 2001, Lee *et al.* 2005, Mazza *et al.* 1996, Sharpe and McAleavey 1998). These values are higher compared to those of bulk Ni (0.2% yield stress of 59 MPa and tensile strength of 317 MPa, *Metals Handbook*, 10th ed., vol. 2, ASTM International). The grain diameter of as-deposited LIGA Ni film is very fine and of the order of several micrometers (Buchheit *et al.* 2003, Hemker and Last 2001), and the high strength of LIGA Ni films is due to the grain refinement. Actually, if the as-deposited Ni film is annealed at high temperature and the grain growth occurs, the tensile strength decreases (Hemker *et al.* 2001). The tensile properties of as-deposited condition seem to be very scattered. The strength of the film generally depends on the grain structure including texture and the defects involved in the film, and these depend on processing parameters during electroplating conditions for LIGA Ni. The above-mentioned data were obtained in the films with various grain diameter and film thickness. Further investigation on the relation among mechanical properties, microstructure, and processing conditions is required.

Other metals and alloy thin films including Ni–Fe and Ni–Co have been prepared by the LIGA process; the mechanical properties were measured (Buchheit *et al.* 2003, Cheng and Chen 2004, Stauss *et al.* 2003) and similar results have been obtained. The detailed mechanical properties of LIGA films were also reviewed by (Connolley *et al.* 2005). The mechanical properties measured in selected investigations are shown in **Table 1**.

1.03.2.3.2 Conventional Metallic Thin Films

Pure metals including Al, Au, Ag, Pt, Rh, and Ru thin films have been used as electrode, contact and

Table 1 Mechanical properties of metallic thin films

Material	*Young's modulus, E (GPa)*	*Yield stress, σ_B (MPa)*	*Tensile strength, σ_B (MPa)*	*Reference*
LIGA Ni	158–182	325–420	522–599	Mazza *et al.* (1996)
	202	405	782	Shrape and McAleavey (1998)
	180	360	535	Hemker and Last (2001)
	163	310–370	–	Cho *et al.* (2003)
	184	873	976	Lee *et al.* (2005)
	195	–	–	Majjad *et al.* (1999)
	200	–	–	Stauss *et al.* (2003)
Au	53–55	160–220	–	Espinosa and Prorok (2003)
	33.2–43.7	–	–	Baek *et al.* (2005)
	110.4–122.7	114.7–391.4	–	Son *et al.* (2003)
Ni-P amorphous	107	–	–	Ogi *et al.* (2002)
	115	–	2000	Takashima and Higo (2004)
Fe-Si-B amorphous	150	–	–	Shibata *et al.* (2005)
$Zr_{75}Cu_{19}Al_6$ BMG	83.4	–	1610 (bending)	Hata *et al.* (2001)
$Pd_{76}Cu_7Si_{17}$ BMG	57.9	–	1140 (bending)	Hata *et al.* (2001)

interlayer materials as well as mechanical components, and these metals are usually deposited on Si. Au thin film is one of the attractive materials for MEMS including RF switches because of its high conductivity that is useful for reducing the loss. Mechanical properties of Au films are therefore required to design mechanical components in the devices.

Young's modulus of Au film is a fundamental property for the devices and has been measured by several methods including microtensile and bending testing. Espinosa and Prorok (2003) and Espinosa *et al.* (2004) measured Young's modulus values of Au thin films of thickness 0.1–0.5 μm prepared by e-beam evaporation. The Young's modulus was determined by membrane deflection experiment (MDE). The resultant Young's modulus was 53–55 GPa, which is significantly lower than that of bulk Au (80 GPa, *Metals Handbook*, 10th ed., vol. 2, ASTM International). Baek *et al.* (2005) measured Young's modulus of Au films prepared by electroplating. The Young's modulus was determined by resonance frequency of cantilever beams, which was electrostatically excited. Young's modulus values obtained were 35.2–43.9 GPa and were also smaller than that of bulk Au. In these films, a strong <111> texture in the thickness direction was observed, and this may be one reason that the Young's modulus values were smaller than that of the bulk material. In contrast, Son *et al.* (2003) obtained Young's modulus values of 110–130 GPa for cantilever specimens prepared from e-beam-evaporated Au thin films of thickness 0.56–1.26 μm. Nanoindentation tests were also performed on these films, and Young's modulus values obtained were 110.4–122.7 GPa. These Young's modulus values are higher compared with the bulk value. In these tests, Young's modulus in out-of-plane direction is considered to be measured, and the similar texture was also observed in these films. The difference of Young's modulus is therefore due to the anisotropy of Au thin films, because the Young's modulus of Au in <111> direction is 117 GPa (Courteny 1990).

Yield stress and tensile strength of Au films varied significantly depending on the grain diameter, the film thickness, and the texture (Emery and Povirk 2003a,b, Son *et al.* 2003). The reported yield stress and tensile strength were 210–340 MPa and 230–350 MPa, repectively, for as-deposited Au films, and these values are much higher compared with that of the bulk Au (130 MPa, *Metals Handbook*, 10th ed., vol. 2, ASTM International). The relation between the mechanical properties and the microstructures in Au films is rather complicated and further investigations are required. Some of the mechanical properties measured for Au thin films are given in **Table 1**.

Mechanical properties of other metallic thin films were also investigated; however, the reported values were scattered, because the fabrication methods and conditions affect the microstructures and defects involved, and this changes the mechanical properties of metallic thin films. In addition, size effects on the mechanical properties of thin films should also be considered. In metallic thin films, the deformation process is governed by the dislocation mechanism, and the motion of dislocation is often controlled by specimen size (Espinosa *et al.* 2006). However, these effects have not been completely analyzed yet. Mechanical properties of metallic thin films, foils, and wires are also summarized in the literature (Connolley *et al.* 2005).

1.03.2.3.3 Multilayer Thin Films

Multilayer metallic thin films are expected to reduce the residual stress and to improve the function. However, the mechanical properties of metallic multilayered thin films have not been intensively investigated yet. This may be due to the difficulty in measurement and analysis methods in these films.

1.03.2.3.4 Amorphous

Amorphous metals or metallic glasses have no crystal structures and no grain boundaries. This means that the mechanical properties are homogeneous and isotropic even if the material size decreases to the micrometer range. In addition, amorphous alloys usually exhibit high corrosion resistance compared to conventional metals and alloys. This property is an advantage for the application of bio-MEMS. These characteristics are favorable for microdevices, and amorphous metals or metallic glasses are considered to be one of the candidate materials for future MEMS. For practical applications of these materials, the evaluation of mechanical properties in the thin-film form is required.

Ogi *et al.* (2002) measured elastic properties of a Ni–P amorphous alloy thin film prepared by electroless plating on an Al–Mg alloy substrate and obtained a complete set of effective elastic stiffness coefficients by EMAR method. The resulting coefficients were those of a transverse isotropic material. C_{11} was 140 ± 14 GPa, whereas C_{33} was 85 ± 3.0 GPa (the film growth is along the x_3 direction), and strong anisotropy between the in-plane and normal

directions was observed; the in-plane Young's modulus is larger than the normal Young's modulus by 34%, for example. The anisotropic coefficients can be interpreted by considering a micromechanics model for local incomplete cohesion (thin ellipsoidal voids) aligned parallel to the film surface. Takashima and Higo (2004) measured Young's modulus of electroless-deposited Ni–P amorphous alloy thin film by microtensile testing and obtained Young's modulus of 115 GPa; this value is comparable to that obtained by Ogi *et al.* (2002) for the in-plane direction. Takashima *et al.* (2001) also measured fracture toughness values of Ni–P amorphous alloy thin film and obtained similar anisotropic crack growth resistance between in-plane and out-of-plane directions.

Shibata *et al.* (2005) measured the complete set of elastic moduli of an Fe–Si–B amorphous-alloy thin film using thickness resonance with EMAR and RUS/laser methods. From the thickness-resonance measurements, out-of-plane shear moduli (C_{44} and C_{55}) were found to be 73.2 and 71.0 GPa. This indicates that the amorphous-alloy thin film shows transverse isotropic elastic symmetry. RUS/laser measurements gave the remaining moduli: $C_{11} = 193.6$ GPa, $C_{33} = 234.3$ GPa, $C_{23} = 72.1$ GPa, $C_{12} = 79.8$ GPa, and $C_{44} = 72.1$ GPa. From the ratios of the elastic modulus ($C_{33}/C_{11} = 1.21$ and $C_{44}/C_{66} = 1.26$), the amorphous Fe–Si–B thin film shows elastic anisotropy between in-plane and the out-of-plane directions in the range of 20–30%.

Tensile properties of amorphous thin films have not been widely investigated yet. Takashima and Higo (2004) performed microtensile tests for an electroless-deposited Ni–P amorphous-alloy thin film. Stress–strain relation is linear until fracture and fracture strength was 2 GPa. Shear bands were observed near the fracture surface and vein pattern was observed on the fracture surface. The fracture strength and the deformation behavior were similar to those obtained for Ni–P amorphous alloys. Mechanical properties measured for amorphous thin films are given in **Table 1** together with those for metallic glass thin films described in Section 1.03.2.3.5.

1.03.2.3.5 Metallic Glasses

Metallic glass is a kind of amorphous, but metallic glasses show viscous flow in a certain temperature range called super-cooled liquid region. If the metallic glasses are formed in this temperature range, microcomponents are expected to be easily prepared using a micromold (Saotome *et al.* 2003).

Zr- and Pd-based metallic glasses are expected to be applied for MEMS, because these two systems are easy to prepare thin films by sputtering. Hata *et al.* (2001) measured Young's modulus of Pd-based ($Pd_{76}Cu_7Si_{17}$, atomic%) and Zr-based ($Zr_{75}Cu_{19}Al_6$, atomic%) metallic glass thin films, prepared by an RF magnetron sputtering, using a thermomechanical analyzer and a microbending testing. The measured Young's moduli of the Zr- and Pd-based metallic glasses are 83.4 and 57.9 GPa, respectively. The Young's modulus of Zr-based thin film is comparable to those obtained for Zr-base bulk ones, but the Young's modulus of Pd-based thin film is lower than those reported for bulk ones (Wang 2005). The tensile strength of the Zr-based film was 1.61 GPa and that of Pd-based film was 1.14 GPa, and these values are slightly smaller than those obtained for bulk metallic glasses (Wang *et al.* 2004). A conical spring-type linear actuator, which is capable of stepwise displacement vertical to the substrate, and an on-chip microvariable inductor have been developed using metallic glass thin films (Fukushige *et al.* 2003, Hata *et al.* 2003).

1.03.2.3.6 Intermetallics

MEMS devices for high-temperature use are expected to be developed for aeronautics and space applications. One class of materials currently being expected to be developed for such applications is intermetallics because of their high-temperature strength and creep resistance. Application of these materials on the micrometer-scale requires an understanding of the effect of a reduction in size on their mechanical properties. The investigations of intermetallics for MEMS materials are, however, very limited. Halford *et al.* (2005) performed fracture tests on microsized, lamellar-structured TiAl and revealed that the lamellar orientation at the crack tip affects their fracture behavior. For the microsized, lamellar-structured TiAl, fracture toughness values (K_Q) are 1.5–7 MPa $m^{1/2}$ (Takashima *et al.* 2005) and are significantly smaller than that obtained for bulk material (18 MPa $m^{1/2}$) (Halford 2003). This indicates that extrinsic toughening mechanisms (crack tip shielding effects in the crack wake region) are not activated effectively, as the crack length is only few micrometers for microcomponents in MEMS devices (Takashima *et al.* 2005). This should be considered for applying these kinds of materials to MEMS.

TiAl thin films are also applied to bimorph actuators. Qu *et al.* (2002) prepared TiAl thin films by DC

magnetron sputtering from Ti and Al targets. The Young's modulus of the film was 175 GPa and this is higher than those of conventional metallic films. They also fabricated TiAl/SiO_2/TiAl bimorph structures. Further investigations on these materials are expected for future MEMS.

1.03.2.4 Strengthening Method

In order to further improve the performance and the reliability of MEMS devices, it is desirable to strengthen materials. It is, however, rather difficult to apply conventional strengthening methods to MEMS materials, as the sizes of MEMS are of the order of micrometers and controlling the mechanical properties based upon microsized structural modifications is no longer applicable to such small-scale materials. Strengthening based on nanostructural control would be one of the most promising methods for microsized materials, thus making its development necessary for potential candidate materials for MEMS applications.

1.03.2.4.1 Amorphous Alloy Thin Film

Tarumi *et al.* (2003) found that nanocrystals are formed under irradiation by FIB at an angle of <70° to the surface of Ni–P amorphous-alloy thin film. In contrast, nanocrystals are not formed under FIB irradiation at an angle between 70° and 90° to the surface, although Ga ion implantation does occur in the amorphous layer. The formation of nanocrystals in the amorphous phase is expected to be effective for improving the mechanical properties of amorphous thin film, with ion implantation also being a technique for improving the mechanical properties of bulk materials. In addition, the beam diameter of FIB can be focused to <100 nm, and this technique can irradiate microcomponents in MEMS. This indicates that FIB is applicable for strengthening microsized components in MEMS.

Figure 1 shows a transmission electron micrograph (dark field image) of Ni–P amorphous thin film after FIB irradiation at an angle of 15° from its surface. Nanocrystals with a diameter of approximately 10 nm are clearly observed in the irradiated area, and their crystallographic orientation is aligned to one direction as shown in Figure 1. This type of nanostructural modification is expected to be a new strengthening method for MEMS components.

A selective strengthening method for microcomponents in MEMS by ion implantation through FIB irradiation has been developed. Ion implantation by

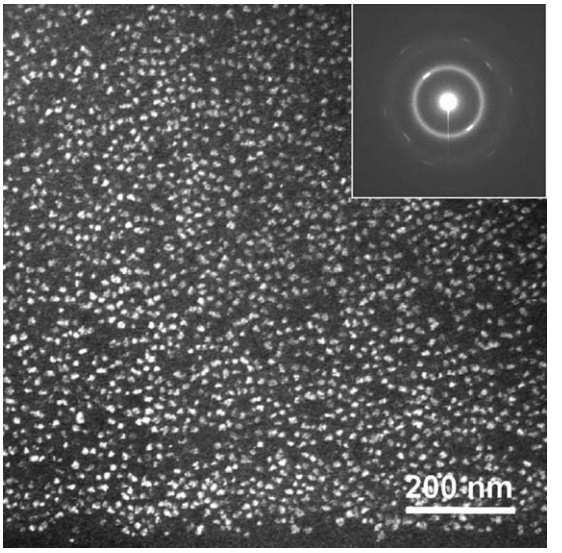


Figure 1 Selected area diffraction pattern (SADP) and dark-field electron micrograph (DFEM) taken from the focused ion beam (FIB)-irradiated plane of Ni–P amorphous alloy thin film. In the DFEM, uniform distribution of a large amount of nanocrystals is observed throughout the irradiated plane.

FIB was also performed on the surface of microcantilever specimens of Zr–Ni ($Zr_{65}Ni_{35}$) amorphous alloy (Kubota *et al.* 2005). In this irradiation, the Ga concentration increased to 8 at.% near the surface. Bending tests were carried out on specimens and compared (Figure 2). The bending strength of a

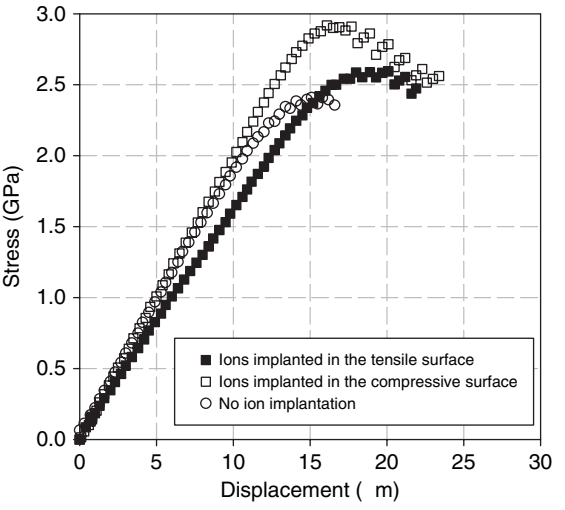


Figure 2 Load-displacement response of a microsized sample of Zr–Ni ($Zr_{65}Ni_{35}$) with Ga ions implanted in the compressive surface, with Ga ions implanted in the tensile surface, and without ion implantation.

Zr–Ni alloy has been shown to increase from 2.4 to 2.6 GPa by implantation on the cantilever's tensile surface or to 2.9 GPa by implanting the compressive surface. Ions implanted on the tensile surface and on the compressive surface are confirmed to increase the bending strength in comparison to samples with no ion implantation. These results suggest that ion implantation causes an increase in strength of the microsized material tested.

The small areas over which this implantation can be applied allow the capability to selectively strengthen regions of MEMS by implantation of ions on the surfaces of amorphous materials. Control over variables such as which surface is implanted, the dose applied, and the angle at which it is applied also allow further potential of developing this technique to locally customize properties. This can be achieved with only a minimal addition to the processing of batch-produced MEMS products.

1.03.2.4.2 Nanocomposite Thin Film

Deposited metallic thin films generally exhibit superior mechanical properties compared to those of counter bulk materials. This is considered to originate from the nanosized grain structures of the deposited thin films. These mechanical properties would be further enhanced by forming nanocomposite in the thin film layer. To date, several ceramic nanoparticle-reinforced metallic thin films have been prepared. Li and Li (2003) fabricated nanosized, Si_3N_4-reinforced NiFe nanocomposite layer of thickness up to 400 μm on a Cu substrate by electroplating technique. Permalloy (79% Ni, 21% Fe) was electroplated in the solutions with different percentages of nanosized amorphous Si_3N_4 particles (<30 nm). The microhardness of NiFe nanocomposite increased with increasing Si_3N_4 contents in as-deposited condition. Shi *et al.* (2006) prepared Ni–Co/SiC nanocomposites with various contents of SiC nanoparticles by electrodeposition. Ni–Co/SiC nanocomposite layer of thickness 20 μm were prepared on a Cu substrate by electrodeposition in a nickel–cobalt-plating bath containing SiC nanoparticles. The microhardness, wear, and corrosion resistance of the nanocomposite layer also increased with increasing content of the nano-SiC particles in as-deposited condition. Ceramic nanoparticle-reinforced nanocomposite films are expected to be applied to MEMS to improve their performances.

1.03.2.5 Standardization

Mechanical properties of structural thin films for MEMS are extremely important for designing and fabricating the devices, because the mechanical properties directly determine their performance, reliability, and durability. In developing MEMS, the fabrication of prototype device is relatively expensive and the development of simulation-based design tool is required. The database of mechanical properties on thin films that supported the simulation tool is needed to realize the simulation-based design. The mechanical properties of thin films are, however, dependent on their fabrication processes, as the microstructure of thin films is varied widely by processing. In addition, mechanical properties of these thin films have been measured by researcher's own methods, and it is rather difficult to compare the data obtained in each institute. This is due to the fact that there are no standard tensile testing methods including preparation of specimens, specimen size, and testing procedures for thin film materials.

In Japan, a round-robin test on tensile testing of thin films including metallic materials was performed and the cross-comparison of Young's modulus, tensile strength, and fracture strain was made (Tsuchiya *et al.* 2005). The materials used were single and polycrystalline Si, and Ti and Ni thin films deposited on Si wafers and microsized specimens were prepared from the thin film layer on the same wafer using a photolithography technique. The geometry of each test piece was such that it fits the testing machine at the institute, but the dimensions of the parallel parts of each test piece were standardized at 20 and 50 μm in width and 100 and 500 μm in gauge length, and these dimensions are approximately 1/1000 of the test pieces specified in ISO 6892 for tensile testing of metallic bulk samples. Methods for gripping the test piece, including electrostatic gripping (Tsuchiya *et al.* 1998), adhesives, mechanical gripping, and on-chip methods (Sato *et al.* 1998a), were examined by each institute. Mechanical properties including Young's modulus, tensile strength, and fracture strain were measured at each institute and compared. The results of round-robin tests revealed that there were no apparent differences between measuring methods, and the measured properties and deviations had almost the same values. The details of the round-robin test results are described in the literature (Tsuchiya *et al.* 2005).

Based on their results, the proposed draft of tensile testing of thin films was submitted to IEC, and the

international standard was specified as IEC 62047-2 (2006–08) *Semiconductor devices – Micro-electromechanical devices – Part 2: Tensile testing method of thin film materials*. In addition to tensile testing, standardization for axial fatigue testing of thin films was also submitted to IEC and accepted as a new proposal (IEC 62047-6 *Semiconductor devices – Micro-electromechanical devices – Part 2: Axial fatigue testing methods of thin film materials*).

Nanoindentation measurement procedure is also standardized in ISO standards (ISO 14577-1, 2, 3). These testing methods are specified for instrumented indentation, but this practice can be applied to nanoindentation. If these standardized testing methods are used to evaluate mechanical properties of thin films, the cross-comparison of the mechanical properties will be possible, and this is useful for developing MEMS. Standardization for other mechanical-related properties including elastic modulus, residual stress, fatigue, and fracture is desired.

1.03.2.6 Summary

Many research results show that metallic thin films deposited on substrate have anisotropic mechanical properties. This anisotropy originates from the grain structure of the thin films, which is often columnar, and the textures (preferred crystallographic orientations). The texture of thin films generally depends on the preparation process and its condition. Furthermore, some defects including microcracks are often formed during the processing. This anisotropic mechanical property is also found for amorphous alloy thin films that have no crystal form.

In addition to the anisotropy, there are several length-scale effects on the mechanical properties. These length-scale effects include intrinsic mechanism-dependent effects (attributed to the dislocation motion in metallic materials) and extrinsic effects such as surface roughness and dimension errors. These length-scale effects also affect the mechanical properties of metallic thin films.

These facts indicate that the mechanical properties of metallic thin films for MEMS applications should be measured using samples prepared by the same processing of the applied devices. The development of a database for mechanical properties including microstructures and processing is expected for the development of design technology in MEMS.

1.03.3 Shape-Memory Alloy Thin Films

1.03.3.1 Introduction

In the past decade, SMA thin films formed by sputtering have been attracting great interest as powerful actuators in MEMS such as microvalves, microfluid pumps, and micromanipulators (Bellouard 2002, Fu *et al.* 2004, Ishida and Martynov 2002, Krulevitch *et al.* 1996, Miyazaki and Ishida 1999) since the SMA thin films possess the largest available force and displacement as compared to the other type of actuators including electrostatic, electromagnetic, and piezoelectric ones (**Table 2**). The large surface-to-volume ratio in the thin films reduces the response speed to several tens of milliseconds, though bulk SMA alloys cannot be cycled at a rate of less than 1 Hz. Furthermore, they have compatibility with batch processing technology of silicon micromachining, so the actuators can be powered and controlled electrically by resistive heating of the films, utilizing the potential of integrated microelectronics.

Microdevices fabricated by the silicon micromachining process are demanded in a variety of fields such as medicine, biotechnology, semiconductor industry, and other industrial fields. The first microvalve using SMA thin films emerged in 1992 (Ray *et al.* 1992). Since then, intensive research works have been done to clarify the properties of SMA thin films and, simultaneously, many efforts to fabricate some other microdevices driven by SMA thin films have been made. This article reviews the properties of SMA thin films and introduces MEMS applications.

Table 2 Comparison of microactuators

Driving principle	*Range of motion*	*Force*	*Work output per volume ($J\,m^{-3}$)*	*Response speed*	*Driving voltage (V)*
Electrostatic	Large	0.04 N	1.8×10^5	3 kHz	5–100
Piezo	0.1%	40 MPa	1.2×10^5	10 kHz	20–30
Shape-memory effect (thin film)	5%	600 MPa	3.0×10^7	100 Hz	1–10

1.03.3.2 Mechanism of Shape-Memory Effect

Shape-memory effect (SME) refers to the ability of certain materials to remember an original shape even after rather severe deformation. Some special alloys such as TiNi, AuCd, and CuAlNi spontaneously return to their original, predeformed shape when heated, even if they are deformed beyond the elastic region at a low temperature (typically, up to about 5% for a Ti–Ni film). An example is shown in **Figure 3**. The film was first bent and then put into the hot water. **Figure 3(b)** shows that the film reverted to the original, straight shape when heated in the hot water. At this time, the film produces a large force. Using this recovery force, the film can move or grasp a small object.

The SME originates from the change of the crystal structure associated with the martensitic transformation. The martensitic transformation is defined as a diffusionless transformation described by a shear-like mechanism, as shown in **Figure 4(a)** and **4(b)**. In **Figure 4**, the crystal structure of the parent phase (a high-temperature phase) is represented by a square (**Figure 4(a)**) and that of the martensite phase (a low-temperature phase) by a parallelogram (**Figure 4(b)**), for simplicity. As the crystal structure of the martensite has lower symmetry than that of the parent phase, multiple formation of martensites with the same structure but in different orientations is possible, as shown in **Figure 4(b)**. Upon cooling, they form at the same amount to minimize the total elastic energy in the material and, as a result, macroscopic deformation cannot be detected in the stress-free state. The deformation associated with the change of the crystal structure, however, appears under a stress, since the deformation of the martensite proceeds in the way of the conversion from one martensite to another martensite (**Figure 4(c)**). The lattice distortion due to the change of the crystal structure generally amounts to about 20 times more than the elastic deformation and thus a large deformation is obtained. Even after unloading, the shape still remains deformed, but it reverts to the original shape as soon as the crystal structure returns to that of the parent phase by heating (**Figure 4(a)**). The important information obtained from **Figure 4** is that an external stress is necessary to utilize the transformation strain. As the martensite is very easy to deform when compared with the parent phase, an SMA film shows a reversible and repeatable shape change under a certain load (usually called a bias force). This bias force is a critical issue in designing a thin-film SMA actuator, as discussed later.

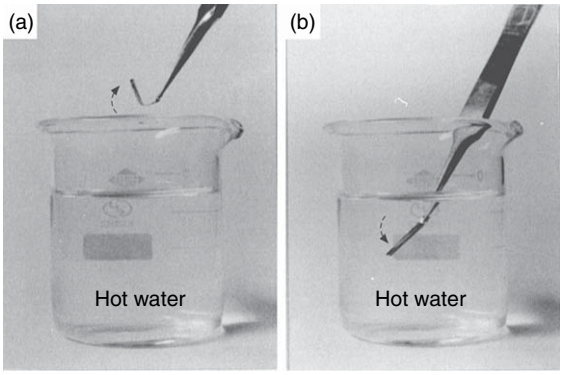


Figure 3 Shape-memory effect (SME) of sputter-deposited TiNi Film: (a) before and (b) after immersion in hot water.

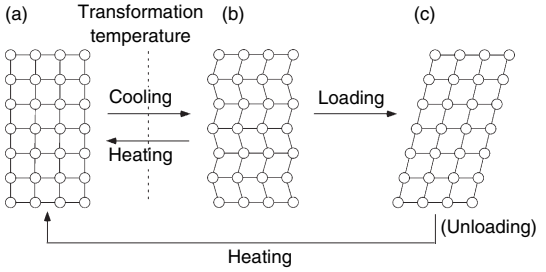


Figure 4 Schematic illustration of shape-memory effect (SME) due to change of crystal structure through transformation: (a) at high temperatures; (b) at room temperature; and (c) under a stress at room temperature.

1.03.3.3 Sputtering Process

SMA thin films of thicknesses ranging from 0.5 to 20 μm can be produced by sputtering. They are deposited on various substrates such as silicon wafers, polyimide films, and glass plates. **Figure 5** shows the schematic illustration of a sputtering apparatus. In the sputtering process, positively charged argon ions strike the target biased negatively and remove target atoms. These sputtered atoms condense on the substrate to form a thin film.

The early difficulties to overcome in fabricating thin films were porous structure, oxygen contamination, composition deviation, and so on. Sputtering parameters such as substrate temperature, sputter gas pressure, target-to-substrate distance, and target power play an important role in the structure and composition of deposited films. For example, if

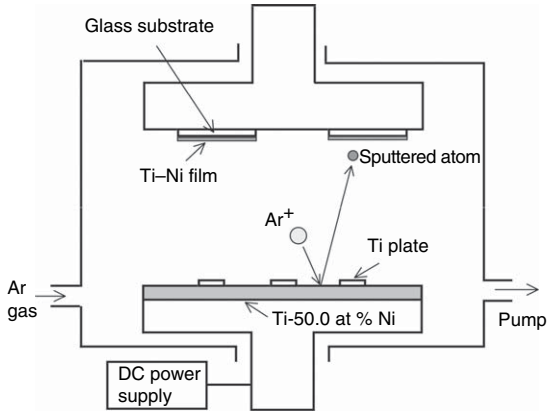


Figure 5 Schematic illustration of a sputtering apparatus.

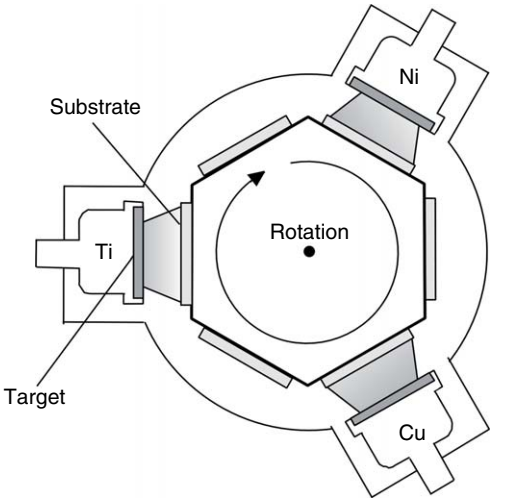


Figure 6 Schematic illustration of a carrousel-type magnetron sputtering apparatus (the view from above).

substrates are not heated, as-sputtered thin films are amorphous. These films require postsputtering heat treatment for crystallization. On the other hand, heated substrates (typically above 600 K) produce crystalline films with $<110>_{TiNi}$ fiber texture and fine-grained structure (Hou *et al.* 1995).

Sputtering gas pressure is another important parameter that influences the film structure. High gas pressure produces porous columnar structure, resulting in brittleness (Ishida *et al.* 1993). Oxygen contamination also causes brittleness and composition shift of deposited films by the formation of Ti-rich oxide. To avoid this contamination, a high-vacuum system (base pressure, 2×10^{-5} Pa) is necessary, but more important is presputtering of a Ti target with the shutter closed. Sufficient presputtering prior to sputtering can reduce the oxygen partial pressure effectively because of the getter effect of titanium.

The deviation of film composition from the target composition and its variation both in the plane and across the thickness are still troublesome problems, since a small shift of film composition causes a large change of the martensitic transformation temperature. The composition shift arises from the differences in sputtering yield, angular flux distribution, and lateral diffusion due to the collision with Ar gas and so on between sputtered species. In addition, the development of localized target wear and the variation in target composition with sputtering time make composition control difficult. After all, precise composition control is experientially achieved depending on the excellent reproducibility of sputtering. In most cases, the composition of Ti–Ni films shifts to the Ni-rich side because of higher sputtering yield of nickel. The compensation is done by placing titanium plates on top of the Ti–Ni target (**Figure 5**) or by using multiple targets. The former method involves a difficulty in reproducing film composition, though it is convenient. On the other hand, the latter method is suitable for precise control of film composition, since the supply of a constituent element from each target can be controlled independently. **Figure 6** shows a carrousel-type magnetron sputtering apparatus employed by the authors to produce Ti–Ni–Cu ternary alloy thin films (Ishida A *et al.* 2005). This apparatus has a rotating substrate holder at the center, which is surrounded by several targets. The power of each target can be controlled independently, allowing any composition.

1.03.3.4 Microstructure and Shape-Memory Behavior of Sputter-Deposited Thin Films

Amorphous films offer the opportunity to design the shape-memory behavior of the films by heat treatment. Therefore, most of the material science research works have been focused on the amorphous films.

For Ni-rich Ti–Ni thin films, conventional heat treatment, i.e., solution treatment followed by aging treatment, is available to stabilize the shape-memory behavior, in the same manner as for bulk materials. This heat treatment produces the same microstructure as that in bulk materials (i.e., distribution of fine Ti_3Ni_4 in the TiNi matrix) (Ishida *et al.* 1996).

In contrast, Ti-rich Ti–Ni thin films show peculiar fine microstructures, which have not been reported in bulk materials. **Figure 7** summarizes the microstructures of the films crystallized from the amorphous state (Kawamura *et al.* 2000). When an amorphous Ti-48.2 at.%Ni thin film is heat-treated at various temperatures, the microstructure in the TiNi matrix changes drastically in the sequence of (d) plate precipitates along {100} planes of the TiNi matrix, (c) plate precipitates and spherical Ti_2Ni precipitates, and (b) spherical Ti_2Ni precipitates, with increasing heat-treatment temperature (Ishida *et al.* 1997). The high-resolution electron microscopy revealed that the plate precipitates have coherency with the matrix, while the spherical Ti_2Ni precipitates are semicoherent with the matrix. Besides semicoherent Ti_2Ni precipitates, Ti_2Ni particles that do not show any specific orientation relationship form in films containing higher Ti content (**Figure 7(a)**).

The drastic change in microstructure with heat treatment and composition severely affect the shape-memory characteristics. **Figure 8** demonstrates the useful shape memory characteristics of the thin films containing the coherent plate precipitates (**Figure 7(d)**). In these experiments the strain of the film was measured during cooling and heating under a constant stress of 400 MPa. When cooled, the film shows elongation due to the martensitic transformation at a certain temperature. The strain associated with the martensitic transformation should completely recover due to the reverse martensitic transformation on heating. However, if plastic deformation is introduced during transformations, the residual strain remains after a thermal cycle, meaning that such a film is not suitable for an actuator. The Ti-50 at.%Ni thin film, which has no precipitates, shows a large transformation strain but a large plastic strain after a thermal cycle. On the other hand, the Ti-45.2 at.%Ni thin film shows no residual strain but a small transformation strain. The Ti-47.9 at.%Ni thin film annealed at 723 K for 1 h exhibits both a large transformation strain, which is comparable to

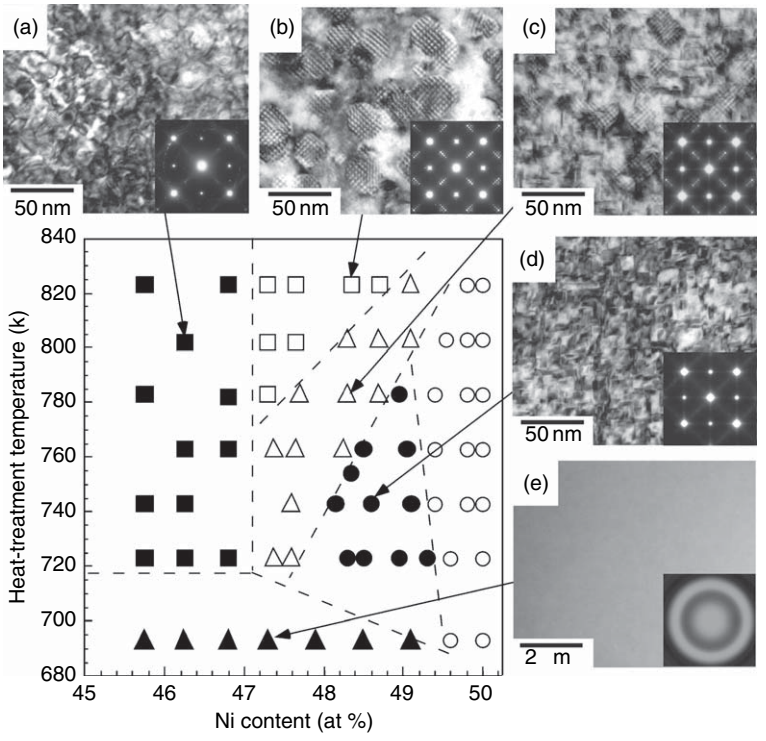


Figure 7 Various precipitates in the Ti–Ni matrix of Ti-rich Ti–Ni thin films heat-treated at various temperatures for 1 h: Solid squares (■) are Ti_2Ni particles with random orientation (a); open squares (□) are Ti_2Ni precipitates with the same orientation as that of the matrix (b); open triangles (△) are plate precipitates and oriented Ti_2Ni precipitates (c); solid circles (●) are plate precipitates (d); open circles (○) indicate no precipitates; and solid triangles (▲) are amorphous films (e). (Source: Kawamura Y, Gyobu A, Saburi T, Asai M 2000 Structure of sputter-deposited Ti-rich Ti–Ni alloy films. *Mater. Sci. Forum* **327–328**, 303–6.)

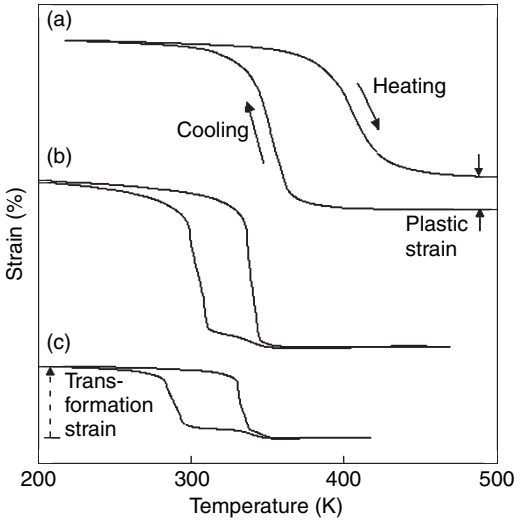


Figure 8 Effects of heat treatment and film composition on strain–temperature curves under a constant stress of 400 MPa for Ti-rich Ti–Ni thin films: (a) Ti-50 at.%Ni film annealed at 773 K for 1 h; (b) Ti-47.9 at.%Ni film annealed at 723 K for 1 h; and (c) Ti-45.2 at.%Ni film annealed at 773 K for 1 h.

that of the Ti-50 at.%Ni thin film, and no permanent strain after a cooling–heating cycle, which is comparable to the shape-memory behavior of the Ti-45.2 at.%Ni thin film. Annealing temperatures higher than 773 K, however, caused the appearance of plastic strain and the decrease of transformation strain. Obviously, the fine microstructures shown in **Figure 7(d)** are effective to improve the shape-memory behavior (Kajiwara *et al.* 1996).

Basically, Ti-rich Ti–Ni thin films exhibit higher transformation temperatures than Ni-rich thin films. In bulk alloys, however, the shape-memory behavior of Ti-rich Ti–Ni alloys has not been utilized and even not studied, since they are very brittle due to coarse grain boundary precipitates formed during casting. In contrast, the shape-memory behavior and mechanical properties of thin films are drastically improved by peculiar fine structures, as is obvious from **Figures 6** and **7**. Therefore, Ti-rich Ti–Ni thin films seem to be more attractive than Ni-rich Ti–Ni thin films from the viewpoint of practical use as well as fundamental aspects.

Ternary SMA thin films (Miyazaki *et al.* 1997) such as Ti–Ni–Cu, Ti–Ni–Pd, Ti–Ni–Hf, and Ti–Ni–Zr are also fabricated by sputtering. Substitution of Cu for Ni decreases temperature hysteresis, while substitution of Pd for Ni increases martensitic transformation temperatures with narrow temperature hysteresis, as in bulk materials. These characteristics are effective to achieve a high-response actuation. Recently, it was reported that a Ti–Ni–Zr thin film shows a perfect SME under a high stress at temperatures above 373 K, though bulk alloys with the same composition are so brittle that the shape-memory behavior has not been measured (Sawaguchi *et al.* 2004). The diversity of SMA thin films is one of the advantages of thin-film SMA actuators.

Besides ordinary SME, two-way SME (Kuribayashi 1990) has also been reported. Furthermore, a perfect superelasticity (Miyazaki *et al.* 1994) has been achieved in the form of thin film as well.

As discussed above, the shape-memory behavior of sputter-deposited thin films is different from that of conventional bulk alloys and is sensitive to heat treatment, composition, and bias stress. This makes it difficult for mechanical engineers and researchers to use SMA thin films. The information on shape-memory behavior of sputter-deposited thin films is, however, available in several material science journals (Ishida and Miyazaki 1999, Ishida *et al.* 1996, 2001, Sato *et al.* 1998b) and establishment of the database is in progress.

1.03.3.5 Mechanical Properties of Sputter-Deposited Thin Films

The measurement of stress–strain curves has been performed for shape-memory thin films with equiatomic, Ni-rich, and Ti-rich compositions (**Figure 9**) (Ishida *et al.* 2000). A Ti-50.0 at.%Ni thin film shows an elongation of more than 50%, while a Ti-51.5 at.%Ni thin film containing a large amount of

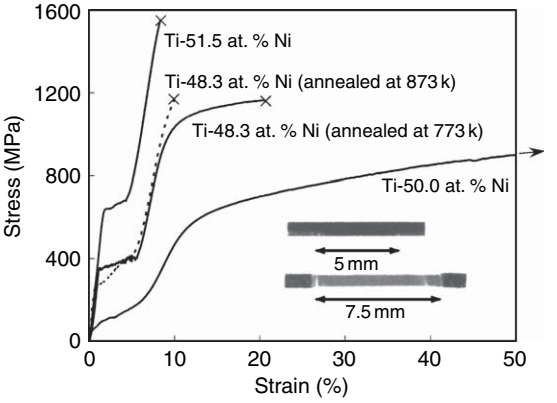


Figure 9 Stress–strain curves of Ti-48.3, 50.0, and 51.1 at.%Ni films. (X means fracture.) The inset shows the Ti-50 at.%Ni films before and after deformation.

Ti_3Ni_4 shows a fracture stress of 1.6 GPa. These values are comparable to those in bulk materials. Furthermore, a Ti-48.3 at.%Ni thin film annealed at a low temperature (773 K) shows an elongation of almost 20%. It is known that a bulk material with the same composition does not show any plastic deformation. This improvement of plasticity is attributed to suppression of grain boundary precipitation, since a Ti-48.3 at.%Ni thin film annealed at a high temperature (873 K) shows grain boundary precipitation, resulting in no plastic deformation. In general, the yield stress for plastic deformation of thin films seems higher than that of bulk materials, owing to the small grain size (<4 μm). The grain size of bulk materials is typically several tens of micrometers.

It has been found that the advantages of crystallized thin films with respect to structure are (1) small grain size, (2) formation of unique metastable precipitates, and (3) suppression of coarse grain boundary precipitates, which bring out stable SME and reliable mechanical properties compared with the conventional shape-memory bulk alloys.

1.03.3.6 Thin-film SMA Microactuators

Ordinary SME is one-way, in that a deformed shape in the martensite phase recovers the original shape in the parent phase. To achieve reversible actuation, a bias force is required to deform the film in the martensite phase, as discussed in Section 1.03.3.2. Depending on the source of a bias force, thin-film SMA microactuators can be classified into four types: (1) bimorph, (2) freestanding, (3) two-way SME, and (4) differential types, as shown in **Figure 10**. For bimorph actuators, the bias force is provided by a thermal stress exerted from the substrate, while for freestanding actuators, there is an external source of stress such as bias pressure and a spring. The bimorph actuators can produce large deflections at low forces while the freestanding ones are capable of exerting large forces over a substantial stroke. In general, the bimorph actuators are used for microgrippers (**Figure 13**), micromirrors and so on, whereas the freestanding actuators are used for microvalves (**Figure 14**) and micropumps (**Figure 15**).

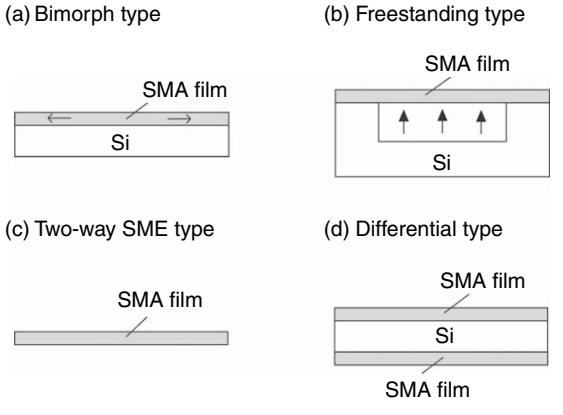


Figure 10 Various types of thin-film shape-memory alloy (SMA) microactuators.

Two-way SME utilizes the two-way SME. Two way means spontaneous shape change on both cooling and heating, thus displaying a repeatable shape change without a bias force. This effect is, however, obtained by aging under constraint, which is difficult to incorporate into the MEMS fabrication process. Further, control of the shape changes and recovery forces on both cooling and heating is difficult. For these reasons, this type of actuators is rarely used.

In differential-type actuators (Benard *et al.* 1998), the two SMA actuators work antagonistically, i.e., if one actuator is heated to shrink, the other actuator at room temperature extends. The advantages of these actuators are a large force and a large efficiency of energy conversion, but the problem is thermal isolation between the two actuators.

1.03.3.7 Technical Issues in Thin-film SMA Microactuators

The application of SMA thin films as microactuators produces additional challenging problems. The understanding of stress evolution and relaxation in thin films during annealing and micromachining is inevitable to prevent detachment of films and to control the actuation of SMA thin films especially for bimorph-type actuators (Grummon and Zhang 2001). The stress in SMA thin films consists of intrinsic and extrinsic stresses; the former is caused by deposition and crystallization, while the latter comes from differential thermal expansion. In addition to the stress in films, characteristics of the film/surface interface, such as the presence of reaction products, affect the film adhesion.

Film thickness is an important parameter in designing a thin-film SMA actuator. **Figure 11** shows the effect of grain size-to-film thickness ratio on the shape-memory behavior and mechanical properties of thin films. It is apparent that both of them are affected by the surfaces if the film thickness is less than 4 μm, which corresponds to a mean grain size of this film (Ishida and Sato 2003).

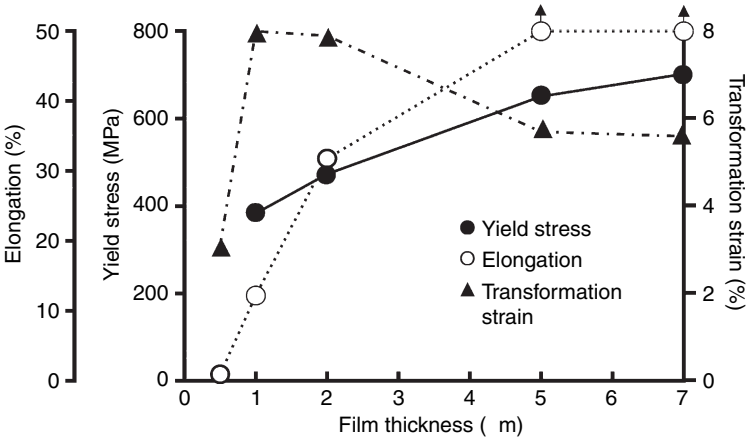


Figure 11 Thickness effects on shape-memory behavior and mechanical properties of Ti-50 at.%Ni film.

1.03.3.8 Microfabrication of SMA Thin Films

The possibility of integrating TiNi SMA thin films into silicon micromachining process was first demonstrated in 1990 (Walker *et al.* 1990). They deposited a Ti–Ni thin film onto a polyimide layer, wet etched the Ti–Ni thin film to make a serpentine shape, and released it by reactive ion etching (RIE) of the sacrificial polyimide layer. **Figure 12** shows the fabrication process of the microvalve shown in **Figure 14(b)** (Ishida and Martynov 2002):

(1) The backside of a Si wafer is anisotropically wet-etched to produce a thin membrane of silicon (~40 μm) for supporting an island of silicon, which eventually becomes the poppet of the valve.
(2) A layer of TiNi is sputter-deposited on the Si membrane and subsequently annealed for crystallization.
(3) A layer of gold is sputter-deposited on top of the TiNi layer, for facilitating masking in TiNi wet etching and for providing electrical contact to the conductive path of the actuator.
(4) The Au and TiNi films are patterned with wet etchants.
(5) The gold on the active parts of the actuators is removed.
(6) The remaining silicon is etched out to release the poppet and active parts of the SMA actuator.
(7) Finally, a microvalve is assembled (see **Figure 14(b)**).

The main issues to be addressed to integrate the sputter-deposited TiNi film into a silicon micromachining process are the following:

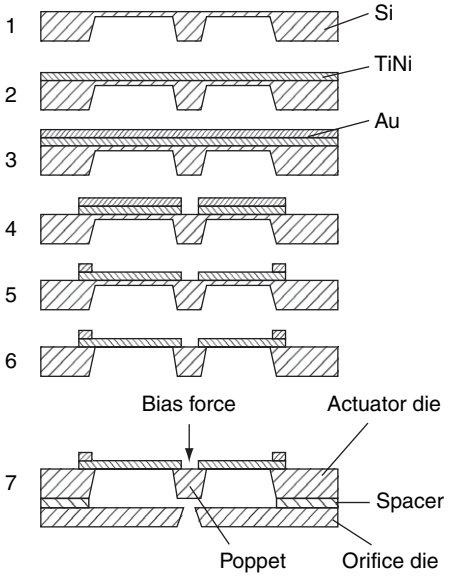


Figure 12 Micromachining steps of the microvalve are shown in (b): (1) backside Si etching; (2) deposition of TiNi film followed by crystallization annealing; (3) deposition of gold; (4) patterning and etching of gold and TiNi films; (5) removal of gold on the active parts; (6) etching of Si to release TiNi films; and (7) assembling. (Source: Ishida A, Martynov V V 2002 Sputter-deposited shape-memory alloy thin films: Properties and applications. *MRS Bull.* **27**, 111–14; reproduced with permission from MRS Bulletin.)

(1) Compatibility with the high annealing temperature (above 773 K) of SMA films
(2) Patterning methods for SMA films
(3) Selective release of SMA films from the substrate
(4) Compatibility with bonding process
(5) Internal stress and adhesion of SMA films

As described earlier, an amorphous film of TiNi is crystallized above 773 K, which causes no problem with a silicon wafer, but may not be compatible with materials that become unstable at high temperatures. Deposition of crystalline films onto heated substrates may avoid this problem. Hou *et al.* (1995) demonstrated the feasibility of producing crystalline SMA films on polymeric substrates by heating the substrate.

For etching of Ti–Ni thin films, $HF/HNO_3/H_2O$ solutions are usually used. A solution of $HF:HNO_3:H_2O=1:1:4$ allows selective etching of a Ti–Ni film deposited on a Si substrate (Makino *et al.* 2000). The problem of this etchant is undercut and backetching, since this process is isotropic etching (Gill *et al.* 2001). When compared with this chemical etching, electrolytic photoetching shows some advantages such as a high etch rate, good surface finish, and near vertical sidewall. As an electrolyte, 5% sulfuric acid in methanol (Kohl *et al.* 1996) and lithium chloride in ethanol (Mineta and Makino 2005) have been reported. RIE (3 sccm of $SiCl_4$ gas, 5 mTorr, 250 W, 20 °C, 90 min) is applicable to the etching of Ti–Ni (Nakamura *et al.* 1997). Ion milling (Gill *et al.* 2001), liftoff (Nakamura *et al.* 1996), and laser machining (Davies *et al.* 2002, Kohl *et al.* 1995) have been also reported for etching of Ti–Ni thin films. These methods can produce a vertical sidewall and avoid back-etching unlike wet etching.

On the other hand, an SMA film should not be damaged during etching of silicon. Though potassium hydroxide (KOH) and ethylene diamine pyrochatecol (EDP) are the most common anisotropic etchant of silicon and are commonly used in fabricating devices with SMA thin films, they (especially KOH) are known to attack the TiNi thin film (Krulevitch *et al.* 1996). To avoid this problem, RIE or deep RIE (Fu *et al.* 2005, Gill *et al.* 2002) is employed. Makino *et al.* (2001) measured the etch rates of Si, TiNi thin film, as well as those of SiO_2, Cr and positive photoresist layers as mask material during RIE in SF_6 at 30 sccm, with 100 mTorr, and at an RF power of 100 W. The result showed that silicon was etched at about 480 nm min^{-1} whereas no etching was observed for TiNi thin films. As mask material, the Cr layer showed superior results, whereas the SiO_2 layer and the photoresist were etched at about 1 and 60 nm min^{-1}. This method allows only Si to be etched without etching of a TiNi thin film.

A bonding temperature is also selected carefully so that the SME is not damaged. A high-temperature process causes the coarsening of the structure of a SMA film, resulting in an increased plastic strain after a thermal cycle. Makino *et al.* (2000) carried out anodic bonding between a TiNi film and a Pyrex glass at 623–673 K for 1–3 h in vacuum and diffusion bonding between two TiNi films under a bonding pressure of 210 MPa at 473–673 K for 1–5 h in vacuum. They were reported to obtain sufficient bond strengths without damaging the TiNi thin films.

1.03.3.9 Application of SMA Thin Films

So far, a number of microdevices activated with TiNi thin-film microactuators have been reported: microvalves (Kohl *et al.* 1997, 1999), micropumps (Benard *et al.* 1998, Makino *et al.* 2001, Xu *et al.* 2001), microreley (Johnson 1999), micromirrors (Fu *et al.* 2005, Johnson and Martynov 1997), out-of-plane spacers (Johnson and Martynov 1997), microgrippers (Roch *et al.* 2003, Wang *et al.* 2002), and microclips (Mineta *et al.* 2005, Takeuchi 2000). **Figure 13** shows an SMA-actuated microgripper fabricated by Krulevitch *et al.* (1996). The size of the microgripper is 1000 μm × 200 μm × 380 μm. Thermal stress between the Ti–Ni–Cu film and the Si substrate provides a bias force. The recovery force of the SMA films deflects each side of the microgripper up to 55 μm for a total gripping motion of 110 μm when the SMA films are heated to 70°C. When fully open, the force exerted by the film corresponds to a 13 mN gripping force at the tip of the gripper. The maximum switching rate for fully opened to fully closed is approximately 100 Hz in air, with an input power of 30 mW in air and 150 mW in water.

The microvalve with a thin-film SMA actuator was first reported in 1992. **Figure 14** shows the

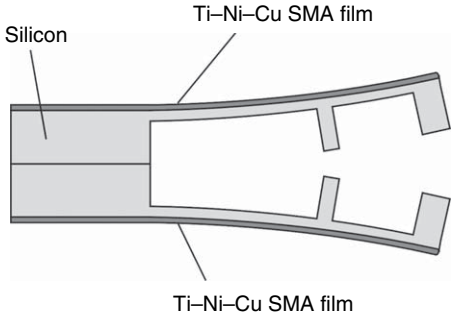


Figure 13 Microgripper driven by shape-memory alloy (SMA) film. (Source: Krulevitch P, Lee A P, Ramsey B R, Trevino J C, Hamilton J, Northrup M A 1996 Thin film shape memory alloy microactuators. *J. Microelectromech. Syst.* **5**, 270–82.)

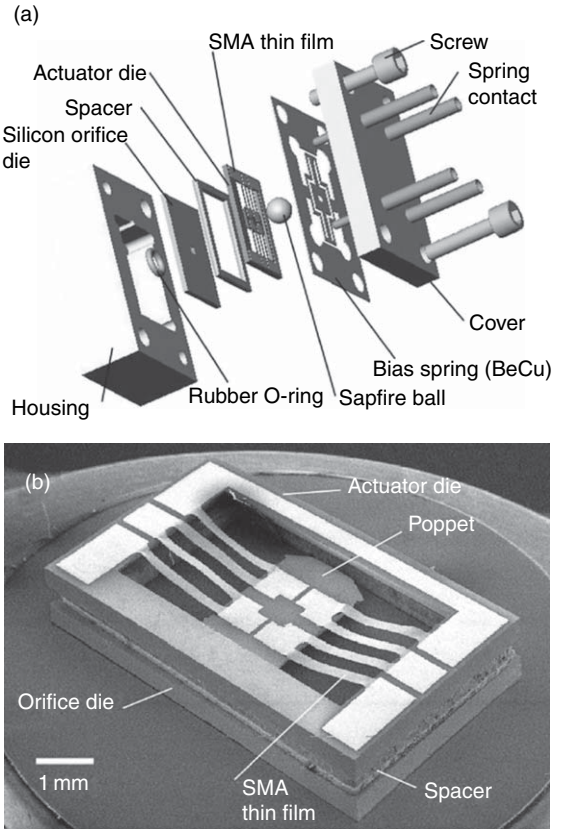


Figure 14 Microvalve driven by shape-memory alloy (SMA) film: (a) exploded view and (b) subassembly of micromachined parts. (Source: Ishida A, Martynov V V 2002 Sputter-deposited shape-memory alloy thin films: Properties and applications. *MRS Bull.* **27**, 111–14; reproduced with permission from MRS Bulletin.)

microvalve manufactured by the TiNi Alloy Company (Johnson 1999). It consists of an actuator die with a poppet controlled by eight TiNi thin-film strips of 3.5 μm thick and 250 μm wide (**Figure 12(b)**), a silicon orifice die, a spacer, and a bias spring. All the elements are assembled in a plastic package (**Figure 12(a)**). The bias spring forces the poppet towards the orifice, so when the current is off the TiNi strips supporting it are elongated allowing the poppet to move down and close the orifice. Resistive heating of the SMA thin films causes it to transform from the martensite phase to the parent phase. By this transformation, TiNi strips recover the original length, lifting the poppet against the bias force and opening the valve. When the current is turned off again, the TiNi strips supporting the poppet are cooled down, transform into the martensite phase, and are easily deformed by the bias spring causing the valve to close. The device has a poppet displacement of ~100 μm and bias force of 0.5 N. It is operated with an electric current of 50–100 mA. The response time is about 10 ms; maximum flow is up to 2000 sccm at 0.13 MPa. This valve survives more than two million cycles.

Recently, a micropump actuated by an SMA thin-film diaphragm was fabricated by Shin *et al.* (2005). **Figure 15** illustrates a cross-sectional view of a pumping chamber. The NiTi membrane of radius of 7 mm is placed above the inflow and the outflow ports. The o-ring and the faceplate clamp the membrane onto the chamber body, leaving a small cavity between the NiTi membrane and the chamber body. Under a bias pressure, the membrane deflects upward at room temperature. When heated, the membrane transforms from the martensite phase to the parent phase and the deflection decreases. The net volume change under the membrane produces a flow rate. In this pump, hot liquid is forced out of the cavity on heating and cool liquid enters the cavity on cooling. This forced cooling mechanism increases the operating frequency of this pump up to 100 Hz, leading to a large flow rate. Their device, which is cubic with a side length of 2 cm and which has NiTi membranes at each of its four faces, achieved a high flow rate of 152 $cm^3\ min^{-1}$ and an output force of 100 N.

Figure 16 shows some devices fabricated by Ishida *et al.* (2005). **Figure 16(a)** shows a microgripper. A similar device was reported by Gill *et al.* (2001). **Figure 16(b)** and **16(c)** shows a microconveyer and a legged robot. The microconveyer can transport an aluminum plate of 48 mg. The microrobot can walk with a load on its back. The walking speed is 400 $\mu m\ s^{-1}$ under a voltage of 8 V and a frequency of 1 Hz. The operation principles of the

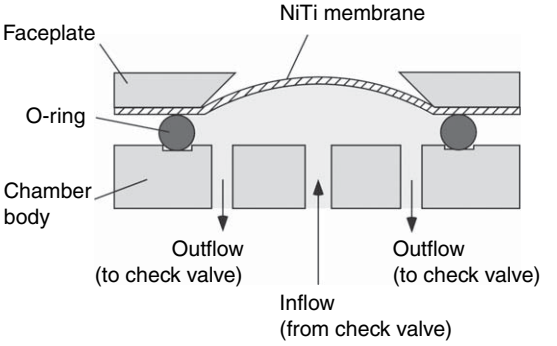


Figure 15 Micropump. (Source: Shin D D, Mohanchandra K P, Carman G P 2005 Development of hydraulic linear actuator using thin film SMA. *Sens. Actuators A* **119**, 151–6.)

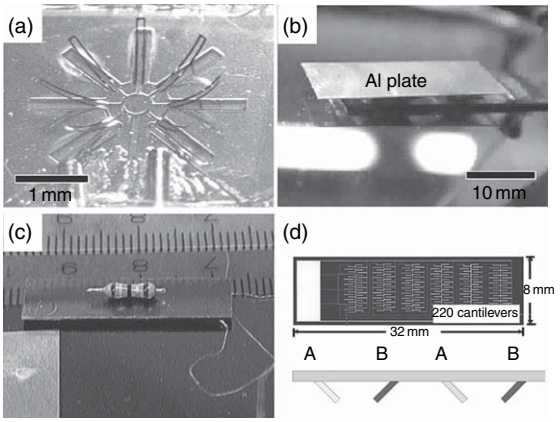


Figure 16 (a) Microgripper; (b) microconveyer transporting aluminum plate; (c) legged robot carrying resistance; and (d) mask pattern for the legged robot and operation principle.

microconveyer and the microrobot are the same, and are shown in **Figure 16(d)**. These devices consist of two groups of actuator arrays arranged oppositely, as shown by the A group and the B group in the figure. By driving the actuator arrays asynchronously, the device can obtain a dynamic motion. The microrobot in **Figure 14(c)** has six pairs of cantilever arrays arranged in a row, each array consisting of 18 cantilevers. Thus, 216 cantilevers (200 μm in width and 1000 μm in length) were fabricated within an area of 25 mm × 8 mm for the walking robot. Integrating many actuators onto a single silicon chip produces many advantages such as enhanced overall function of microsystems, low fabrication cost, and mass productivity. In this respect, shape-memory thin films are desired actuators in comparison with piezoactuators, since they are easier to etch for a fine pattern. Furthermore, the robustness of shape-memory thin films is an important factor for making a microrobot, since the main problem associated with the fabrication of silicon robots has been to achieve sufficient strength in the movable parts.

1.03.3.10 Future Perspectives

Understanding of SMA thin films and development of SMA microactuators have greatly advanced in recent years. Early difficulties of film fabrication such as porosity and oxygen contamination have been solved and composition control technique has shown an advance. Schematic and in-depth study revealed that sputter-deposited thin films possess stable SME and reliable mechanical properties even superior to those of bulk alloys. Further development of ternary alloy thin films is still underway. Microfabrication techniques have also been developed by many researchers. The fabrication technology of microdevices using SMA thin films is, therefore, ready to be transferred to manufacturing. Medical applications and microfluidic components are potential markets. However, in order that shape-memory thin films are widely accepted as promising microactuators, development of device design technique based on the material database is necessary. A complicated device design to introduce a bias force and a complicated behavior of a thin-film SMA actuator as a function of bias stress, heat treatment, and composition make it difficult for mechanical engineers to use SMA thin films as a common actuator. The issue of film adherence remains to be addressed and it sometimes causes a problem in processing. A relatively low frequency and efficiency may be crucial disadvantages in some applications. Even with these disadvantages and problems to be solved, shape-memory thin films are considered to contribute to advances in MEMS in which large force and stroke are essential. In the near future, they will find practical applications owing to their unique properties, which cannot be replaced with other actuating mechanisms.

References

Allameh S M 2003 An introduction to mechanical-properties-related issues in MEMS structures. *J. Mater. Sci.* **3**, 4115–23

Allameh J L S, Buchheit T, Soboejo W O 2003 An investigation of the effects of thickness on mechanical properties of LIGA nickel MEMS structures. *J. Mater. Sci.* **38**, 4129–35

Baek C-W, Kim Y-K, Ahn Y, Kim Y-H 2005 Measurement of the mechanical properties of electroplated gold thin films using micromachined beam structures. *Sens. Actuators A* **117**, 17–27

Bellouard Y 2002 Microrobotics, microdevices based on shape-memory alloys. In: Schwartz M (ed.) *Smart Materials.* Wiley-Interscience, New York, Vol. 2, pp. 620–44

Benard W L, Kahn H, Heuer A H, Huff M A 1998 Thin-film shape-memory alloy actuated micropumps. *J. Microelectromech. Syst.* **7**, 245–51

Buchheit T E, Glass S J, Sullivan J R, Mani S S, Lavan D A, Friedmann T A, Janek R 2003 Micromechanical testing of MEMS materials. *J. Mater. Sci.* **38**, 4081–6

Cheng C-M, Chen R-H 2004 Experimental investigation of fabrication properties of electroformed Ni-based micro mould inserts. *Microelectron. Eng.* **75**, 423–32

Cho H S, Hemker K J, Lian K, Goettert J, Dirras 2003 G Measured mechanical properties of LIGA Ni structures. *Sens. Actuators* **A103**, 59–63

Connolley T, Mchugh P E, Bruzzi M 2005 A review of deformation and fatigue of metals at small size scales. *Fatigue Fract. Eng. Mater. Struct.* **28**, 1119–52

Courteny T H 1990 *Mechanical Behavior of Materials*. McGraw-Hill, New York

Davies S T, Harvey E C, Jin H, Hayes J P, Ghantasala M K, Roch I, Buchaillot L 2002 Characterization of micromachining processes during KrF excimer laser ablation of TiNi shape memory alloy thin sheets and films. *Smart Mater. Struct.* **11**, 708–14

Emery R D, Povirk G L 2003a Tensile behavior of free-standing gold films. Part I. Coarse-grained films. *Acta Mater.* **51**, 2067–78

Emery R D, Povirk G L 2003b Tensile behavior of free-standing gold films. Part II. Fine-grained films. *Acta Mater.* **51**, 2079–87

Espinosa H D, Prorok B C 2003 Size effects on the mechanical behavior of gold thin films. *J. Mater. Sci.* **38**, 4125–8

Espinosa H D, Prorok B C, Peng B 2004 Plasticity size effects in free-standing submicron polycrystalline FCC films subjected to pure tension. *J. Mech. Phys. Solids* **52**, 667–89

Espinosa H D, Panico M, Berbenni S, Schwar K W 2006 Discrete dislocation dynamics simulations to interpret plasticity size and surface effects in freestanding FCC thin films. *Int. J. Plast.* **22**, 2091–117

Fu Y Q, Luo J K, Hu M, Du H J, Flewitt A J, Milne W I 2005 Micromirror structure actuated by TiNi shape memory thin films. *J. Micromech. Microeng* **15**, 1872–7

Fukushige T, Yokoyama Y, Hata S, Masu K, Shimokohbe A 2003 Fabrication and evaluation of an on-chip micro-variable inductor. *Microelectron. Eng.* **67–68**, 582–7

Gill J J, Chang D T, Momoda L A, Carman G P 2001 Manufacturing issues of thin film NiTi microwrapper. *Sens. Actuators A* **93**, 148–56

Gill J J, Ho K, Carman G P 2002 Three-dimensional thin-film shape memory alloy microactuator with two-way effect. *J. Microelectromech. Syst.* **11**, 68–77

Grummon D S, Zhang J P 2001 Stress in sputtered films of near-equiatomic TiNiX on (100) Si: Intrinsic and extrinsic stresses and their modification by thermally activated mechanisms. *Phys. Stat. Sol. (a)* **186**, 17–39

Halford T P 2003 Fatigue and fracture of a high strength, fully lamellar γ-TiAl based alloy. Ph.D. thesis, The University of Birmingham.

Halford T P, Takashima K, Higo Y, Bowen P 2005 Fracture tests of microsized TiAl specimens. *Fatigue Fract. Eng. Mater. Struct.* **28**, 695–701

Hata S, Liu Y, Wada K, Shimokohbe A 2001 Fabrication of thin film metallic glasses and their properties. *J. Jpn. Soc. Precision Eng.* **67**, 1708–13 (in Japanese)

Hata S, Kato T, Fukushige T, Shimokohbe A 2003 Integrated conical spring linear actuator. *Microelectron. Eng.* **67–68**, 574–81

Hemker K J, Last H 2001 Microsample tensile testing of LIGA nickel for MEMS applications. *Mater. Sci. Eng.* **A319–321**, 882–6

Hou L, Pence T J, Grummon D S 1995 Structure and thermal stability in titanium–nickel thin films sputtered at elevated-temperature on inorganic and polymeric substrates. *Mater. Res. Soc. Symp. Proc.* **360**, 369–74

Hruby J 2001 LIGA technology and Applications. *MRS Bull.* **26**, 33–340

Hutchinson J W 2000 Plasticity at the micron scale. *Int. J. Solids Struct.* **37**, 225–38

Ishida A, Martynov V V 2002 Sputter-deposited shape-memory alloy thin films: Properties and applications. *MRS Bull.* **27**, 111–14

Ishida A, Miyazaki S 1999 Microstructure and mechanical properties of sputter-deposited Ti–Ni alloy thin films. *J. Eng. Mater. T. ASME* **121**, 2–8

Ishida A, Sato M 2003 Thickness effect on shape memory behavior of Ti-50.0at.%Ni thin film. *Acta Mater.* **51**, 5571–8

Ishida A, Takei A, Miyazaki S 1993 Shape memory thin film of Ti–Ni formed by sputtering. *Thin Solid Films* **228**, 210–14

Ishida A, Sato M, Takei A, Nomura K, Miyazaki S 1996 Effect of aging on shape memory behavior of Ti-51.3at.pct Ni thin films. *Metall. Mater. Trans. A* **27A**, 3735–59

Ishida A, Ogawa K, Sato M, Miyazaki S 1997 Microstructure of Ti-48.2at.pct Ni shape memory thin films. *Metall. Mater. Trans. A* **28A**, 1985–91

Ishida A, Sato M, Kimura T, Miyazaki S 2000 Stress–strain curves of sputter-deposited Ti–Ni thin films. *Philos. Mag. A* **80**, 967–80

Ishida A, Sato M, Kimura T, Sawaguchi T 2001 Effects of composition and annealing on shape memory behavior of Ti-rich Ti–Ni thin films formed by sputtering. *Mater. Trans.* **42**, 1060–7

Ishida A, Sato M, Tabata O, Yoshikawa W 2005 Shape memory thin films formed with carrousel-type magnetron sputtering apparatus. *Smart. Mater. Struct.* **14**, S216–22

Johnson A D 1999 Thin film shape-memory technology: A tool for MEMS. *Micromach. Devices* **4**, 1–5

Johnson A D, Martynov V V 1997 Application of shape-memory alloy thin film. *Proc. 2nd Int. Conf. Shape Memory and Superelastic Technologies, Asilomar Conference Center*, Pacific Grove, CA, USA, pp. 149–54

Kahn H, Benard W L, Huff M A, Heuer A H 1997 Titanium–nickel shape memory thin film actuators for micromachined valves. *Mater. Res. Soc. Symp. Proc.* **444**, 227–32

Kajiwara S, Kikuchi T, Ogawa K, Matsunaga T, Miyazaki S 1996 Strengthening of Ti–Ni shape-memory films by coherent subnanometric plate precipitates. *Philos. Mag. Lett.* **74**, 137–44

Kawamura Y, Gyobu A, Saburi T, Asai M 2000 Structure of sputter-deposited Ti-rich Ti–Ni alloy films. *Mater. Sci. Forum* **327–328**, 303–6

Kohl M, Skrobanek K D, Quandt E, Schloβmacher P, Schüßler A, Allen D M 1995 Development of microactuators based on the shape memory effect. *J. Phis. IV,* C8-1187-92

Kohl M, Skrobanek K D, Goh C M, Allen D M 1996 Mechanical characterisation of shape memory micromaterials. *SPIE Proc.* **2880**, 108–18

Kohl M, Dittmann D, Quandt E, Winzek B, Miyazaki S, Allen D M 1999 Shape memory microvalves based on thin films or rolled sheets. *Mater. Sci. Eng.* **A273–275**, 784–8

Krulevitch P, Lee A P, Ramsey B R, Trevino J C, Hamilton J, Northrup M A 1996 Thin film shape memory alloy microactuators. *J. Microelectromech. Syst.* **5**, 270–82

Kubota H, Koyama S, Takashima K, Higo Y 2005 Effect of ion-implantation on the mechanical properties of a Zr–Ni amorphous alloy. *J. Metastable Nanocrystalline Mater.* **24–25**, 221–4

Kuribayashi K 1990 Reversible SMA actuator for micron sized robot. *Proceedings of the IEEE Micro Electro Mechanical Systems Workshop*, Napa Valley, CA, USA, pp. 217–21

Last H R, Hemker K J, Witt R 2000 MEMS material microstructure and elastic property modeling.In: de Boer M P, Heuer A H, Jacobs S J, and Peeters E (eds.) *Materials Science of Microelectromechanical Systems (MEMS) Devices II, MRS Proceedings, Vol. 605,* pp. 191–6

Lee Y, Tada J, Isono Y 2005 Mechanical characterization of single crystal silicon and UV-LIGA nickel thin films using tester operated in AFM. *Fatigue Fract. Eng. Mater. Struct.* **28**, 675–86

Li X, Li Z 2003 Nano-sized Si_3N_4 reinforced NiFe nanocomposites by electroplating. *Mater. Sci. Eng. A* **357**, 107–13

Majjad H, Basrour S, Delobelle P, Schmidt M 1999 Dynamic determination of Young's modulus of electroplated nickel used in LIGA technique. *Sens. Actuators* **74**, 148–51

Makino E, Mitsuya T, Shibata T 2000 Micromachining of TiNi shape memory thin film for fabrication of micropump. *Sens. Actuators A* **79**, 251–9

Makino E, Mitsuya T, Shibata T 2001 Fabrication of TiNi shape memory micropump. *Sens. Actuators A* **88**, 256–62

Mazza E, Abel S, Dual J 1996 Experimental determination of mechanical properties of Ni and Ni–Fe microbars. *Microsyst. Technol.* **2**, 197–202

Mineta T, Makino E 2005 Electrochemical etching with high etch factor for micromachining of shape memory alloy. *Proceedings of the 5th Euspen International Conference*, Montpellier, France, pp. 695–8

Mineta T, Hirota K, Makino E, Sugawara T, Toh S, Shibata T 2005 Shape memory thin film for blood vessel holding actuator of thrombus detector. *Key Eng. Mater.* **297–300**, 244–9

Miyazaki S, Ishida A 1999 Martensitic transformation and shape memory behavior in sputter-deposited TiNi-based thin films. *Mater. Sci. Eng. A* **273–275**, 106–33

Miyazaki S, Nomura K, Zhirong H 1994 Shape memory effect and superelasticity developed in sputter-deposited Ti–Ni thin films. *Proc. Int. Conf. Shape Memory and Superelastic Technologies (SMST-94)*, Pacific Grove, CA, USA, pp. 19–24

Miyazaki S, Nomura K, Ishida A, Kajiwara S 1997 Recent development in sputter-deposited Ti–Ni-based shape memory alloy thin films. *J. Phys. IV* 7, C5-275-80

Nakamura Y, Nakamura S, Ataka M, Fujita H 1996 Micromachining process for thin-film SMA actuators. *Proc. IEEE Conf. Emerging Technologies and Factory Automation: ETFA'96*, Kauai, HI, USA, pp. 493–7

Nakamura S, Nakamura Y, Ataka M, Fujita H 1997 A study on patterning method of TiNi shape memory thin film. *T. IEE Japan* **117-E**, 27–32

Ogi H, Nakamura N, Tanei H, Hirao M 2005 Advanced resonant-ultrasound spectroscopy for studying anisotropic elastic constants of thin films. In: Buchheit T E, Minor A M, Spolenak R, and Takashima K (eds.) *Thin Films – Stresses and Mechanical Properties XI. MRS Proceedings,* Vol. 875, Warrendale, PA, USA, pp. 3–14

Ogi H, Shimoike G, Hirao M, Takashima K, Higo Y 2002 Anisotropic elastic-stiffness coefficients of an amorphous Ni–P film. *J. Appl. Phys.* **91**, 4857–62

Qu A A, Zhang Q X, Zou Q B, Balasubramanian N, Yang P, Zeng K Y 2002 Characterization of TiAl alloy films for potential application in MEMS bimorph actuators. *Mater. Sci. Semicond. Process.* **5**, 35–8

Ray C A, Sloan C L, Johnson A D, Busch J D, Petty B R 1992 A Silicon-based memory alloy microvalve. *Mater. Res. Soc. Symp. Proc.* **246**, 161–166

Roch I, Bidaud Ph, Collard D, Buchaillot L 2003 Fabrication and characterization of an SU-8 gripper actuated by a shape memory alloy thin film. *J. Micromech. Microeng.* **13**, 330–6

Saotome Y, Imai K, Sawanobori N 2003 Microformability of optical glasses for precision molding. *J. Mater. Process. Technol.* **140**, 379–84

Sato K, Yoshioka T, Ando T, Shikida M, Kawabata T 1998a Tensile testing of silicon film having different crystallographic orientations carried out on a silicon chip. *Sens. Actuators* **A70**, 148–52

Sato M, Ishida A, Miyazaki S 1998b Two-way shape memory effect of sputter-deposited thin films of Ti51.3at.% Ni. *Thin Solid Films* **315**, 305–9

Sawaguchi T, Morio S, Ishida A 2004 Grain-size effect on shape-memory behavior of $Ti_{35.0}Ni_{49.7}Zr_{15.4}$ thin films. *Metall. Mater. Trans. A* **35A**, 111–19

Sharpe W N Jr., McAleavey A 1998 Tensile properties of LIGA nickel. In: Friedrich C R and Vladimirsky Y (eds.) *Materials and Device Characterization in Micromachining, Proceedings of SPIE,* Vol. 3512, Bellingham, WA, USA, pp. 130–7

Sharpe W N, Jr., Bagdahn J, Jackson K, Coles G 2003 Tensile testing of MEMS materials – Recent progress. *J. Mater. Sci.* **38**, 4075–9

Shi L, Sun C, Gao P, Zhou F, Liu W 2006 Mechanical properties and wear and corrosion resistance of electrodeposited Ni–Co/SiC nanocomposite coating. *Appl. Surf. Sci.* **252**, 3591–9

Shibata A, Tarumi R, Ogi H, Hirao M, Takashima K, Higo Y 2005 Evaluation of the elastic constants of an amorphous thin film by using EMAR and RUS/Laser methods. *J. Soc. Mater. Sci. Jpn* **54**, 1012–15

Shin D D, Mohanchandra K P, Carman G P 2005 Development of hydraulic linear actuator using thin film SMA. *Sens. Actuators A* **119**, 151–6

Son D, Jeong J H, Kwon D 2003 Film-thickness considerations in microcantilever-beam test in measuring mechanical properties of metal thin film. *Thin Solid Films* **437**, 182–7

Spearing S M 2000 Materials issues in microelectromechanical systems (MEMS). *Acta Mater.* **48**, 179–96

Stauss S, Schwaller P, Bucaille J-L, Rabe R, Rohr L, Michler J, Blank E 2003 Determining the stress–strain behaviour of small devices by nanoindentation in combination with inverse methods. *Microelectron. Eng.* **67–68**, 818–25

Takashima K, Halford T P, Rudinal D, Higo Y, Takeyama M 2005 Micro fracture toughness testing of TiAl based alloys with a fully lamellar structure. In: Mills M J, Inui H, Clemens H, and Fu C-L (eds.) *MRS Symposium. Proceedings.* Vol. 842**,** Warrendale, PA, USA, pp. 175–80

Takashima K, Higo Y 2004 Characterization of micro-sized specimens. *Kensa-Gijutsu (Inspection Eng.)* **9**, 1–5 (in Japanese)

Takashima K, Ogura A, Ichikawa Y, Higo Y 2001 Anisotropic fracture behavior of electroless deposited Ni–P amorphous alloy thin film. In: Kahn H, de Boer M, Judy M, and Spearing S M (eds.) *Materials Science of Microelectromechanical Systems (MEMS) Devices III. MRS Proceedings,* Vol. 657**,** Warrendale, PA, USA, pp. EE5.12.1–6

Takeuchi S 2000 A three-dimensional shape memory alloy microelectrode with clipping structure for insect neural recording. *J. Microelectromech. Syst.* **9**, 24–31

Tarumi R, Takashima K, Higo Y 2003 Crystallographic features of oriented nanocrystals induced by focused-ion-beam irradiation for an amorphous alloy. *J. Appl. Phys.* **94**, 6108–15

Tsuchiya T, Tabata O, Sakata J, Taga Y 1998 Specimen size effect on tensile strength of surface micromachined polycrystalline silicon thin films. *J. Microelectromech. Syst.* **7**, 106–13

Tsuchiya T, Hirata M, Chiba N, Udo R, Yoshitomi Y, Ando T, Sato K, Takashima K, Higo Y, Saotome Y, Ogawa H, Ozaki K 2005 Cross comparison of thin-film tensile-testing methods examined using single-crystal silicon, polysilicon, nickel, and titanium films. *J. Microelectromech. Syst.* **14**, 1178–86

Walker J A, Mehregany M, Gabriel K J 1990 Thin-film processing of TiNi shape memory alloy. *Sens. Actuators A* **21–23**, 243–6

Wang W H 2005 Elastic moduli and behaviors of metallic glasses. *J. Non-Crystalline Solids* **351**, 1481–5

Wang R X, Zohar Y, Wong M 2002 Residual stress-loaded titanium–nickel shape-memory alloy thin-film micro-actuators. *J. Micromech. Microeng.* **12**, 323–7

Wang W H, Dong C, Shek C H 2004 Bulk metallic glasses. *Mater. Sci. Eng. R* **44**, 45–89

Xu D, Wang L, Ding G, Zhou Y, Yu A, Cai B 2001 Characteristics and fabrication of NiTi/Si diaphragm micropump. *Sens. Actuators A* **93**, 87–92

Biographies

Kazuki Takashima received the B.Eng. degree in Metallurgy from Kumamoto University, Japan, in 1978 and the M.Eng. and Ph.D degrees in Metallurgy from Tokyo Institute of Technology, Japan, in 1980 and 1984, respectively. He is currently a professor of Department of Materials Science and Engineering, Kumamoto University, Japan. His research interests include the evaluation of mechanical properties of microsized materials for MEMS applications.

Akira Ishida is the leader of the actuator function group in the sensor materials center at the National Institute for Materials Science in Tsukuba, Japan. His most recent research involves martensitic transformations, physical vapor deposition, mechanical properties of thin films, and microelectromechanical systems (MEMS). He has also studied superalloys, hot corrosion and coating, and hard coatings formed by ion plating. He has received Ph.D. degrees in materials science from Osaka University.

1.04 Polymers

Ponnambalam Ravi Selvaganapathy, McMaster University, Hamilton, ON, Canada

Glossary

Alkanethiol A class of organic compounds that consist of a carbon chain (alkane) attached to a sulfur-based chemical group (thiol). These compounds have a known propensity to form self-assembled monolayers on gold surfaces.

Autoclave A device that uses high-pressure steam to kill microorganisms and clean medical equipment.

Base pair (bp) Constituents of the DNA strand. The common nucleic acid bases such as adenine (A), guanine (G), thymine (T), cytosine (C), and uracil (U) pair selectively with each other in nucleic acids. The A–T (or T–A) and C–G (or G–C) base pairs form the ''rungs'' of the DNA double helix, holding the two complementary strands together.

Cationic polymerization A process in which the active end of the growing polymer molecule is a positive ion. The polymerization proceeds via a radical chain reaction with intermolecular transfer of the radical center.

Collagen An insoluble fibrous protein, which is the principal component in connective tissue. It provides the rigid framework for tissue and cells. It is found in tissues, cartilage, and bones. Collagen is formed when tropocollagen molecules polymerize.

Da (Daltons) Atomic Mass Unit. It is used to express atomic and molecular masses. It is defined as 1/12th of the mass of one atom of carbon-12.

DMD Digital Mirror Display. These consist of thousands of MEMS mirrors whose angle could be individually controlled, hence producing an image.

DNA Deoxyribonucleic acid, a high molecular weight molecule composed of two chains coiled into a double helix. DNA is an universal component of living matter. It is located on the chromosomes in the cell nucleus, and it consists of nucleoproteins. The set of DNA molecules forms the chromosomes, which contain all the information of the genetic code.

EDM A process for machining metal parts by a very rapid succession of electrical discharges in an insulating liquid. The tool, usually in the form of a wire, is taken over the workpiece according to a design and the part is formed.

Genomic sequencing Process of elucidating the sequence in which base pairs occur in a particular strand of DNA.

Lewis acid An electron pair acceptor. This substance acts in a similar manner as H^+ ion in accepting a pair of electrons.

LIGA Abbreviation for lithography, electroforming, and molding, which are the fundamental process stages of this technology. The lithography method used is X-ray deep lithography, which used to define high aspect ratio structures in polymers. Electroplating in this mold creates a master, which can be used for precision replication.

LoC Lab-on–a-chip devices that were first developed in the early 1990s that miniaturized and automated the sequence of steps required to perform complex but repetitive laboratory-based analysis of biomolecules.

PCR Polymerase chain reaction, a method that produces several million copies of a single fragment of DNA in several amplification cycles.

III–V materials Elemental materials that occupy groups III and V of the periodic table of the elements. The combination of these materials in a crystalline structure produces semiconducting properties, which is used in electronics.

φX-174/HaeIII A DNA marker sample with DNA fragments ranging in size from 72 to 1353 bp. This is widely used to determine the size of unknown fragments of DNA.

1.04.1 Introduction

Miniaturization of laboratory-based analytical techniques into tiny lab-on-a-chip (LoC) devices has the capability to dramatically improve performance, throughput, and the cost associated with these methods (Dittrich and Manz 2006). This in turn has far-reaching consequences in areas as diverse as medical diagnostics, drug delivery, drug discovery, analytical chemistry, and molecular diagnosis. It was observed that miniaturization of the analysis equipment reduced the need for a large volume (expensive) of samples and reagents. Similarly it reduces the time of analysis significantly as unit operations like heating, mixing, and metering can be performed more accurately and quickly with smaller volumes. Once the

operation sequence of an analysis is determined and miniaturized, these devices can be produced in thousands and can be used to perform thousands of parallel analysis leading to higher throughput. Miniaturization and the microfabrication techniques used allow precise replication of these devices leading to repeatable performance, which is critical for comparing the analysis by different operators. This eliminates the human factor in the analysis and increases confidence in the analysis (Selvaganapathy *et al.* 2003).

Polymers are ideally suited – and hence are widely investigated – for use in LoC devices for biomedical and microfluidic applications because of their low cost, chemical inertness, low electrical and thermal conductivity, suitability for surface modification, and compatibility with biological materials. Biomedical devices including sensors and LoC systems are required to be disposable to avoid cross contamination between samples. Polymer substrates cost ~2 cents per square centimeter while glass costs ~10–20 cents per square centimeter (Becker and Gartner 2000). Other microfluidic applications typically involve a large footprint of several square centimeters and cannot be conveniently miniaturized without loss of performance (Becker and Gartner 2000). The low cost of polymeric materials and their processing techniques are the twin advantages that provide the impetus for developing processing technologies for the microfabrication of polymeric microelectromechanical systems (MEMS)/microfluidic systems. This chapter discusses some of the widely used polymers in MEMS, their properties, and their fabrication methods.

1.04.1.1 Polymers and Their Classification

Polymers are macromolecules of high molecular masses (higher than 1000 Da and having more than 100 repeat units) made from smaller molecules called monomers. Polymerization is the process of producing polymers from monomeric units through chemical reactions. These reactions can lead to linear polymers or cross-linked polymers depending on the position of the reaction groups in the monomer and the cross-linker. The latter case leads to a large 3D network structure. The process is also statistically dependent and hence a range of polymer chain lengths are produced. Hence polymers usually do not have a defined melting temperature but soften above a range called the melt interval. A wide variety of polymers from different starting monomers have been polymerized, and they are classified according to their structures and behavior (Nicholson 1997).

Polymers have traditionally been classified according to their response to thermal treatment. Thermoplastics (such as polyethylene and polystyrene) are mostly linear or branched polymers that melt upon the application of heat and resolidify when cooled. They do not have cross-linking and hence their thermal behavior is reversible – they undergo the same phase transition without hysteresis. Above the glass transition temperature (T_g), these materials become plastic and they can be molded into specific shapes that they retain at lower temperatures.

Elastomers (such as polydimethylsiloxane (PDMS)) are weakly cross-linked polymers. They can be easily stretched and revert back to their original shape upon release. Since they are cross-linked they do not melt before reaching their decomposition temperature. Thermosetting polymers (such as bakelite and vulcanized rubber) are heavily cross-linked and hence the molecular movement for elasticity is not possible. They are normally rigid and brittle and do not soften significantly at higher temperatures. The key parameter of interest in microfabrication is the glass transition temperature (T_g). This is the temperature below which the polymer solidifies and becomes hard and brittle. Thus, the polymer can be heated above T_g and molded using a mold insert. Then the mold can be cooled to below the glass transition temperature and the plastic component demolded.

1.04.2 Microfabrication

Microfabrication originated in the semiconductor industry as a process for constructing, layer by layer, using different materials, the structure of a transistor and its interconnections. Materials are deposited or removed from a planar substrate using thin-film deposition or etching techniques, producing structures in the order of 0.1–5 μm. The definition of regions for selective removal or deposition is done using photolithography. These techniques have been modified and adapted to produce high aspect ratio microstructures in the order of 20–200 μm using methods broadly classified as bulk and surface micromachining. Bulk micromachining describes processes that involve microfabrication aimed at selective removal of material from the substrate. One popular example is wet etching of silicon. Other methods like dry etching using the Bosch

processes (Larmer and Schilp 1996) or other micromachining methods like LIGA (German acronym for X-ray lithography (X-ray Lithographie), Electroplating (Galvanoformung), and Molding (Abformung)) process (Ehrfeld *et al.* 1988) produce structures with high aspect ratios (~10). Surface micromachining involves processes aimed at selective addition of thin films of a sacrificial material (to produce topology) and several structural materials. The substrate is not affected during the processing and provides only a stable base layer. The sacrificial material is removed at the end of processing to produce a 3D structural feature. These microfabrication techniques generally use silicon and III–V materials as the working materials primarily due to their origin and their development in the semiconductor industry. Since the early 1990s (Manz *et al.* 1990) traditional semiconductor microfabrication techniques have been adapted for the construction of LoC devices in silicon and glass (Harrison *et al.* 1993, Liang *et al.* 1996). These devices perform a complex series of chemical and biochemical reactions and unit operations such as metering, mixing, and pumping in the microscale. Genomic sequencing, drug discovery, and disease diagnostics have been the main drivers in this effort to produce LoC devices, which enabled the low-cost, high-throughput, low-reagent consumption analysis to be performed faster and more efficiently.

However, silicon as a substrate and structural material has several disadvantages. Silicon-based devices have high substrate electrical and thermal conductivities. Special insulation layers (oxide, nitride, or polymers) have to be provided in the instance of electrokinetic pumping, which is a widely used actuation mechanism in microfluidic systems. Similarly, thermal insulation or isolation islands have to be provided for microreaction chambers to prevent heat transfer between chambers or interaction with biomolecules in the microchannels. Glass provides excellent isolation properties, but it is very difficult to machine high aspect ratio complex 3D structures. Polymers, in contrast, are moldable to complex shapes and dimensions and have isolation properties comparable to glass.

Silicon also interacts with the biomolecules and fluids in the channels, leading to adhesion, accumulation, and fouling. Typically, surface modification (silanization, polyelectrolyte assembly, atom-free radical polymerization) is performed to prevent interactions. However, surface modification leads to extra processing cost, is variable in performance, and degrades over time. Polymeric materials, in contrast, can be made inert to interaction with biomolecules (e.g., polyethylene glycol) and can be modified to have a uniform surface property for electroosmotic flow (EOF).

Polymers are better suited for these applications because of their low cost, chemical inertness, low electrical and thermal conductivity, suitability for surface modification, and compatibility with biological materials. Further, biochemical analysis has been traditionally performed in either glass equipments that could be autoclaved or in plastic that could be disposed after use. However, the processing techniques used in the semiconductor industry – such as dry etching, sputter deposition, chemical vapor deposition – are high-temperature processes, and hence they are generally unsuitable for polymeric processing. Microfabrication of LoC and other MEMS/microfluidic devices therefore require the development of techniques for polymer microfabrication and replication, which are cheap and cost-effective for producing thousands of low-cost disposable devices. Polymeric materials are also available in a compliant form at low temperatures, which allows other techniques such as hot embossing, injection molding, and casting to be used for microfabrication. These are bulk micromachining techniques where the substrate made of the polymer is fashioned as a negative replica of the master, which retains its structural integrity at these temperatures. Fabrication costs of molded microparts are insensitive to the complexity of design. Once a mold insert (master) has been made, several thousand parts can be molded with little effort. The replication can be performed outside the clean room, while still obtaining a resolution of the device that is similar to that in photolithography. This reduces the cost of the devices significantly – making the technology applicable for disposable biomedical devices. An excellent review of these alternative microfabrication technologies, their applications, and limitations has been published in Heckele and Schomburg (2004). Surface micromachining techniques have also been adapted to suit low-temperature processing for polymeric materials. These allow construction of devices with multiple materials and construction of devices with complex functionality. A brief description of the different methods widely used to structure 3D objects in the microscale is given in the subsequent sections.

1.04.2.1 Injection Molding

Injection molding is a well-established technique in macroscopic production of products such as CDs and DVDs. Hence the machines, technology, and process know-how have been rapidly adapted for microscale molding (Hanemann *et al.* 2000, Larsson *et al.* 1997). A number of groups have demonstrated the feasibility of replication in thermoplastic polymers of microstructures (Despa *et al.* 1999, Larsson *et al.* 1997, Madou *et al.* 2001, Piotter *et al.* 1997, 2001, Weber *et al.* 1996). An injection molding process starts with the formation of molds. Silicon micromachining (Manz *et al.* 1990), LIGA process (Ruprecht *et al.* 1995), SU-8 photolithography (Jian *et al.* 2001), and microEDM are used to produce microscale features, which serve as mold inserts. The mold cavity with the mold inserts is evacuated and heated to a temperature above the glass transition temperature (T_g) of the polymer. A horizontal injection unit heats the polymer and injects the viscous melt into the mold cavity. The entire cavity is then cooled below T_g and subsequently the molding part is demolded. **Figure 1(a)** shows the process flow. Injection molding is the most ideal technique adapted to industrial production due to short cycle times (~minutes). The quality of surface features within the fabricated device is entirely dependent on the quality and the precision of the mold insert.

1.04.2.2 Hot Embossing

Hot embossing refers to the replication of a microstructure from a hot mold insert (master) onto a thin film of thermoplastic material. **Figure 1(b)** depicts the process. In this method, a thermoplastic film is inserted into the embossing mold with mold inserts on either side. The tool is evacuated and the mold inserts are pressed against the thermoplastic film with great force. The film and the mold insert are heated above T_g. The plastic fill is squeezed from the elevated parts and fills the valleys and generally confirms to the contours of the mold inserts. The setup is cooled and the mold insert is removed from the plastic (Heckele *et al.* 1998).

Unlike injection molding, hot embossing requires the polymer flow only over short distances. Two major effects of this are the reduction in stress and the faster cycle time. Hot embossed parts have very low stress as there is only local modification of the surface of the film. Hence these are ideal for membrane structures and optical components. Since the flow is local, the viscosity of the film need not be reduced as much as in the case of injection molding. This reduces the need to heat the film to a high temperature thereby reducing the heating and cooling times. The stress and temperature effects reduce shrinkage during cooling and enhance demolding. Thus, microstructures with high aspect ratios and thin structural features, like membranes, can be fabricated using hot embossing as compared to injection molding (Heckele and Schomburg 2004). Since high aspect ratio structures (>10) can be replicated very easily in thin films (1–100 μm), this technique has been adapted to replicate patterns on photoresist films using a process called nanoimprint lithography.

1.04.2.3 Casting

In this process a material is introduced into a mold in its liquid state, allowed to solidify inside the mold, and then removed, thus producing a replica of the microstructured master. **Figure 1(c)** illustrates the process. Techniques similar to those described earlier for producing injection molding mold inserts are used to produce the microstructured master. Casting materials used could be hot, liquid metals or meltable plastics (such as thermoplastics). A variation of this process is what is termed as reaction casting. Here the plastic – instead of being melted and cast into the mold – comes as two parts, which are mixed together. The mixing with heat or UV light input coldsets the polymers through cross-linking – increasing its viscosity by several orders of magnitude. The initial low viscosity of the two-component mixture allows the polymer to flow and wet the entire surface of the mold insert and accurately replicate the shape. This process is typically used for nonthermoplastic materials such as thermosetting plastics (bakelite), epoxies (Sethu and Mastrangelo 2004), and elastomers (PDMS) (Xia and Whitesides 1998). The essential requirements of this process are that complete mixing and distribution of the two components have to take place and the chemical reaction (cross-linking) has to take place throughout the bulk for uniform strength. This can be achieved only by comparatively long cycle times (~hours). However, features in the scale of a few nanometers can be easily replicated.

1.04.2.4 Stereolithography

Stereolithography is a rapid prototyping technology first developed in the 1980s (Hull 1986) for creating a 3D prototype for automotive and aeronautical

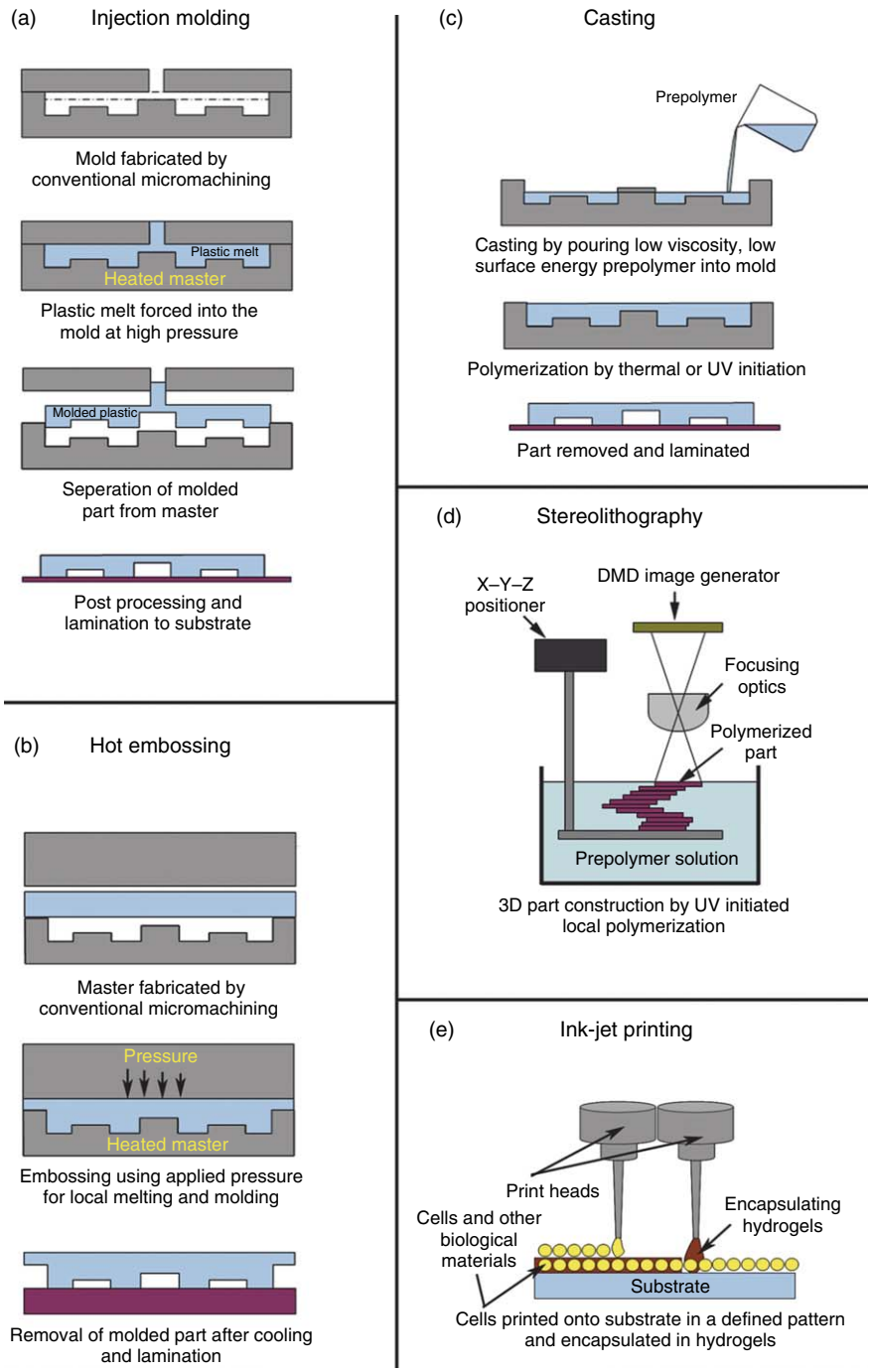


Figure 1 Injection molding, hot embossing, casting, stereolithography and inkjet printing process flow.

applications mainly for fluid dynamic testing and visualization. The object to be formed is deconstructed into a series of 2D layers and reproduced by laser-induced polymerization of a resin. Figure 1(d) illustrates the process. In this, a laser beam is focused on the free surface of a photosensitive liquid, which induces polymerization of the local region of the liquid, transforming it to a polymerized solid. The laser is scanned across the surface with control from a software program to trace a 2D feature. Resolution depends on the laser, the scanning apparatus used, and the diffusion in the photosensitive liquid. The

penetration depth of the laser light into the liquid determines the thickness of the layer. This produces the shape of one layer of the object to be formed. Fresh liquid is spread on top of this photodefined layer and the second layer is produced similarly. The resolution in this technique is ~150–200 μm in three dimensions. The effort to improve resolution and prototype of smaller objects started in 1993 (Ikuta and Hirowatari 1993, Takagi and Nakajima 1993). The thickness of individual layers was reduced by using a glass container attached to an X–Y–Z micropositioner to determine the thickness of the resin with which the light interacts. The laser illumination mechanism along with the shutters for patterning are attached to the X–Y–Z positioner and moved with it. The major problem with this approach is that the polymerized resin adheres to the glass plate. Free surface polymerization is then done wherein the part formed is attached to the X–Y–Z positioner and moved close to the free surface of the resin/air interface. The gap between the part formed and the free surface determines the thickness of the layers (Zhang *et al.* 1998). Increasing the throughput of the method by using optical fibers to split the laser light and to manufacture several parts simultaneously has been investigated (Ikuta *et al.* 1996). Micrometer scale resolutions have been obtained using these methods. LCD displays (Bertsch *et al.* 1997) and digital mirror display (DMD) devices (Bertsch *et al.* 1998) have been used to generate patterns and eliminate X–Y movement and to provide a lithography-type process that decreases the time to generate layers and increases throughput. However, these methods require the resin to be spread uniformly on the already formed part, time for stabilization of the resist after vertical movement of the part, support for the part, and an inability to form freely moving structures.

Two-photon absorption, a nonlinear optical phenomenon, has been used to produce parts inside the bulk of the resin instead of close to the interface. This eliminates most of the problems mentioned above and, when combined with LCD and DMD masking technology, provides a high throughput method for fabrication (Maruo and Kawata 1998, Maruo *et al.* 1997). Nonphotoreactive materials including ceramics can be fabricated by incorporating them as suspension particles in the photosensitive resin. This, once polymerized, can be fired at a high temperature to burn off the resin, leaving behind a sintered 3D microstructure (Zhang *et al.* 1999). Structures with lateral dimension of ~100 nm have been obtained using this technique (Sun and Kawata 2003). An excellent review of stereolithography and its applications has been published in Bertsch *et al.* (2001).

1.04.2.5 Ink-jet Printing

Ink-jet printing has been widely used for the reproduction of text by propelling microscale droplets of ink through a narrow nozzle. Recently efforts have been made to adapt this technology for the printing of polymeric liquids other than inks. The print head of the ink-jet printer consists of a reservoir with an actuator and a nozzle. The reservoir holds the ink, which is propelled through the nozzle upon actuation. The transducer produces volumetric change either due to the flexure of the piezoelectric material or due to thermal expansion by heating. This creates pressure waves, which force the fluid to move toward the exit. The droplet size is typically of the size of the nozzle (Wallace 1989) and drops with diameters ranging from 20 to 120 μm can be generated. The critical aspects of this method are the composition of the ink and its viscosity. The viscosity should be <20 m Pa s so that the energy imparted by the actuator to the fluid is not viscously dissipated. Hence polymers have typically been printed as melts (such as waxes) or in the colloidal form (Wong *et al.* 1988). Hot melt waxes solidify on contact with the substrate and retain their shapes. Spots (20–40 μm diameter) with drop placement accuracy of 1 μm have been used to fabricate masks for metal layer etching of transistors (Wong *et al.* 2002).

For biopolymers such as DNA, proteins, and cells, this is the only viable method for pattern definition. These polymers are highly temperature sensitive and lose their properties on exposure to other solvents. Ink-jet printers have been used to deliver specific reagents to individual spots on a slide to synthesize a different DNA sequence at each site (Lemmo *et al.* 1997). One of the issues with printing of biological materials, especially proteins and cells, is the loss of functionality due to dehydration on the substrate. This has been solved by printing proteins and cells embedded in hydrogels (Kesari *et al.* 2005). **Figure 1(e)** shows the schematic of the use of ink-jet printing of biological materials such as cells in a hydrogel matrix.

1.04.3 Polymers and Processes

A wide variety of polymers are used to fabricate microstructures using both traditional micromachining processes such as bulk and surface micromachining and nontraditional processes such as injection molding, casting, hot embossing, and stereolithography. Thermoplastic polymers such as polymethylmethacrylate, polycarbonate, polyvinylidenefluoride, polysulfone, polystyrene, polyvinylchloride, polypropylene, polyetheretherketone, polyoxymethylene, and poly-amide have been successfully used to fabricate microstructures using injection molding-, hot embossing-, and casting-based approaches (Heckele and Schomburg 2004). However, the most widely used polymeric materials for MEMS/microfluidics include parylene (*para*-xylylene), PDMS, SU-8, hydrogels, porous polymers, biodegradable polymers, polyesters, polyimide, and paraffin. This section provides a detailed description of the properties of these polymers and their fabrication processes.

1.04.3.1 Parylene

Parylene is the common name for a class of polymers called poly-*para*-xylylene. The basic member of the series, called parylene N, is a completely linear, highly crystalline material. It is commercially available in three forms, which are variations of a basic polymer backbone of xylylene formed by the replacement of 1–4 atoms in the benzene ring as shown in **Figure 2**. Parylene C and D are other variations, which have chlorine atoms at various positions in the basic backbone. The commercialization of parylene coating was initiated with the realization that the dimeric form was stable and with the development of a vapor-phase deposition process (Gorham 1966). It has been used in a wide variety of applications such as encapsulation for microelectronic circuits (Lin and Wong 1992, Olson 1989), as interlayer dielectrics (Selbrede and Zucker 1997), and for strengthening wire bonds (Flaherty 1995) in microchip packaging. Similarly, a number of applications of parylene in microfluidics include microchannels (Man *et al.* 1997; Webster and Mastrangelo 1997), microvalves (Carlen and Mastrangelo 2002, Rich and Wise 1999, Wang *et al.* 1999), membrane filters (Yang *et al.* 1998), and other micromachined devices.

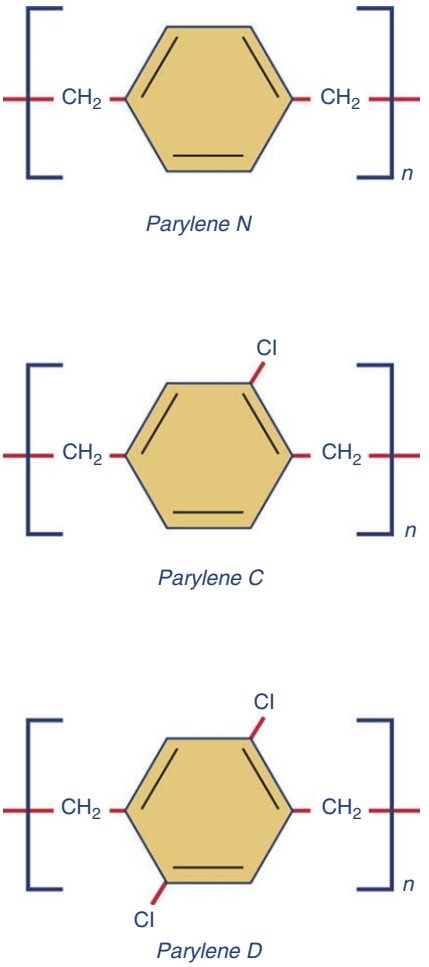


Figure 2 Chemical structure of different kinds of commercially available parylene polymers. Reproduced with permission from Specialty Coatings Systems Inc.

1.04.3.1.1 Polymerization

Parylenes have to be polymerized *in situ* and cannot be formed by extrusion or molding techniques due to their high molecular weight (~500 000), which causes their melting temperature and crystallinity to be high. Parylene polymers are deposited by vapor deposition at room temperature and in vacuum. **Figure 3** shows the schematic of the process flow. The dimer form of parylene, a solid at room temperature, is the source. It is sublimated at 140–160°C and vaporized into the pyrolysis chamber where it splits into two monomers at 680°C. The reactive monomer is then allowed to flow into a deposition chamber where it deposits and polymerizes on surfaces that are below 100°C. The deposition pressure is ~100 mtorr and hence the deposition is conformal. This allows the formation of uniform conformal thin films.

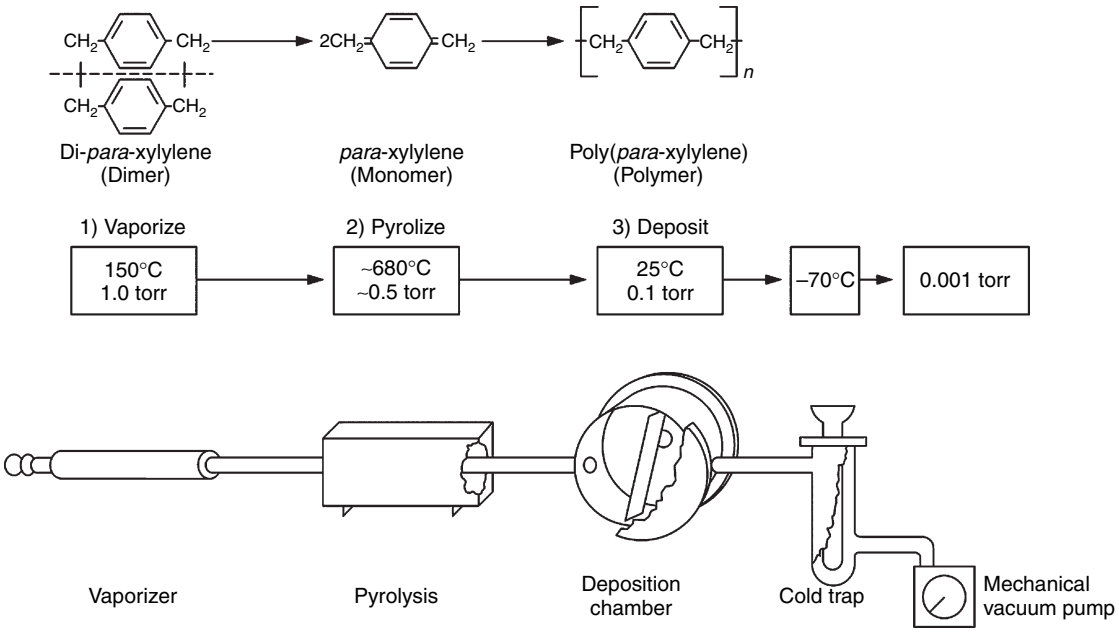


Figure 3 Parylene polymerization process flow. Reproduced with permission from Specialty Coatings Systems Inc.

1.04.3.1.2 Properties

Parylene thin films are extremely conformal even with high aspect ratio structures due to the vapor-based deposition process. Under an operating pressure of 0.1 torr, the mean free path of the molecules is much smaller than the feature size, leading to uniform condensation and excellent control over thickness. Low surface roughness and pinhole-free coating can be obtained for films of thickness >0.5 pm. The films also are chemically pure due to the lack of initiators and catalysts in the polymerization process. The room temperature deposition and polymerization ensure formation of stress-free films. These films have excellent dielectric breakdown properties, especially below 1 μm.

With regard to microfluidic devices, parylenes are clear and possess low autofluorescence allowing visualization of biological and chemical interactions through optical means. They exhibit very little absorption in the visible region and are, therefore, transparent and colorless. Below 280 nm, both parylene N and C absorb strongly. They are chemically and biologically inert and hence resistant to damage by acids, bases, corrosive body fluids, electrolytes, enzymes, and proteins.

The low solubility in organic or other media allows them to maintain the integrity and concentration of the biomaterials and chemicals in organic solvent without significant absorption. Parylenes have superior barrier properties when compared to silicones and other polymeric materials. Their moisture vapor permeability is $1.7 \times 10^{-16}\,\mathrm{kg\,m\,N^{-1}\,s^{-1}}$, which is an order of magnitude lower than for silicones. They resist chemical attack at room temperature and are insoluble in all organic solvents up to 150°C. Parylene C can be dissolved in chloronapthalene at 175°C, and parylene N is soluble at the solvent's boiling point (265°C).

Surface properties of parylene are modified on exposure to strong chemicals. For instance, etchants such as buffered hydrofluoric acid (BHF) and hydrofluoric acid (HF) make the surface more hydrophobic. Chromium etchant solution makes the surface hydrophilic. Substituted parylenes that are capable of reacting with biomolecules to promote strong attachment have been developed recently (Lahann *et al.* 1998). Parylene is biocompatible as evidenced by its use in cochlear implants (Bell *et al.* 1997) as well as for PCR (Man *et al.* 1997) and biochemical reactions (Brahmasandra *et al.* 1998).

1.04.3.1.3 Adhesion

Adhesion of parylene to silicon, SiO_2, and Si_xN_y surfaces is greatly improved by the silanation of silicon dioxide and silicon nitride surfaces. While the silanation of silicon itself will improve parylene adhesion, parylene adheres better to the silicon dioxide and nitride films. Silanation can be done either by dip-coating the silane (vinyltrichlorosilane

in isopropyl alcohol) solution and spin-drying the substrate or by vapor deposition prior to parylene deposition. Adhesion of parylene to most polymers isimproved by exposing the polymer surface to 250-mtorr, 100-W oxygen plasma for 1 min. This plasma treatment etches the surface of the polymer and exposes the active polymer nucleation sites for parylene to bind. Using this treatment, parylene C has been found to adhere to polyimide, photoresist, and polycarbonate. Metals such as Cr, Au, and Ti adhere quite well to parylene while Al does not. Silanation process described above can be used to improve the adhesion of Al to parylene.

1.04.3.1.4 Microfabrication

As mentioned before, parylene films cannot be formed by molding or casting and vapor deposition is the only technique for forming thin conformal films. Solvent-resistant property enables photoresists, developers, and other chemicals to be used directly on parylene. This enables the formation of complex multilayer structures using parylene as a structural material and photoresist as a sacrificial material. No wet etchants are available for parylene. Dry etching in O_2 and CF_4 atmospheres at low pressures produces etch rates of 0.2 μm min^{-1} at 250 W and 50 mtorr (Tacito and Steinbruchel 1996).

A surface micromachined process for construction of parylene microchannels has been demonstrated (Webster *et al.* 1998). The surface micromachined fabrication process is shown in **Figure 4**. In this, the substrate merely provides mechanical support and hence materials other than silicon such as glass, polycarbonate, and other plastics can be used. The process begins with the initial coating of the substrate with a thin conformal parylene layer. Adhesion is usually obtained by plasma treatment as described above. Then, electrodes and other functional parts of the microfluidic and MEMS device are deposited and patterned. Thick (20 μm) photoresist is spun-cast and photolithographically patterned to define the microchannel – its microstructure, shape, and dimensions. A second layer of parylene is then conformally deposited to form the top and sides of the channel. The adhesion of this layer is also assisted by a short oxygen plasma treatment. A thick photoresist layer is used as a mask to etch the reservoir, and contact pad openings are etched using oxygen plasma. A photodefinable silicone rubber (polysiloxane) is used to make reservoir rings. The wafer is partially released in acetone to remove the photoresist mask and treated with an oxygen plasma for a short duration to improve adhesion. The channel is then released in acetone for 36 h without agitation, rinsed and dried, and then immersed in chrome etchant several times to etch the remaining chromium at the bottom of the channel. The release time can be shortened by constant agitation or by heating the acetone. A number of devices such as DNA separation devices (Webster *et al.* 1998), PCR systems (Man *et al.* 1997), cochlear implants (Bell *et al.* 1997), microvalves for drug delivery (Carlen and Mastrangelo 2002, Rich and Wise 1999), and microneedles have been fabricated using these techniques.

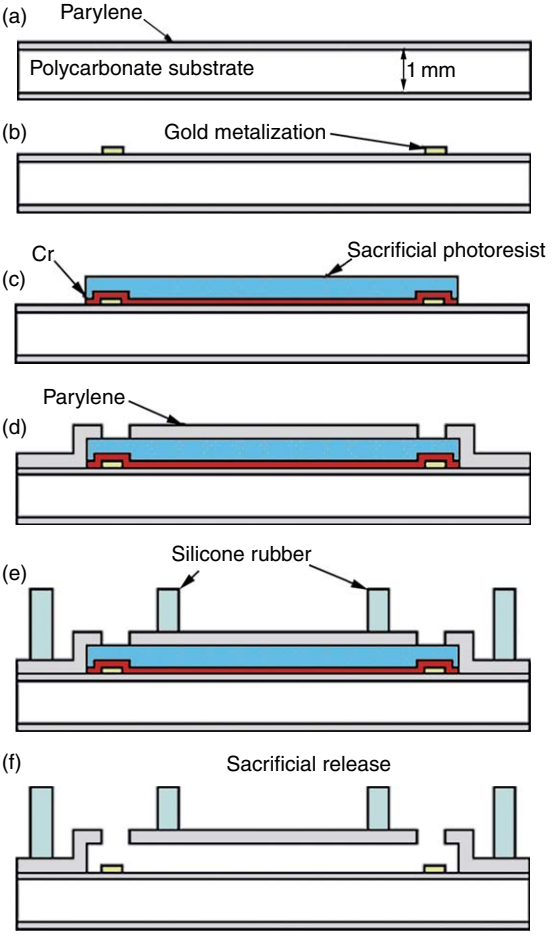


Figure 4 Parylene surface micromachined process flow. (a) Parylene coated on the substrate forming the bottom of the microchannel structure; (b) electrodes patterned using gold; (c) Cr layer to protect the electrode and lithography of photoresist to form microchannel structure; (d) parylene deposition to form top and side walls of the microchannels along with dry etching at reservoirs; (e) silicone rubber photodefinition to form tall reservoirs; (f) sacrificial removal of resist to form hollow microchannel. Reproduced with permission from Webster J R, Burke D T, Burns M A, Mastrangelo C H 1998 An inexpensive plastic technology for microfabricated capillary electrophoresis chips. *Proc. Tas '98 Workshop*, Banff, Canada, pp. 249–52, © 1998 Kluwer.

1.04.3.2 Polydimethylsiloxane

Polydimethylsiloxane is one of the most widely used polymers in MEMS, especially in microfluidic and bioMEMS. It belongs to a class of polymers called silicones, which are synthetic polymers containing a Si–O backbone. To this backbone, organic groups are frequently attached to the silicon atoms via a Si–C bond. PDMS has a repeating $(CH_3)_2SiO$ unit. Depending upon the number of repeat units in the polymer chain and the degree of cross-linking, the material can be produced in different rheological forms such as fluids, emulsions, lubricants, resins, elastomers, and rubbers (Brook 2000). The low rigidity of the backbone allows the methyl groups to be easily exposed, resulting in low intermolecular interactions of PDMS and also its low surface tension.

1.04.3.2.1 Polymerization

Silicones are made from silicon and methyl chloride in a process known as the direct reaction yielding methyl chlorosilanes. They are distilled, and the dimethyldichlorosilane ($(CH_3)_2SiCl_2$) is hydrolyzed to produce PDMS (Noll 1968). Both linear and cyclic siloxane oligomers are formed simultaneously in the hydrolysis reaction. These are the basic raw materials (monomers) from which a number of useful silicone products can be made through cross-linking.

As mentioned previously, siloxane polymers occur in a variety of common forms – fluids, gels, elastomers, and resins. Fluids are typically linear chains of PDMS, which are terminated with a trimethylsilyl group. Gels are lightly cross-linked PDMS fluids, where the cross-link is introduced either through a trifunctional silane – such as CH_3SiCl_3 giving a branched silicone structure – or through a chemical reaction between a Si-vinyl group on one polymer chain and a hydrogen-bonded silicon in another (Noll 1968). This produces a 3D network that can be swollen with PDMS fluids to produce a gel-like structure. Elastomers are highly cross-linked fluids with very little free fluid in the matrix. Fillers, such as amorphous silica, are frequently added to the matrix to give greater reinforcement to the network and thereby increase its strength.

PDMS in its elastomeric form is generally used for microfluidic and bioMEMS applications. The commercial form is available as a kit with the PDMS base solution and a curing agent to cross-link fluid PDMS with an elastomer by an organometallic cross-linking reaction. The siloxane base oligomers (prepolymer) contain vinyl-terminated end groups ($CH_2{=}CH–$). The base solution also contains a platinum-based catalyst and a silica filler (dimethyl-vinylated and trimethylated silica). The cross-linking oligomers, typically dimethyl-methylhydrogen siloxane, tetrakis (dimethylsiloxyl silane) (TDS) $(HSi(CH_3)_2O)_4Si$), or a copolymer poly(dimethylsiloxane-*co*-methylhydrogenosiloxane) $(CH_3)_3Si–O–[Si(HCH_3)–O]_x–[Si(CH_3)_2–O]_y–Si(CH_3)_3$, contain at least 4–11 silicon hydride bonds each. The curing agent also contains an inhibitor (tetramethyl tetravinyl cyclotetrasiloxane). The platinum-based catalyst catalyzes the addition of the SiH bond across the vinyl groups, forming $Si–CH_2–CH_2–Si$ linkages (Fu *et al.* 2003). The multiple reaction sites on both the monomer chain and the cross-linking chain enable 3D cross-linking, as shown in **Figure 5**. One advantage of this type of addition reaction is that no waste products such as water are generated. If the ratio of the curing agent to the base is increased, a harder, more cross-linked elastomer is obtained. Heating also accelerates the cross-linking reaction.

1.04.3.2.2 Microcontact printing

One of the first applications of this polymer has been in microcontact printing. In 1993, the Whitesides group demonstrated that PDMS inked with an alkanethiol and when brought into contact with a gold-coated surface can form a monolayer of these molecules in the areas of contact (Kumar and Whitesides 1993; Kumar *et al.* 1995). The schematic of the process is described in **Figure 6**. Exact pattern replication by this printing process requires both accurate definition of a pattern and conformal contact of the substrate to facilitate ink transfer from the master. PDMS – due to low surface tension and elastomeric properties – has ideal properties to satisfy both these criteria.

A photoresist layer patterned by conventional optical lithography can be used to produce the master. High aspect ratio patterns have been accomplished using SU-8 epoxy-based photoresist, which after sufficient heat treatment tends to form a stable master capable of numerous replications. The PDMS prepolymer (1:10 mixture of the cross-linker and PDMS base) is cast over this master and allowed to cross-link at room temperature for 1 h and heated to 65°C for 1 h. The PDMS base having low surface energy wets the entire surface down to the molecular level forming an exact negative replica of the relief. The cross-linking of the base then solidifies the material into an elastic monolith. The 1:10 ratio allows easy delamination of the PDMS from the master by peeling. The master is used to produce several PDMS molds, thus replicating micro- and

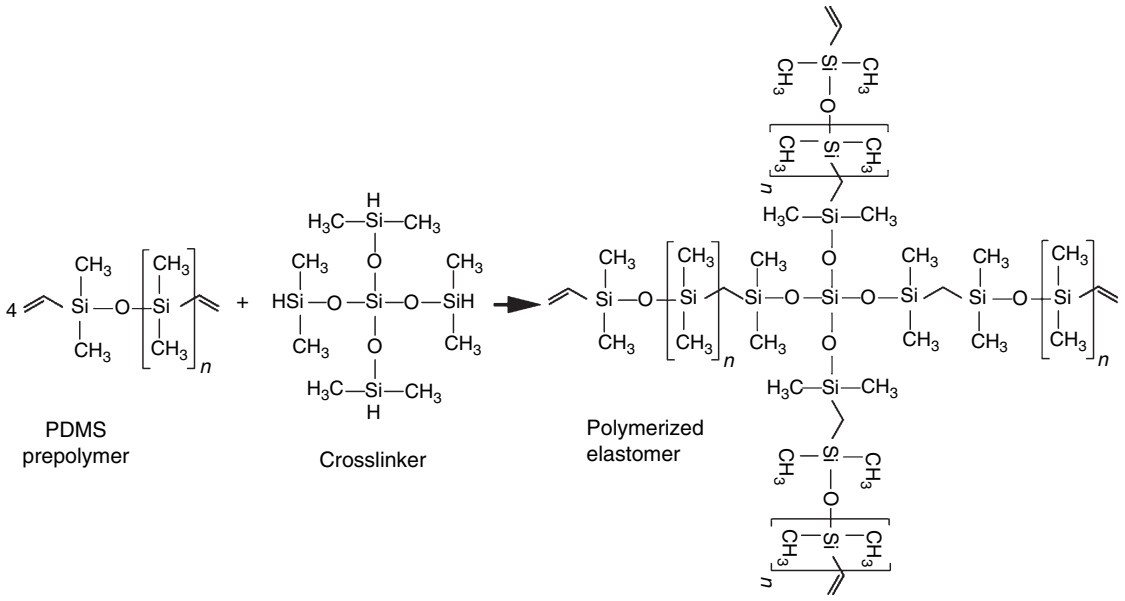


Figure 5 Chemical structure of PDMS and its polymerization.

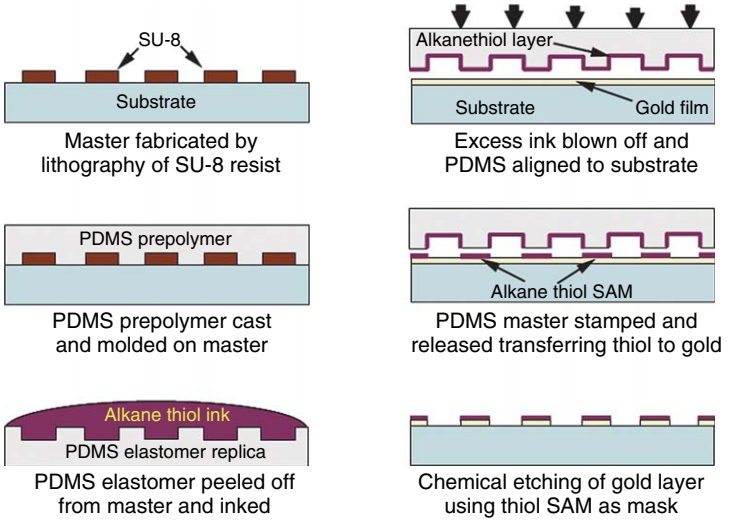


Figure 6 Microcontact printing process.

nanoscale features without expensive clean room processes (Kumar and Whitesides 1993).

This mold can then be inked with alkanethiols either using an ink pad or by solvent evaporation (Xia *et al.* 1996). These methods allow the deposition of a thin layer of the material on the features – either on the elevated surfaces or conformally – thereby maintaining the topography. This inked mold is then stamped on the gold film substrate. The thiol group has a natural affinity for the gold surface and tends to form self-assembled monolayers on the surface. Because of the topography, only those regions that are elevated in the mold imprint thiols on the gold surface. Hence the pattern on the PDMS mold is replicated on the gold surface in alkanethiols. Etching in gold etchant allows the replication of this pattern in gold (Xia *et al.* 1996).

This method allows the direct transfer of a micrographic pattern present on the flexible PDMS layer to be replicated on any substrate without expensive microscale definition processes like UV lithography. A number of different materials have been patterned using this technique, including metals and dielectrics (Joon *et al.* 1995, St. John and Craighead 1996,

Wang *et al.* 1997a). The important feature of using PDMS as the feature replication mold is its flexibility and its ability to achieve conformal contact between the PDMS stamp and the substrate. The mold can be used to replicate patterns on a nonplanar surface. These techniques are able to generate features on both curved (Jackman *et al.* 1995) and reflective substrates and rapidly pattern large areas, which are not possible in conventional lithography. A comprehensive review of microcontact printing (Xia and Whitesides 1998) and its modifications (Deng *et al.* 2000, Huck *et al.* 2000, Jackman *et al.* 1999, Jeon *et al.* 1999, Qin *et al.* 1999, Shah *et al.* 2000) provides greater detail about the process and its applications.

1.04.3.2.3 Microfluidic channels

The method mentioned earlier has been adapted for constructing microfluidic channels using PDMS elastomers as the structural material (Effenhauser *et al.* 1997). Till then microfluidic devices were made of silicon and glass-based substrates (Harrison *et al.* 1993, Liang *et al.* 1996). These were rigid and their fabrication processes require a series of cumbersome and expensive steps using traditional microfabrication techniques. However, the use of a microfabricated template of silicon to cast PDMS and the use of replica molding for high aspect ratio structures reduced the cost of the process (Effenhauser *et al.* 1997). The process is shown in **Figure 7**. **Figure 8** shows the replicated microchannel in PDMS.

In this process, lithographic patterning and wet chemical etching of silicon are used to produce a rigid high aspect ratio (20-μm-high) master. The silicon wafers are silanized in 3% (v/v) dimethyloctadecylchlorosilane in toluene for 2 h to facilitate peeling off the PDMS replica. A 10:1 mixture of PDMS base and its curing agent is poured over the wafer. After curing the PDMS for 4 h at 65°C, the PDMS replica is peeled off the master. Holes serving as reservoirs and providing access to the channels are punched through the bulk material. The device is then placed on a thin slab of PDMS in order to form a closed channel system of four equivalent walls. The hermetically sealed microchannels are readily formed by mere adhesion without applying external force (Effenhauser *et al.* 1997). The chips manufactured using this technique withstood pressures of 1 bar without clamping. The adhesion between the PDMS layers is reversible and the two layers can be separated easily by using a blade. This is useful in cleaning and reusing the microchannels. Buffer solution and separation media are pipetted into the reservoirs, and the microchannels are filled by applying vacuum to one of the reservoirs. Capillary electrophoresis separation of φX-174/HaeIII DNA is performed in this device as shown in **Figure 9** to demonstrate the feasibility of microchannel construction and its use in bioMEMS (Effenhauser *et al.* 1997).

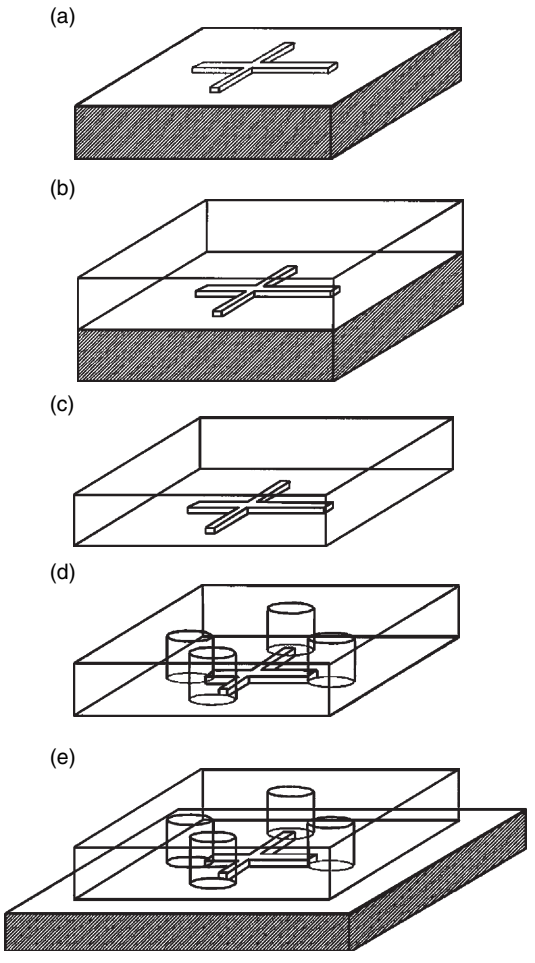


Figure 7 Microchannel construction in PDMS. (a) Silicon master wafer with positive surface relief, (b) premixed solution of Sylgard 184 and its curing agent poured over the master, (c) cured PDMS slab peeled from the master wafer, (d) PDMS slab with punched reservoir holes, and (e) ready-to-use device placed on a slab of PDMS. Reproduced with permission from Effenhauser C S, Bruin G J, Paulus A, Ehrat M 1997 Integrated capillary electrophoresis on flexible silicone microdevices: Analysis of DNA restriction fragments and detection of single DNA molecules on microchips. *Anal. Chem.* **69**, pp. 3451–7, © 1997 American Chemical Society.

However, this process produced microchannels that are hydrophobic and leak under higher pressures. Further, producing a master is expensive, involving single-crystal silicon wafers. A technique for cheaper

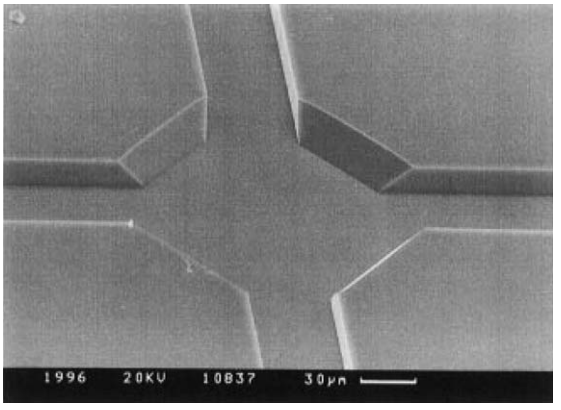


Figure 8 SEM of constructed PDMS microchannels. Reproduced with permission from Effenhauser C S, Bruin G J, Paulus A, Ehrat M 1997 Integrated capillary electrophoresis on flexible silicone microdevices: Analysis of DNA restriction fragments and detection of single DNA molecules on microchips. *Anal. Chem.* **69**, pp. 3451–7, © 1997 American Chemical Society.

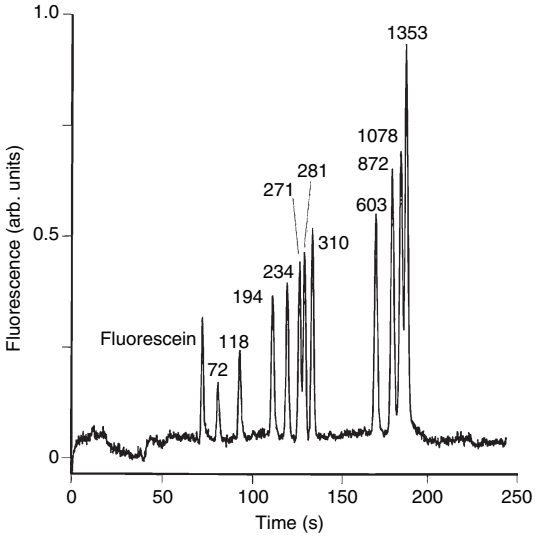


Figure 9 Separation of biomolecules (DNA fragments) in PDMS microchannels. DNA concentration, 1.3 ng/μL (370 pM); separation medium, 0.5% HPC in 100 mM Tris/100 mM boric acid in the presence of 100 nM YOYO-1, 0.5 μM 9-aminoacridine, 25 mM NaCl, and 21 nM fluorescein. L = 35 mm; E = 500 V/cm. Reproduced with permission from Effenhauser C S, Bruin G J, Paulus A, Ehrat M 1997 Integrated capillary electrophoresis on flexible silicone microdevices: Analysis of DNA restriction fragments and detection of single DNA molecules on microchips. *Anal. Chem.* **69**, pp. 3451–7, © 1997 American Chemical Society.

production of master using photoprinting (Qin *et al.* 1996, Xia and Whitesides 1998) of the masks and using photolithographic definition of master has been developed (Duffy *et al.* 1998). In this technique, a computer-aided design (CAD) file is printed on a transparency sheet by using a high-resolution laser printer. The file contains the pattern, which determines the regions blocked from exposure to UV light, and the printed transparency is used as a mask in contact photolithography of SU-8 high aspect ratio-negative resists. The transparency mask is 10–50 times cheaper to make and has a fast turnaround time. However, the resolution is limited to the accuracy of the photoprinter (>20 μm) while that of a chrome mask is ~500 nm (Duffy *et al.* 1998). After development, the pattern of SU-8 photoresist is used as the master to cast replicas of PDMS. Apart from providing high aspect ratio patterns, SU-8 has sufficient adhesion to the substrate to provide reliable replication on repeated use. Polyurethane and other epoxies have also been cast on the relief to provide more reliable single-piece masters (Duffy *et al.* 1998). The PDMS processing is similar to the process described for microcontact printing. The difference between the processes is in the type of SU-8 resist used and the height of microstructures formed. Microcontact printing typically can be accomplished with structures that are 5–10 μm high. However, microchannel structures typically require heights in the range of 20–100 μm.

1.04.3.2.4 PDMS bonding

A method to irreversibly bond PDMS to a substrate in order to construct devices to withstand higher pressures has also been developed (Duffy *et al.* 1998). The technique involves oxidizing the surface groups using oxygen plasma. In its natural state the PDMS surface is hydrophobic with CH_3 groups on the surface. This makes the surface inert and hence affects adhesion with other substrates. Surface oxidation of the PDMS in plasma converts some of the methyl groups on the PDMS backbone into hydroxyl groups – changing the surface from hydrophobic to hydrophilic. Bringing the two oxidized PDMS surfaces into conformal contact forms a tight, irreversible seal, improving adhesion of the substrates, and pressure tolerance of microchannels. The Si–OH groups at the surface of the PDMS slabs form a covalent Si–O–Si bond between them. The thickness of the oxidized surface layer is approximately 130–160 nm and consists of SiO_x, i.e., silicon bonded to three or four oxygen atoms. However, the oxidized layer becomes thinner and starts to crack on prolonged plasma treatment (>30 s). If the treated samples are left to age in an ambient atmosphere, they become hydrophobic again. This is due to low

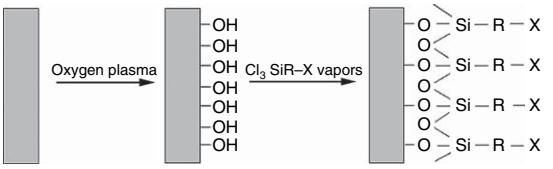


Figure 10 Surface functionalization of PDMS by silyl chloride vapor.

molecular weight PDMS migrating to the surface (Hillborg *et al.* 2000).

An important consequence of oxidizing the PDMS surfaces for microfluidic systems that use electroosmotic pumping is that the walls of the channels present a layer of silanol (SiOH) groups to an aqueous solution in contact with them. These surfaces are charged (SiO–) when the solution is neutral or basic, and so the channels support EOF toward the cathode. These OH groups can be further reacted with different terminal groups in the presence of silyl chloride vapors to form a surface with desirable properties (Xia and Whitesides 1998), as shown in **Figure 10**.

Although plasma oxidation is popular, other techniques such as corona discharge (Efimenko *et al.* 2002) and UV light exposure (Hillborg and Gedde 1998) have also been used to reduce hydrophobicity, particularly to reduce protein adsorption in microchannels.

PDMS surfaces can also be rendered hydrophilic by chemical treatment (Jo *et al.* 2000) using sodium silicate, which produces a permanent hydrophilic surface inside the microchannels. Sodium silicate has been used as a low-temperature adhesive for glass bonding applications (Ito *et al.* 2002, Wang *et al.* 1997b) and it has demonstrated excellent hydrophilic and adhesive properties. Other methods include sol–gel techniques (Roman *et al.* 2005), silanization (Papra 2001), chemical vapor deposition (Lahann *et al.* 2003), atom transfer radical polymerization (Xiao *et al.* 2002), and polyelectrolyte multilayers (Liu *et al.* 2000). Polyelectrolyte multilayers have also been cross-linked to provide greater stability to the surface layer (Makamba *et al.* 2005).

1.04.3.2.5 Multichannel structures

The basic microchannel fabrication technique described above can be modified to produce multiple stacked layers with microscale features embedded in them using a technique known as multilayer soft lithography that combines soft lithography with the capability to bond multiple patterned layers of elastomer. The resulting microstructure is shown in **Figure 11** (Unger *et al.* 2000). Multilayer structures are constructed by bonding

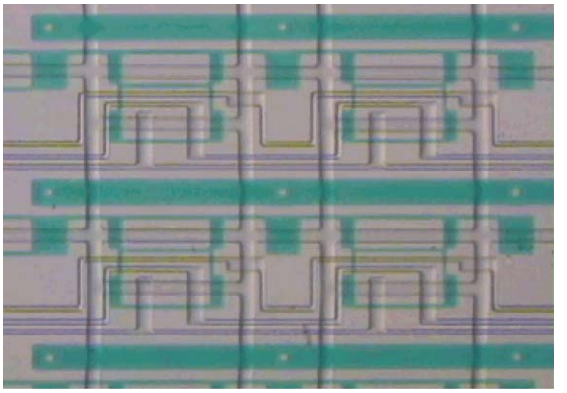


Figure 11 Multilayer microfluidic channels fabricated in PDMS. Reproduced with permission from Liu Y, Fanguy J C, Bledsoe J M, Henry C S 2000 Dynamic coating using polyelectrolyte multilayers for chemical control of electroosmotic flow in capillary electrophoresis microchips. *Anal. Chem*. **72**, 5939–44, © 2000 American Chemical Society.

layers of elastomer, each of which is separately cast from a micromachined mold. This technique exploits the two-component curing process to obtain PDMS elastomers. The bottom layer has an excess of one of the components (PDMS base), whereas the upper layer has an excess of the other (cross-linker). After separate curing of the layers, the upper layer is removed from its mold and is placed on top of the lower layer, where it forms a hermetic seal. Because each layer has an excess of one of the two components, reactive molecules remain at the interface between the layers. Further curing causes the two layers to irreversibly bond. The strength of the interface equals the strength of the bulk elastomer (Unger *et al.* 2000). A monolithic 3D patterned structure composed entirely of elastomer is produced by this process. Additional layers are added by simply repeating the process. Each time the device is sealed on a layer of opposite polarity and cured, another layer is added. Valves and pumps using pneumatic control have been constructed using this approach (Unger *et al.* 2000).

1.04.3.3 SU-8

Photoresists are polymeric solutions whose properties can be selectively altered using UV light. They have been typically used as the selective definition layer on other thin films in semiconductor applications. They have also been used as a sacrificial layer for the construction of microchannels (Man *et al.* 1997). However, it is with the development of SU-8 – a negative tone resist – that photoresists have been used as a structural layer for MEMS and microfluidic applications. SU-8 was developed by IBM (Lee *et al.* 1995) for LIGA

applications. A modified form of this material was used to produce high aspect ratio structures for MEMS applications using conventional lithography (Lorenz *et al.* 1997).

1.04.3.3.1 Polymerization

SU-8 resist consists of EPON SU-8 resin (an epoxy) as its main constituent. The resin is an aromatic hydrocarbon with an epoxide group at either end. The solvent used is γ-butyrolacton. The ratio of the solvent to the resin determines the viscosity and thereby the thickness of the resist layer at a particular spinning speed. The versatility of this resist was evidenced by the fact that thicknesses from 750 nm to 450 μm can be obtained by a single spin using a conventional spinner by varying the solvent/resin ratio (Lorenz *et al.* 1997). The resist also consists of 10 wt.% of photoinitiator, which is a photo Lewis acid generator.

A typical processing sequence of the resist is as follows. SU-8 is spun on a silicon wafer and the solvent is allowed to dry by prebaking at 95°C. The prebake time depends on the thickness of the resist and is typically 15 min for a 100-μm thickness. Exposure of the resist is performed on a UV aligner and the UV light interacts with the photoinitiator, converting it into a Lewis acid. The resin has very low UV absorption and hence a uniform exposure dose, and Lewis acid generation throughout the bulk of the resist can be achieved. The resin then undergoes cationic polymerization induced by the Lewis acid to form a cross-linked epoxy layer in the areas exposed, as shown in **Figure 12**. A dose of 300–400 mJ cm^{-2} at 365 nm is required for complete cross-linking of a 100-μm-thick resist layer. Removal of the unexposed resist is done using propylene glycol monomethylether acetate solution and rinsed in

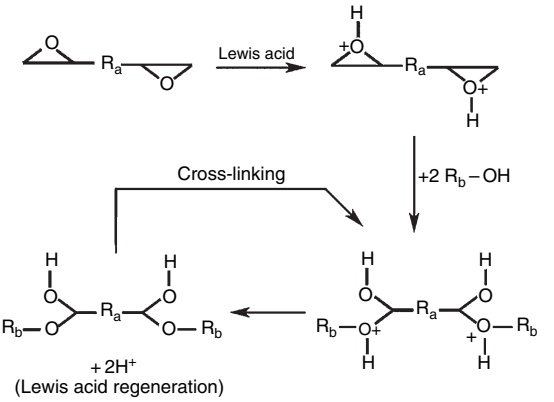


Figure 12 SU-8 polymerization process.

isopropyl alcohol. The development time depends on the thickness and is typically 15 min for a 100-μm-thick resist layer. The resist can be stripped at this stage using hot 1-methyl-2-pyrrolidon solution. Further hard baking of the resist at higher temperatures of 120–200°C for 30 min makes it resistant to metal etchants and other solvents typically used in MEMS fabrication (Lorenz *et al.* 1997).

High aspect ratio structures (>18) with a thickness range from 80– to 1200 μm with conventional mask aligners have been discussed in Lorenz *et al.* (1998). Higher aspect ratios (>60) have been achieved using X-ray sources for LIGA applications (Jian *et al.* 2003). Another feature of SU-8 is that taller structures can be constructed by using multiple layers. A 100-μm layer can be spun, prebaked, and exposed, and this can then be used as a substrate for the next layer. Multiple layers with alignment using different masks can produce complex 3D structures. After the final layer has been exposed and baked, all the layers can be developed at the same time. **Figure 13** shows a number of microstructures fabricated using SU-8

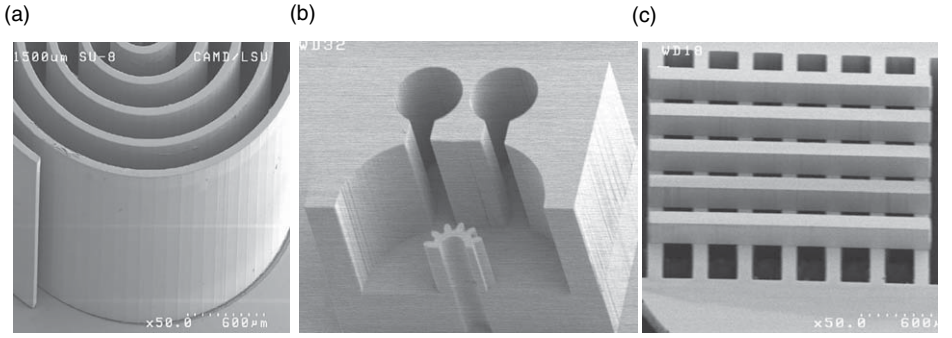


Figure 13 Part manufactured in SU-8: (a) 1500 μm tall microstructure, (b) cross section of engine housing, (c) multi-level 3D microstructures made by the combined use of SU-8 and PMMA resists. Reproduced with permission from Jian L, Desta Y M, Goettert J, Bednarzik M, Loechel B, Yoonyoung J, Aigeldinger G, Singh V, Ahrens G, Gruetzner G, Ruhmann R, Degen R 2003 SU-8 based deep X-ray lithography/LIGA. *Micromach. Microfabricat. Process Technol. VIII, Proc. SPIE* **4979**, 394–401, © 2003 SPIE.

including a microrotary engine housing, which has a size of 1.6 mm × 1.6 mm, and three multilevels of 0.3, 0.6, and 0.9 mm, with the smallest feature size of 15 μm (Jian *et al.* 2003).

1.04.3.3.2 Channel fabrication

There are a number of surface micromachining methods for the fabrication of microchannels using SU-8 as depicted in **Figure 14**. One of the early methods involves a process flow as shown in **Figure 14(d)**. Here the SU-8 layer is spun, prebaked, exposed, and developed to produce a pattern of the microchannel. A dry photoresist film – Riston (Dupont) – is laminated over the SU-8 layer. Access holes are then patterned in the Riston layer to provide fluidic connection to the SU-8 microchannel (Heusckel *et al.* 1998).

Another technique, depicted in **Figure 14(a)**, involves the use of a sacrificial material (typically positive tone photoresist) to define the microchannel structure. SU-8 is cast over this sacrificial mold and the access holes are lithographically defined. Then the sacrificial resist is removed to produce an open microchannel (Guerin *et al.* 1997). A variation of the methods is depicted in **Figure 14(b)**. Here a layer of SU-8 is spun, prebaked, and exposed into patterns of microchannels. Then a thin metal layer is deposited on top to protect the unexposed regions at the bottom SU-8 layer from exposure. A second SU-8 layer is then spun, prebaked, and exposed to a pattern to form access holes for the microchannels. Subsequently the second SU-8 layer is developed, the metal layer etched, and the unexposed resist in the first SU-8 layer also developed to produce the channel structure (Guerin *et al.* 1997). The fourth technique, as shown in **Figure 14(c)**, starts with spinning of a double layer of SU-8. A long exposure to a 2-MeV proton beam through a mask is done to produce a

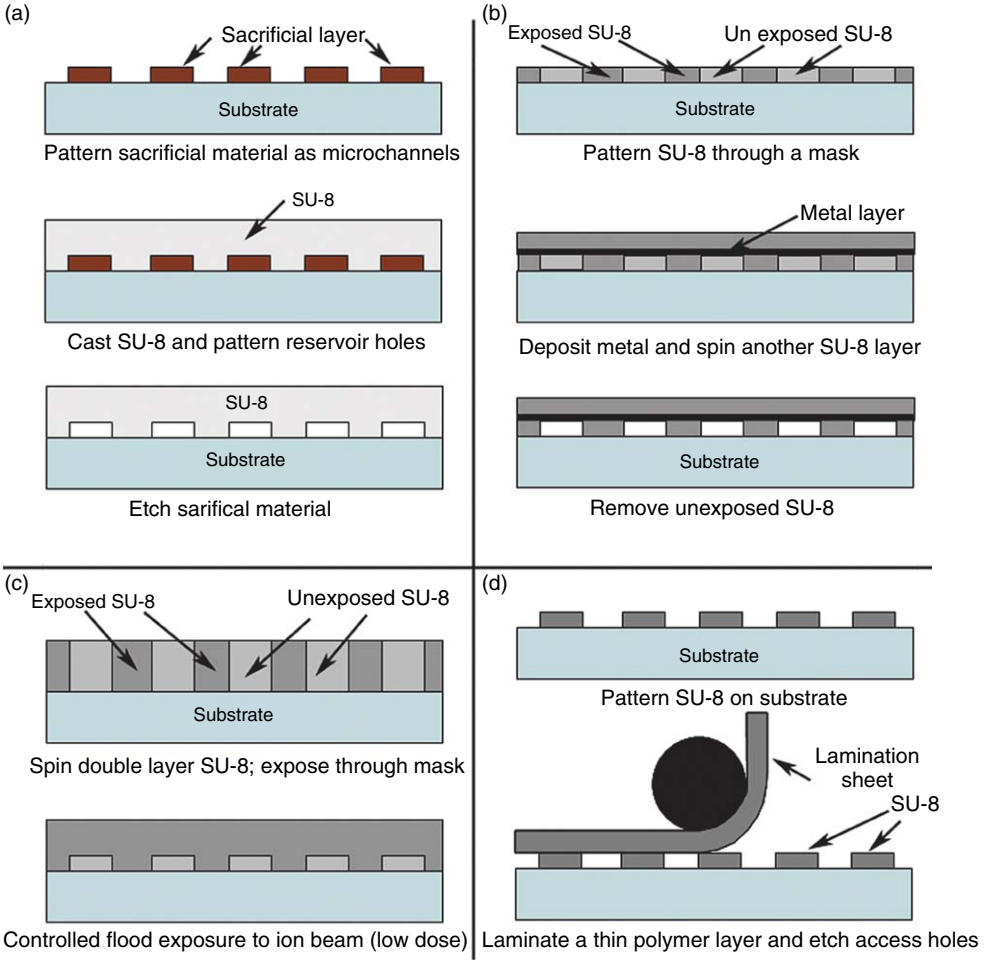


Figure 14 SU-8 microchannels construction techniques.

pattern of the microchannels. In this exposure the entire thickness of the resist is exposed. Subsequently the wafer is flood exposed to a 0.6-MeV beam. These low-energy protons have a thickness range of 10 µm and hence the top layer of SU-8 is exposed while the bottom layer is not. Subsequently the unexposed regions are developed to produce the embedded microchannel structure (Tay *et al.* 2001).

1.04.3.4 Hydrogels

A gel is generally defined as a loose network of cross-linked polymers that retains much of its solvent. They are of two types: physical and chemical. Physical gels may be colloidal suspensions, which aggregate together and adhere to form the gel. These are not held together tightly and can be dispersed again by agitation or mixing. Another type of physical gel is where individual long-chain polymers aggregate either through helix formation or by crystallization. A prime example of this is collagen – a protein that in its denatured state is called gelatin. With a temperature stimulus, individual long-chain gelatin polymers aggregate through helix formation (due to hydrogen bonding or electrostatic attraction). Less rapid aggregation, with the help of fibroblast cells, allows tighter aggregation in the form of a triple helix, leading to strong fibers of collagen.

Chemical gels, on the other hand, are aggregated or gelated with the help of cross-linking reactions between the long polymer chains. An example of this is polyacrylamide, which has been used as a matrix for gel electrophoresis of proteins. Typically, a 1–10% aqueous solution of acrylamide monomer with methylene bisacrylamide as a cross-linking agent at 5–20% of the monomer concentration, a persulfate free-radical catalyst, and a tertiary amine (usually TEMED, tetramethylethylenediamine) as an activator (Heskins and Guillet 1968) is used for gel formation. The activator causes decomposition of the catalyst to produce free radicals that initiate the polymerization. The cross-link density then depends on the ratio of cross-linking agent to monomer. Gels in which water is the solvent are called hydrogels. Any water-soluble polymer can be prepared as a gel by carrying out the polymerization in the presence of a cross-linking agent in an aqueous solution of the polymer (Dusek 1993).

1.04.3.4.1 Responsive gels

Since these gels have a high solvent content, their properties depend upon the balance of polymer–polymer versus polymer–water interactions. Structural change, size variation, and mechanical actuation can be accomplished by slight changes in the polymer–solvent interaction properties initiated by an external stimulus. These kinds of gels are termed as stimuli-responsive hydrogels, and temperature sensitivity is the most common kind where the polymer gel network displays a temperature-sensitive phase behavior at the swelling solvent (Pelton 2000). A linear polymer that displays cloud point behavior (i.e. loss of transparency in a solution of long change hydrocarbon due to aggregation and particulate formation) when heated can be cross-linked to give a temperature-sensitive gel network. A common example is poly(*N*-isopropylacrylamide) (polyNIPAM) or related copolymers. The structure of *N*-isopropylacrylamide (NIPAM) is shown in **Figure 15**. When this gel is heated, it becomes hydrophobic and shrinks by expelling water above a temperature termed as lower critical solution temperature (LCST). At their LCST (32°C) polyNIPAM chains undergo a rapid and reversible phase transition from extended hydrated chains below the LCST to collapsed hydrophobic coils above the LCST (Heskins and Guillet 1968). At room temperature, the gels have high water content, low refractive index difference with water, and few electrically charged groups at the chain ends. At elevated temperatures, the gel volume is several times smaller. The density of electrically charged groups is higher and the refractive index difference with water is greater (Pelton 2000).

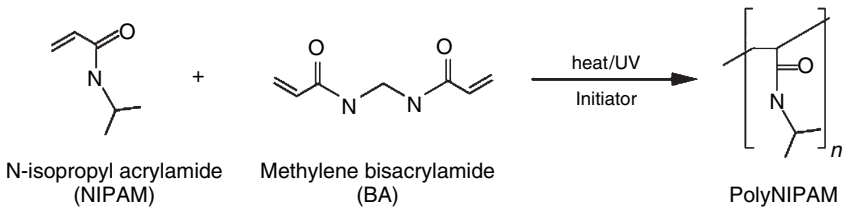


Figure 15 Chemical structure of NIPAM.

MEMS devices utilize transduction principles where a signal in one domain (biological, chemical, mechanical, or fluidic) is converted into another domain. Hydrogels provide a natural phenomenon that can be utilized in microfluidics for intelligent transduction. They are capable of converting chemical signals such as pH to mechanical action such as expansion or contraction. Hydrogels can be tuned to be sensitive to a variety of environmental cues such as pH (Tanaka and Fillmore 1979), temperature (Tanaka *et al.* 1980), electric field (Shiga 1997), glucose concentration (Kataoka *et al.* 1998), and antigens (Miyata *et al.* 1999). The response time for the volume change is proportional to the square of the gel dimension (Gan and Lyon 2001) because diffusion of the stimuli into the bulk of the gel is the key criterion that affects its phase transition. Therefore, faster response can be achieved by reduction in size, making hydrogels attractive materials for microscale actuation.

1.04.3.4.2 Microfabrication

These gels can be formed by photoinitiation, particularly using UV light (Chen *et al.* 1998). This feature makes them suitable for microfabrication through photolithography. One technique for microfabrication involves the use of microchannels to provide spatial resolution (Beebe *et al.* 2000). In this process, microfluidic channels are first constructed using a PDMS elastomer through a casting process as described in Section 1.04.3.2. This method can also be applied to microchannels made of other materials but with sufficient optical transparency. The microchannel is filled with a mixture of monomers and a photoinitiator and allowed to settle to equilibrium. This mixture is irradiated with UV light through a photomask to define the regions in the microchannel for polymerization (Beebe *et al.* 2000). The photopolymerizable liquid consists of acrylic acid and 2-hydroxyethyl methacrylate (in a 1:4 molar ratio), ethylene glycol dimethacrylate (1 wt%), and a photoinitiator (3 wt%) (Beebe *et al.* 2000). Polymerization times vary depending on the light intensity and the photoinitiator and monomer mixture, and is less than 20 s when Irgacure 651 is used as the photoinitiator. When the polymerization is completed, the channel is flushed with water to remove the unpolymerized liquid. This method allows pH-responsive hydrogels of different shapes and sizes to be integrated directly into microfluidic systems (Beebe *et al.* 2000).

1.04.3.4.3 Glucose-responsive devices

In another application, hydrogels responsive to glucose concentration are fabricated using a casting-based approach (Baldi *et al.* 2003) to actuate microvalves. Glucose-responsive hydrogels are synthesized at room temperature in a 200 mg ml^{-1} aqueous solution containing a feed of 18.4 mol% 3-methacrylamidophenylboronic acid, 81.5 mol% acrylamide, and 0.1 mol% methylene bisacrylamide (bis) solution, using ammonium persulfate and TEMED as initiators (Baldi *et al.* 2003). Microfabrication of this gel is done by fabricating a microfabrication mold on top of a membrane as shown in **Figure 16**. The monomer mixture is poured into the cavity, the top layer of which is made of a Teflon sheet. The gel is polymerized by thermal initiation. Then the Teflon sheet is removed and the porous membrane is placed for glucose diffusion. The expansion of the gel in response to glucose concentration regulates the flow in the microchannel. This gel, apart from being sensitive to glucose, is also sensitive to pH, demonstrating a 100% volumetric expansion for pH variation from 7 to 11 and a 30% volumetric expansion for glucose variation from 0 to 100 mM at a physiological pH of 7.4. As noted earlier, the response time of the gel is dependent on its size and valve opening times as low as 7 min have been observed with 30-μm hydrogel layers (Baldi *et al.* 2003).

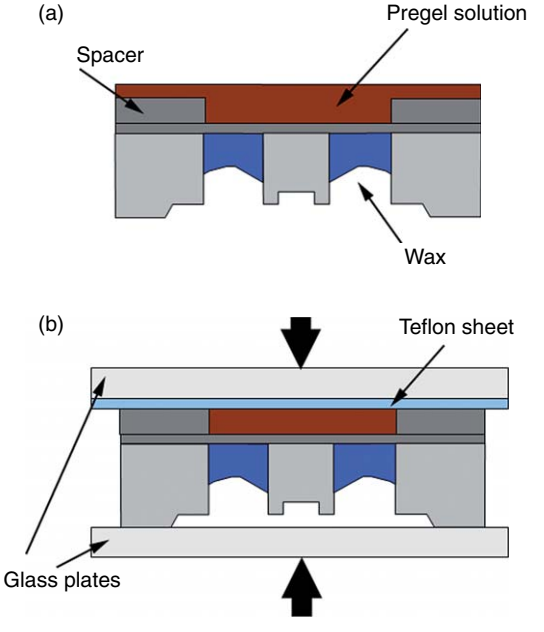


Figure 16 Microfabrication of hydrogel actuator. Reproduced with permission from Baldi A, Gu Y, Loftness P E, Siegel R A, Ziaie B 2003 A hydrogel-actuated environmentally sensitive microvalve for active flow control. *J. MEMS* **12**, 613–21, © 2003 IEEE.

1.04.3.4.4 Thermally responsive devices

In an interesting photoresist-like implementation, a solution of NIPAM, allyl methacrylate, and azobisisobutyronitrile (photoinitiator) in tetrahydrofuran (THF) solution is spin-cast to 1-μm thickness and photolithographically defined using optical photolithography (Mutlu *et al.* 2003). Higher thicknesses are obtained by casting the prepolymer mixture in a mold and heating the solution at 60°C for 10 min. This reduces the solvent, increases the viscosity, and allows the processing of thicker films. Photolithography and development in acetone produce microstructures with 50-μm resolution (Mutlu *et al.* 2003). Adhesion is promoted by dip-coating and by incubation of the substrate in a solution of 3-(trimethoxysilyl)-propyl methacrylate for 1 day. An in-line microvalve has been fabricated around this hydrogel using parylene microchannels constructed as described in Section 1.04.1. A 2.5-fold increase in polymer volume in water is observed when the water temperature decreases from 25°C to 10°C. Response time is reduced to 7 s primarily due to microfabrication of the hydrogel (Mutlu *et al.* 2003).

Hydrogels have been incorporated into the pores of macroporous polymers to provide both structural stability and thermal responsiveness. The macroporous polymer monolith does not change shape and hence provides overall structural stability, while the hydrogel inside the pore undergoes volumetric change, thereby changing the resistance to flow through the porous material (Yu *et al.* 2003). Response times of 4–5 s are observed with a pressure resistance of 200 psi in the closed state. The LCST temperature at which these phase changes occur has also been adjusted to be between 35°C and 74°C with the same characteristics by varying the proportions of the monovinyl monomers in the polymerization mixture (Lu *et al.* 2003).

1.04.3.5 Polyimide

Polyimide has also been used as a structural material for the construction of microfluidic devices (Goll *et al.* 1996, Metz *et al.* 2001). Commercially available polyimide resin has aromatic heterocyclic chains similar to the structure shown in **Figure 17**. They consist of alternating carbonyl (C=O) groups that act as electron acceptors and nitrogen (N) groups that act as electron donors. This charge transfer complex stabilizes the chain and is responsible for its chemical inertness. This kind of stabilization also happens between the chains and causes structural rigidity and high strength. Polyimide synthesis is a two-step process (**Figure 17**) involving a reaction between a dianhydride and a diamine under ambient conditions in a dipolar aprotic solvent such as *N*,*N*-dimethylacetamide (DMAc) or *N*-methylpyrrolidione (NMP) to yield the corresponding polyamic acid. The acid undergoes cyclization to form the final polyimide.

Photosensitivity can be introduced into the synthesis in two ways, with both depending on prevention of cyclization and hence formation of the polyimide backbone. The first method is to attach a photosensitive group to the carboxylic acid groups and prevent cyclization. The photosensitive group detaches from the polyamic acid upon light excitation. The other method is to use an acid–ion linkage between a photosensitive group and the polyamic acid, which can be broken using light (Hiramoto 1990).

1.04.3.5.1 Microfabrication

For applications in biosciences, polyimides offer advantages over other polymers due to their excellent chemical and thermal stability, low water uptake, and good biocompatibility (Richardson *et al.* 1993). In microfluidics, the initial application of polyimide is as membrane materials for microvalves (Goll *et al.* 1996) and pumps (Goll *et al.* 1997). These studies used commercially available thin polyimide membranes to separate the actuation chamber from the working chamber. Later, photosensitive polyimide (PI-2723, a negative tone organic polymer) was used as a structural material for microchannel construction (Glasgow *et al.* 1999). A thin polyimide layer is spun, prebaked, and patterned on a substrate using optical lithography with the desired geometry and vent channels, but not cured. Next, a thin layer of solvent with dissolved precursor is used to coat a soft-baked layer polyimide on a second substrate. The two halves are placed in contact and cured. Channels with a width of 50–1000 μm and a depth of 3–30 μm have been fabricated for chemical analysis and heat transfer devices (Glasgow *et al.* 1999). Significant stress develops in the structure due to solvent evaporation during the final curing.

An adapted lamination technique is used to produce microchannels with embedded electrodes without the problem of solvent evaporation (Metz *et al.* 2001). The process uses a photosensitive polyimide PI-2732, which is microstructured using optical lithography, and a nonphotosensitive polyimide PI-2611, which is microstructured using dry

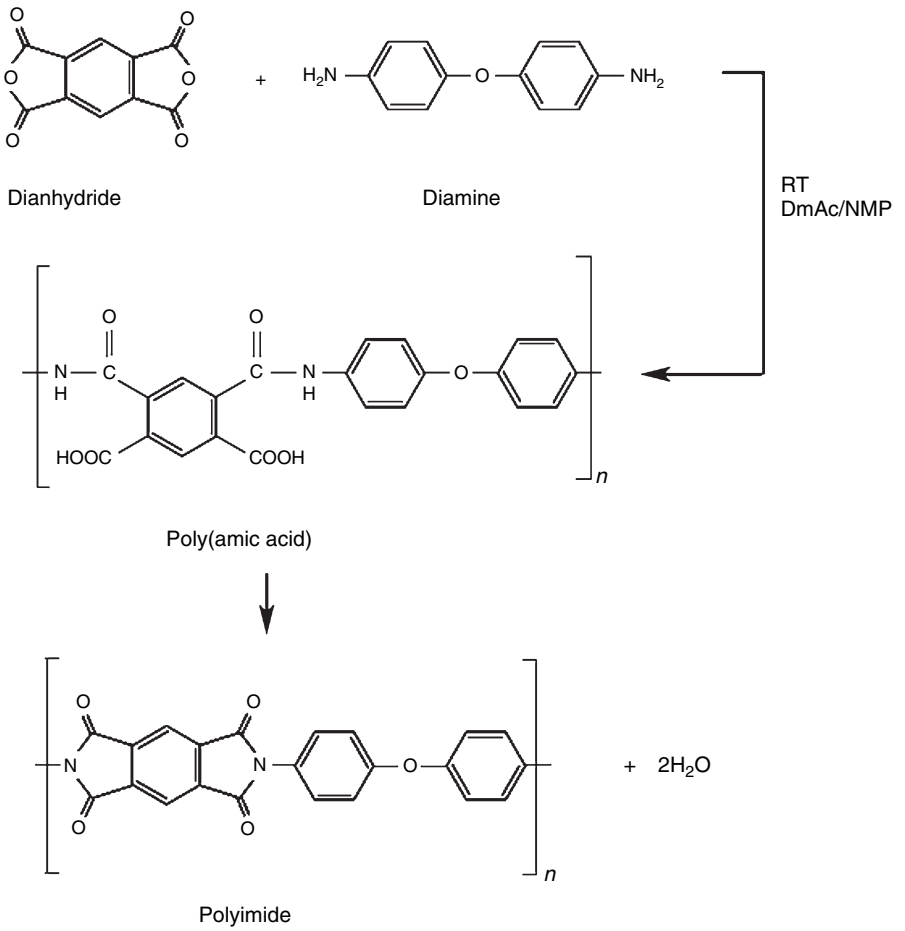


Figure 17 Chemical structure of polyimide.

etching with oxygen plasma with silicon dioxide mask. The process flow begins with the spinning of a 5- to 20-μm layer of polyimide and its curing at 350°C for 1 h. This layer forms the base of the microchannel. A layer of Ti/Pt (50/200 nm) is sputtered and patterned using a positive photoresist. This forms the electrodes inside the microchannels. A second layer of polyimide (5–20 μm), which forms the sidewalls, is spun and partially cured at 100–150°C for 1 h. This layer not only insulates the electrodes but also provides the structure for microchannels. Lamination is done with a nonphotosensitive polyimide layer whose surface is treated with *n*-methyl-2-pyrrolidone (NMP) (swelling agent). The swelling agent causes the partially cured second layer to swell, enabling higher interdiffusion with the lamination layer. A thin Mylar foil (polyimide) is spin-coated with photosensitive polyimide, partially cured, flipped over, and bonded to the substrate by lamination. This layer forms the top of the microchannel. The Mylar layer is peeled off and then the microstructure is laminated again (Metz *et al.* 2001). No interface was observed between the open channel prestructures and the laminated top layer, which indicates good channel sealing as shown in **Figure 18**. The imidization (curing) results in the shrinkage of the polyimide (30–40%), which decreases the height of the film and degrades the sharpness of the sidewalls. The channels thus fabricated withstand pressures as high as 19 bar (Metz *et al.* 2001).

1.04.3.6 Macroporous Polymers

Macroporous polymers were discovered in the 1950s (Abrams 1964; Corte and Meier 1971, Meitzner and Olin 1980). These polymers have a fixed pore size, narrow pore size distribution, extensive cross-linkings, and are structurally stable with their porous structure persisting even in the dry state. The porous

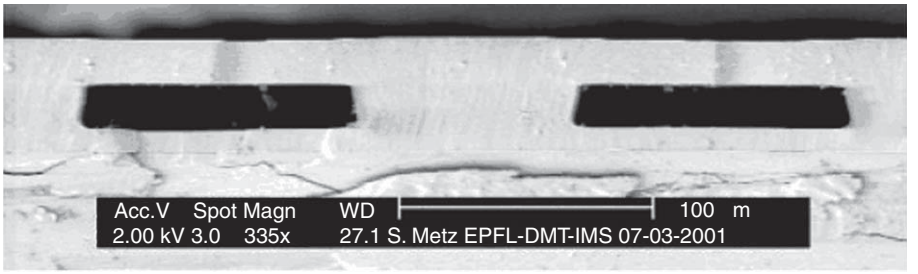


Figure 18 SEM image of microchannels constructed of polyimide. Reproduced with permission from Metz S, Holzer R, Renaud P 2001 Polyimide-based microfluidic devices. *Lab Chip* **1**, 29–34, © 2001 Royal Society of Chemistry.

nature gives them a high surface-to-volume ratio, which is desired in case of catalytic reaction, surface adsorption, chromatography, drug delivery, and electroosmosis.

These materials were initially produced in bead form for applications such as chromatographic packings (Brooks 1990, Yuan *et al.* 1991). The control of pore size and its narrow distribution is accomplished using inert diluents (also called porogens), which phase-separate during polymerization (Svec and Frechet 1995a, Viklund *et al.* 1996). They have also been demonstrated to be effective as monolith support for catalytic reactions (Unger 1990). The fabrication of these monoliths in molded fashion in various tube sizes and tube materials has been discussed in Peters *et al.* (1997a), Svec and Frechet (1995b, c), and Wang *et al.* (1993). These polymers have been used in high-throughput bioreactors with immobilization of various catalytic enzymes on the pores (Petro *et al.* 1996), and now find applications in drug delivery, membranes, and encapsulation of bioreactors (Dusek 1993). The surface properties of these pores formed can be changed from hydrophobic to hydrophilic by addition of grafted side chains (Wang *et al.* 1995). The variation of water permeation in the porous monolith can be controlled with the addition of polyNIPAM side chains (Peters *et al.* 1997b).

1.04.3.6.1 Polymerization

These monoliths can be structured to any given shape by casting the monomeric solution in the mold and by *in situ* polymerization. Microfabrication of these monoliths, using this technique, has been done by using optical lithographic techniques (Mutlu *et al.* 2002). The monomeric mixture consists of ethylene dimethacrylate (EDMA), butyl methacrylate (BMA), and 2-acrylamide-2-methyl-1-propanesulfonic acid (AMPS) as the main monomers, azobisisobutyronitrile (AIBN) as the polymerization initiator, and 1-propanol and 1,4-butenediol as the porogenic solvents. Light initiates radical generation of AIBN and polymerization of the

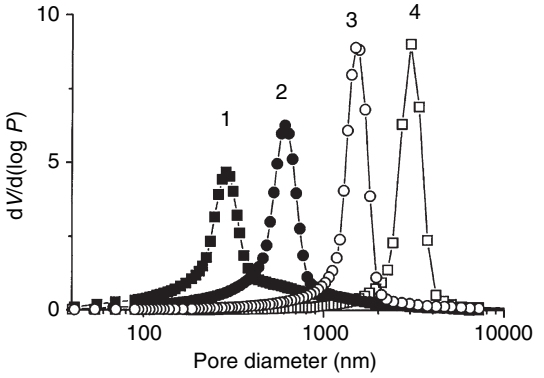


Figure 19 Pore size distribution in porous monolith. Reproduced with permission from Yu C, Svec F, Frechet J M J 2000 Towards stationary phases for chromatography on a microchip: Molded porous polymer monoliths prepared in capillaries by photo-initiated *in-situ* polymerization as separation media for electro-chromatography. *Electrophoresis* **21**, 120–7, © 2000 Wiley.

monomers. Upon polymerization, the polymer phase separates from the porogen phase and polymerizes around the trapped porogens, forming a porous structure. These polymers are usually hydrophobic; however, AMPS ionizes in most pH and hence provides hydrophilicity to the pore surface. The composition of the monomeric mixture determines the pore size and its narrow distribution (Yu *et al.* 2000) as shown in **Figure 19**.

1.04.3.6.2 Microfabrication

The main issues in adapting the technique for thin-film microstructure fabrication are the low viscosity of the monomeric mixture and the effect of oxygen on the polymerization reaction. Low viscosity restricts the use of spinning as a technique for obtaining the defined thickness of films. Very thin layers (~50 nm) and nonuniform distribution of polymer thickness are obtained by spinning. Furthermore, thin layers of the monomeric solution exposed to

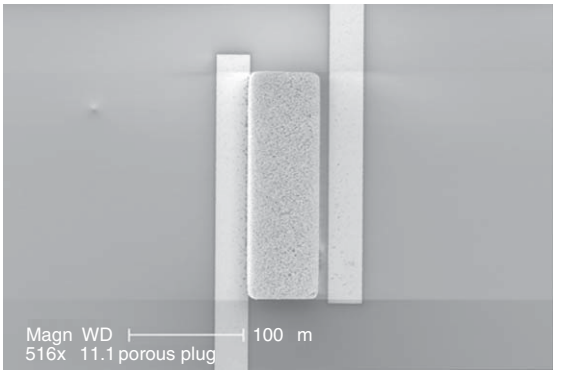


Figure 20 SEM image of a microfabricated porous monolith. Reproduced with permission from Mutlu S, Yu C, Selvaganapathy P, Svec F, Mastrangelo C H, Frechet J M J 2002 Micromachined porous polymer for bubble free electroosmotic pump. *Proc. IEEE Conf. MEMS*, Las Vegas, NV, USA, pp. 19–23, © 2002 IEEE.

the atmosphere absorb oxygen, which inhibits polymerization. These issues are solved using a microstructured glass mold, which is etched to a depth of 20 μm (thickness of a thin-film layer) and is used to enclose the monomeric mixture (Mutlu *et al.* 2002). The substrate to which the polymer monolith is to be structured is used as the capping layer. Glass allows polymerization of the monomer using optical techniques. Differential adhesion is obtained by depositing a monolayer of γ-methacryloxytropyl trimethoxy silane as an adhesion layer on the substrate while coating glass mold with parylene (Mutlu *et al.* 2002). Low viscosity causes diffusion of polymers and loss of resolution. Hence flood polymerization and dry etching techniques are used for producing high-resolution structures as shown in **Figure 20**. These structures have been demonstrated to increase the pumping capacity of microelectroosmotic pumps (Mutlu *et al.* 2002) and have potential applications in microscale catalysis, bioreactors, preconcentrators, and filtration units.

1.04.3.7 Biodegradable Materials

Biodegradable materials have traditionally been used as scaffolds for tissue engineering applications (Freed *et al.* 1994), controlled release mechanism for drug delivery (Langer 1990), and in sutures (Frazza and Schimdt 1971). Commonly used biodegradable polymers are polyglycolic acid (PGA), polylactic acid (PLDA), polycaprolactone (PCL), and poly(lactide-*co*-glycolide) (PLGA). The degradation times of these polymers range from 5 months to more than 24 months (Lu and Chen 2004). These polymers have low toxicity and minimal immune response, tunable degradation rates, and excellent mechanical properties suited for tissue engineering applications. The degradation of these polymers is done by hydrolysis of unstable linkage groups in the backbone, yielding degradation products that are capable of reabsorption by the body through metabolic pathways. For example, the degradation product of PGA is glycolic acid, which is a natural metabolite. Suitable unstable groups include esters, anhydrides, orthoesters, and amides. The structures of these polymers allow the manipulation of their degradation rates.

1.04.3.7.1 Polymerization

Synthesis of these polymers is carried out from their stable dimeric form. **Figure 21** shows the chemical structure and the synthesis of the most common of the biodegradable polymers. The dimer is cyclic and hence stabilized. In the presence of heat and an appropriate catalyst, the dimer undergoes ring-opening polymerization and forms linear chains. The chains take the shape of the mold in which the reaction occurs. Other techniques such as injection molding and extrusion have been used for shaping these polymers. An excellent review of the material properties of various classes of biodegradable materials can be found in Gunatillake and Adhikari (2003).

1.04.3.7.2 Microfabrication

A wide variety of microfabrication methods including imprinting or hot embossing (Lu and Chen 2004), soft lithography (Vozzi *et al.* 2003), direct writing (Vozzi *et al.* 2003), stereolithography (Cooke *et al.* 2002), and laser micromachining (Chen *et al.* 2003, Kancharla and Chen 2002) have also been used for structuring biodegradable materials. Photocurable biodegradable versions of PCL have been developed for stereolithographic applications (Matsuda *et al.* 2000). Microfabrication of these photopolymers using lithography has also been accomplished (Leclerca *et al.* 2004).

Compression molding combined with low-temperature bonding of layers has been used to demonstrate the feasibility of producing microstructures, specifically microchannels, using biodegradable materials (King *et al.* 2004). The application of these devices is in tissue engineering of microfluidic vascular networks, which then form the scaffold for seeding and growth of cells to form tissues with the existing blood supply networks. Biodegradable materials are attractive since they provide the initial structural

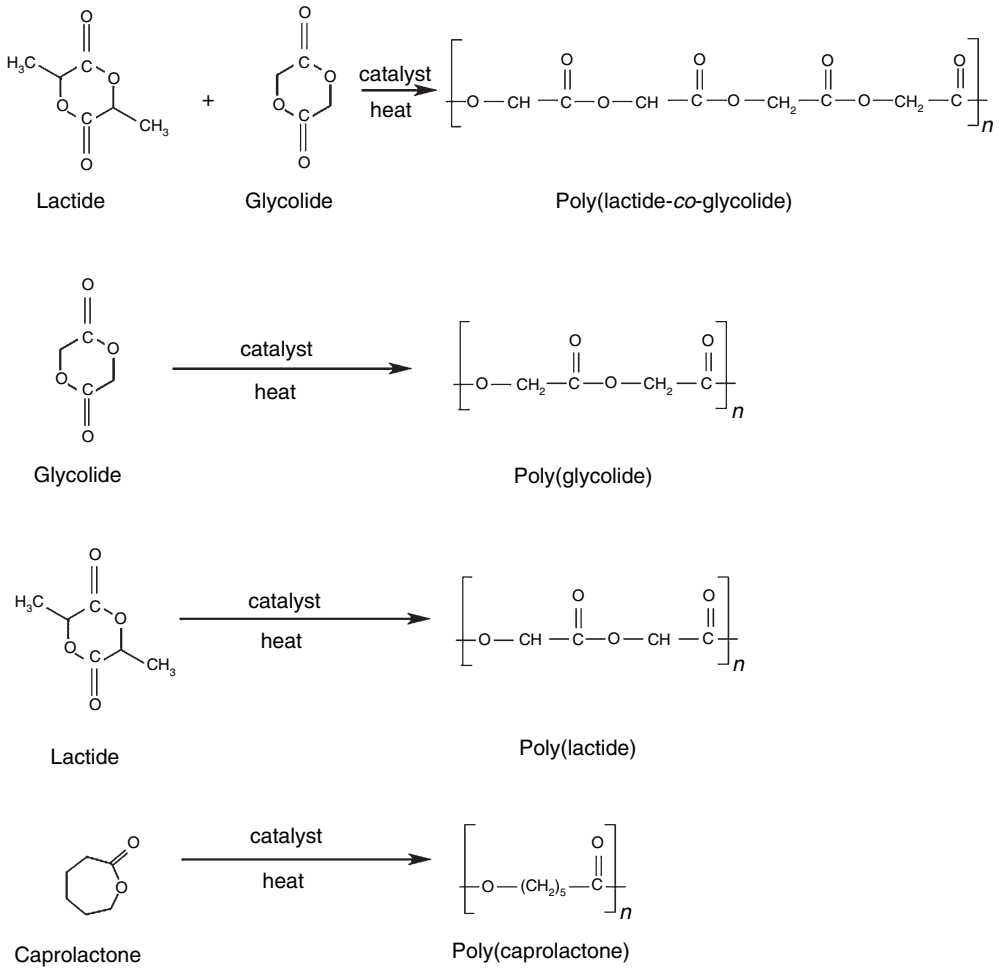


Figure 21 Polymerization process and chemical structure of biodegradable materials.

framework for growth and then wither away when the tissue is able to sustain itself. For molding, the masters containing the fluidic network are produced from PDMS using soft lithography techniques as shown in **Figure 22**. PDMS posts are then attached to the reservoirs in the mold. Another PDMS master with larger holes that are concentric with the attached post is assembled on the first master. PLGA pellets are placed between these two layers and heated to 110°C along with the application of compressive force of 100–500 lbs.

The PLGA becomes a viscous melt and conforms with the microstructure inside the mold cavity. Film thickness of 100–500 μm can be obtained by varying the compression time, the temperature, and the applied force. An unpatterned PLGA film is used to close the microchannel structure using low-temperature fusion bonding at 60°C for 30–60 min. Bonding at temperatures slightly higher than the glass transition temperature maintains structural integrity while still allowing interdiffusion in the areas of contact providing greater bond strength (King *et al.* 2004). Multichannel stacks could be created this way leading to 3D microchannel network as shown in **Figure 23**. A similar method was used to mold PCL material into microchambers for drug delivery (Armani and Liu 2000).

1.04.3.8 Paraffin

Paraffins are saturated hydrocarbon mixtures consisting of alkanes of varying chain lengths. They can be gas, liquid, or solid depending on the pressure and temperature conditions. They occur as simple straight chains or as branched chains. The general chemical formula for straight chain paraffins is

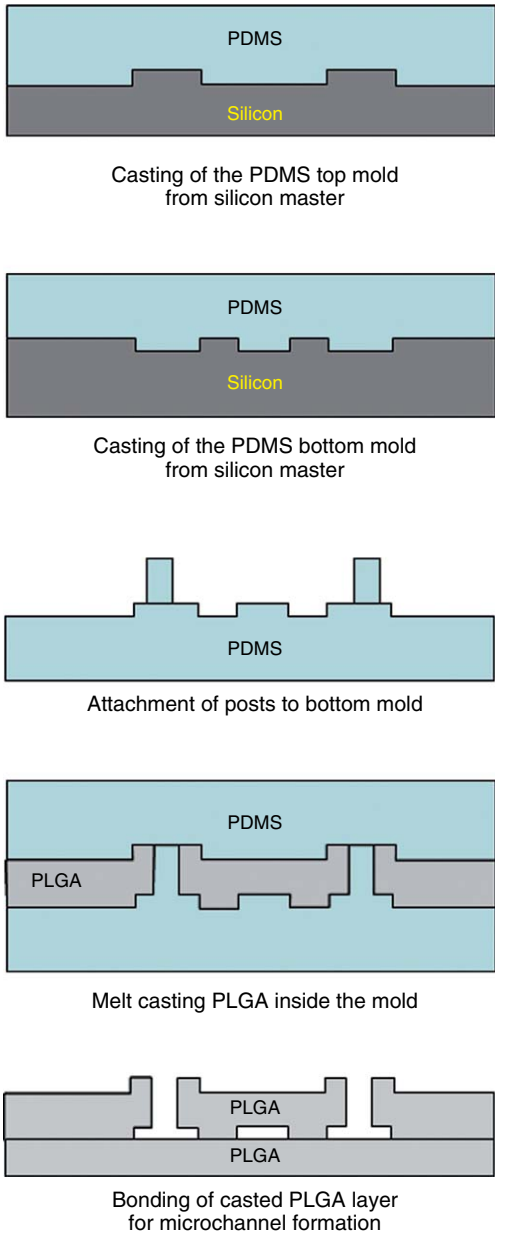


Figure 22 Schematic process flow for the construction of microchannel network in PLGA for creation of vascular networks.

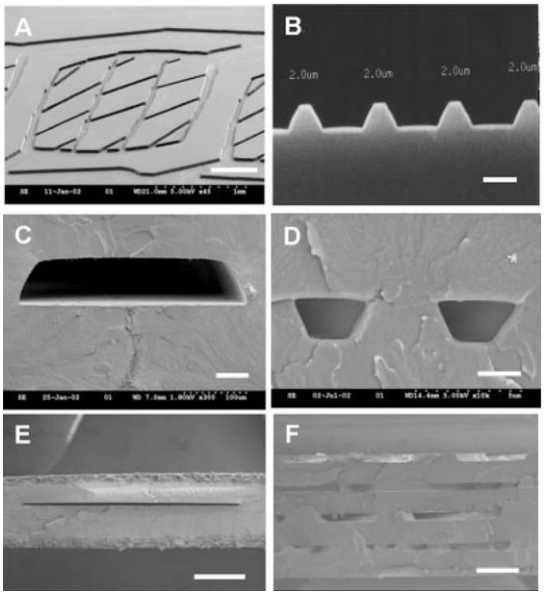


Figure 23 SEM cross-section of biodegradable multichannel network. (A) Complex branching of network of microchannels (scale bar 500 μm); (B) high resolution lines (scale bar 2 μm); (C) nominal microfluidic channels (scale bar 50 μm); (D) high resolution microchannels (scale bar 2 μm); (E) fully bonded high aspect ratio microchannels (scale bar 500 μm); and (F) multilayer microfluidic network (scale bar 200 μm). Reproduced with permission from King K R, Wang C, Kaazenpur M, Vacanti J P, Borenstein J T 2004 Biodegradable microfluidics. *Adv. Mater.* **16**, 2007–12, © 2004 Wiley.

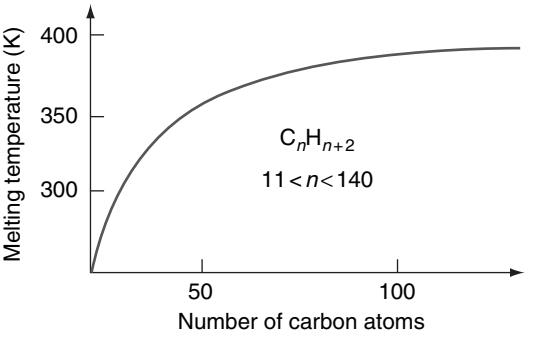


Figure 24 Melting point variation of paraffins with varying chain lengths. Reproduced with permission from Carlen E T, Mastrangelo C H 2002 Electrothermally activated paraffin microactuators. *J. MEMS* **11**, 165–74, © 2002 IEEE.

C_nH_{2n+2} and they have different melting temperatures depending on their chain lengths. **Figure 24** shows the range of melting temperatures for paraffins of different chain lengths.

1.04.3.8.1 Properties

Paraffins have low thermal conductivity, low electrical conductivity, and low chemical reactivity, and hence they are ideally suited for microactuator applications. They undergo phase change from solid to liquid, which is accompanied by volumetric expansion. Differences in volume of 5–40% are observed between phases for various paraffins (Carlen 2001). **Figure 25** shows the specific volumetric expansion of a particular paraffin wax under different pressure and temperature conditions. Paraffins are stable up to 250°C; therefore no

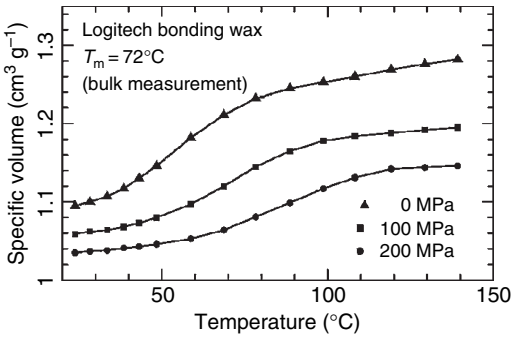


Figure 25 Volumetric expansion of paraffins under varying thermopneumatic conditions. Reproduced with permission from Carlen E T, Mastrangelo C H 2002 Electrothermally activated paraffin microactuators. *J. MEMS* **11**, 165–74, © 2002 IEEE.

boiling occurs even at high temperatures. Pressure loss during actuation is low due to its low viscosity in the liquid phase. Furthermore, liquid paraffins are nonpolar and hence do not mix with the polar liquids such as water, which are commonly used in microfluidic applications. Paraffins are also stable through numerous phase change cycles and their properties remain constant over long periods of time.

1.04.3.8.2 Microfabrication

Phase change actuation has the highest actuation power among all the actuation methods used in microactuators. They provide both high displacements and forces but their integration into microsystems has been cumbersome (Bergstrom *et al.* 1995). Microfabrication of paraffins has enabled ease of integration and development of high-force actuators in the microscale (Carlen and Mastrangelo 2002). Thin films of paraffins have been deposited using a thermal evaporation process. They are easy to evaporate due to their high vapor pressures. Deposition conditions of 5 µtorr and temperature of 150°C provided a deposition rate of 100 nm min^{-1}. Surface roughness of $\leq$42 nm has been measured (Carlen 2001). Adhesion of the deposited film to the substrate was enhanced by silanation and oxygen plasma treatment. Micropatterning of paraffins follows an elaborate procedure because of its dissolution and loss of feature in solvents such as acetone and developer solutions. First, the thin film is encapsulated by a thin layer of parylene. A thin layer of Au is deposited using a cooled-stage sputtering station. The Au layer is patterned using optical lithography and is used as a mask for reactive ion etching (RIE) of parylene/paraffin hybrid layer. RIE under 300 mtorr pressure with 25:75 sccm (CF_4:O_2) at 70-W radio frequency (RF) provided an etch rate of 250–300 nm min^{-1}. It is critical to maintain the substrate at a low temperature in order to maintain the patterned feature on the wafer.

Microactuators consisting of a microheater with a patterned micropatch of paraffin enclosed in a parylene sealed chamber have been developed (Carlen and Mastrangelo 2000), exhibiting a deflection of 2.7 µm at 150 mW. Actuators were assembled into a valve format and a 6-bit microflow controller was developed, which had flow control from 1 to 2 sccm using 5 V, 50 mW power (Carlen and Mastrangelo 2000).

1.04.4 Conclusion

The use of polymers in MEMS and microfluidics has been expanding rapidly since 1996. A number of microfabrication methods, both conventional and nonconventional, have been developed to encompass a wide selection of polymers available. The main driving force is the application of these techniques toward construction of microfluidic devices for medical diagnostics, drug discovery, and analytical laboratory applications. The attractiveness of polymers is their low cost, variety, solvent and biocompatibility, and customizable surface properties. Single devices such as microchannel networks, pumps, and valves have been successfully constructed using the existing polymers and fabrication techniques. The focus of future research is in three main directions:

1. Development of microfabrication technologies for production of huge volumes of disposable devices at low cost.
2. Development of new polymers customized for specific applications in microfluidics.
3. Development of integrated systems involving multiple polymeric materials and fabrication technologies for complex functionality.

The use of polymers has moved MEMS and microfluidics away from its roots in the semiconductor microprocessing industry. Its widespread adoption points toward its popularity and long-term success.

References

Abrams I M 1964 Ion exchange resins, preparations thereof. *US Pat.* 3 122 514

Armani D K, Liu C 2000 Microfabrication technology for polycaprolactone, a biodegradable polymer. *J. Micromech. Microeng.* **10**, 80–4

Baldi A, Gu Y, Loftness P E, Siegel R A, Ziaie B 2003 A hydrogel-actuated environmentally sensitive microvalve for active flow control. *J. MEMS* **12**, 613–21

Becker H, Gartner C 2000 Polymer microfabrication methods for microfluidic analytical applications. *Electrophoresis* **21**, 12–26

Beebe D J, Moore J S, Bauer J M, Yu Q, Liu R H, Devadoss C, Jo B H 2000 Functional hydrogel structures for autonomous flow control inside microfluidic channels. *Nature* **404**, 588–90

Bell T E, Wise K D, Anderson D J 1997 Flexible micromachined electrode array for a cochlear prosthesis. *Int. Conf. Solid-State Sensors and Actuators Proceedings (Tranducers '97)*, Chicago, IL, USA, Vol. 2, pp. 1315–18

Bergstrom P L, Ji J, Liu Y L, Kaviany M, Wise K D 1995 Thermally driven phase change microactuation. *J. MEMS* **4**, 10–17

Bertsch A, Zissi S, Jezequel J Y, Crobel S, Andre J C 1997 Microstereolithography using a liquid crystal display as a dynamic mask generator. *Microsyst. Technol.* **3**, 42–7

Bertsch A, Lorenz H and Renaud P (1998) Combining microstereolithography and thick resist UV lithography for 3D microfabrication. *Proc. IEEE Conf. MEMS*, pp. 18–23

Bertsch A, Bernhard P and Renaud P 2001 Microstereolithography: Concepts and applications. *Proc. IEEE Conf. Emerging Technologies and Factory Automation*, pp. 289–98

Brahmasandra S N, Johnson B N, Webster J R, Burke D T, Mastrangelo C H, Burns M A 1998 On-chip DNA band detection in microfabricated separation systems. *Proc. SPIE Conf. Microfluidic Devices and Systems*, Santa Clara, CA, USA, Vol. 3515, pp. 242–51

Brook M 2000 *Silicon in Organic, Organometallic and Polymer Chemistry*. John Wiley and Sons, New York

Brooks D W 1990 Basic aspects and recent developments in suspension polymerization. *Macromol. Chem. Macromol. Symp.* **35/36**, 121–5

Carlen ET 2001 Electrothermally actuated polymer microvalves. Ph.D. thesis, University of Michigan

Carlen E T, Mastrangelo C H 2000 Paraffin actuated surface micromachined valves. *IEEE Conf. MEMS*, pp. 381–5

Carlen E T, Mastrangelo C H 2002 Electrothermally activated paraffin microactuators. *J. MEMS* **11**, 165–74

Chen G, Imanishi Y, Ito Y 1998 Photolithographic synthesis of hydrogels. *Macromolecules* **31**, 4379–81

Chen S C, Kancharla V, Lu Y 2003 Laser-based microscale patterning of biodegradable polymers for biomedical applications. *Int. J. Mater. Prod. Technol.* **18**, 457–68

Cooke M N, Fisher J P, Dean D, Rimnac C, Mikos A G 2002 Use of stereolithography to manufacture critical-sized 3D biodegradable scaffolds for bone ingrowth. *J. Biomed. Mater. Res. Part B Appl. Biomater.* **64B**, 65–9

Corte H, Meier E 1971 Cation exchanger with sponge structure. *US Pat.* **3** 586 646

Deng T, Wu H K, Brittain S T, Whitesides G M 2000 Prototyping of masks, masters, and stamps/molds for soft lithography using an office printer and photographic reduction. *Anal. Chem.* **72**, 3176–80

Despa M S, Kelly K W, Collier J R 1999 Injection molding of polymeric LIGA HARMs. *Microsyst. Technol.* **6**, 60–6

Dittrich P S, Manz A 2006 Lab-on-a-chip: Microfluidics in drug discovery. *Nat. Rev. Drug Discov.* **5**, 210–18

Duffy D C, McDonald J C, Schueller O J, Whitesides G M 1998 Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). *Anal. Chem.* **70**, 4974–84

Dusek K 1993 *Advances in Polymer Science. Responsive Gels: Volume Transitions II*. Springer-Verlag, Berlin

Effenhauser C S, Bruin G J, Paulus A, Ehrat M 1997 Integrated capillary electrophoresis on flexible silicone microdevices: Analysis of DNA restriction fragments and detection of single DNA molecules on microchips. *Anal. Chem.* **69**, pp. 3451–7

Efimenko K, Wallace W E, Genzer J 2002 Surface modification of Sylgard-184 poly(dimethyl siloxane) networks by ultraviolet and ultraviolet/ozone treatment. *J. Colloid Interf. Sci.* **254**, 306–15

Ehrfeld W, Gotz F, Munchmeyer D, Schelb W, Schmidt D 1988 LIGA process: Sensor construction techniques via X-ray lithography. In: *Solid-State Sensor and Actuator Workshop, 1988*, Technical Digest, IEEE, Hilton Head Island, SC, USA, pp. 1–4

Flaherty M 1995 Conformal polymer film protects circuits, stabilizes solder joints. *Proc. Int. Electronics Packaging Conf.,* Atlanta, GA, USA, pp. 675–9

Frazza E J, Schimdt E E 1971 A new absorbable suture. *J. Biomed. Mater. Res. Symp.* **1**, 43–58

Freed L E, Novakovic G V, Biron R J, Eagles D B, Lesnoy D C, Barlow S K, Langer R 1994 Biodegradable polymer scaffold for tissue engineering. *Nat. Biotechnol.* **12**, 689–93

Fu P F, Glover S, King R K, Lee C I, Pretzer M R, Tomalia M K 2003 Polypropylene–polysiloxane block copolymers via hydrosilylation of monovinylidene capped isotactic polypropylene. *ACS Polym. Prepr.* **44**, 1014–15

Gan D J, Lyon L A 2001 Tunable swelling kinetics in core-shell hydrogel nanoparticles. *J. Am. Chem. Soc.* **123**, 7511–17

Glasgow I K, Beebe D J, White V E 1999 Design rules for polyimide solvent bonding. *Sens. Mater.* **11**, 269–78

Goll C, Bacher W, Bustgens B, Maas D, Menz W, Schomburg W K 1996 Microvalves with bistable buckled polymer diaphragms. *J. Micromech. Microeng.* **6**, 77–9

Goll C, Bacher W, Büstgens B, Maas D, Ruprecht R, Schomburg W K 1997 An electrostatically actuated polymer microvalve equipped with a movable membrane electrode. *J. Micromech. Microeng.* **7**, 224–6

Gorham W F 1966 A new, general synthetic method for the preparation of linear poly-p-xylylenes. *J. Polym. Sci.* **4**, 3027–39

Guerin L J, Bossel M, Demierre M, Calmes S, Renaud P 1997 Simple and low cost fabrication of embedded microchannels by using a new thick film photoplastic. *Proc. Transducers '97,* Chicago, IL, USA, pp. 1419–22

Gunatillake P A, Adhikari R 2003 Biodegradable synthetic polymers for tissue engineering. *Eur. Cells Mater.* **5**, 1–16

Hanemann T, Heckele M, Piotter V 2000 Current status of micromolding technology. *Polym. News* **25**, 224–9

Harrison D J, Glavina P G, Manz A 1993 Towards miniaturized electrophoresis and chemical analysis systems on silicon: An alternative to chemical sensors. *Sens. Actuators B* **10**, 107–16

Heckele M, Schomburg W K 2004 Review on micro molding of thermoplastic polymers. *J. Micromech. Microeng.* **14**, R1–14

Heckele M, Bacher W, Muller K D 1998 Hot embossing – The molding technique for plastic microstructures. *Microsyst. Technol.* **4**, 122–4

Heskins M, Guillet J E 1968 Solution properties of poly(N-isopropylacrylamide). *J. Macromol. Sci. Chem.* **8**, 1441–55

Heuschkel M O, Guerin L, Buisson B, Bertrand D, Renaud P 1998 Buried microchannels in photopolymer for delivering of solutions to neurons in a network. *Sens. Actuators B* **48**, 356–61

Hillborg H, Gedde U W 1998 Hydrophobicity recovery of polydimethylsiloxane after exposure to corona discharges. *Polym. Sci.* **39**, 1991–8

Hillborg H, Ankner J F, Gedde U W, Smith G D, Yasuda H K, Wikstrom K 2000 Crosslinked polydimethylsiloxane exposed

to oxygen plasma studied by neutron reflectometry and other surface specific techniques. *Polymer* **41**, 6851–63
Hiramoto H 1990 Photosensitive polyimides. In *Advanced Electronic Packaging Materials. Proceedings*, MRS, Warrendale, PA, Vol. 167, pp. 87–97
Huck W T, Strook A D, Whitesides G M 2000 Synthesis of geometrically well defined: molecularly thin polymer films. *Angew. Chem. Int. Ed.* **39**, 1058–61
Hull C W 1986 Apparatus for production of three dimensional objects by stereolithography. *US Pat.* **4**, 575–330
Ikuta K, Hirowatari K 1993 Real 3-D microfabrication using stereolithography and metal molding. *Proc. IEEE Workshop MEMS,* Fort Lauderdale, FL, USA, pp. 42–7
Ikuta K, Ogata T, Tsubio M, Kojima S 1996 Development of mass productive micro stereolithography. *Proc. IEEE Conf. MEMS,* San Diego, CA, USA, pp. 301–6
Ito T, Soubue K, Ohya S 2002 Water glass bonding for micro total analysis systems. *Sens. Actuators B* **81**, 187–95
Jackman R J, Wilbur J, Whitesides G M 1995 Fabrication of submicron features on curved substrates by microcontact printing. *Science* **269**, 664–6
Jackman R J, Brittain S T, Adams A, Wu H, Prentiss M G, Whitesides S, Whitesides G M 1999 Three-dimensional metallic microstructures fabricated by soft lithography and microelectrodeposition. *Langmuir* **15**, 826–36
Jeon N L, Choni I S, Whitesides G M, Kim N Y, Laibinis P E, Harada Y, Finnie K R, Girolami G S, Nuzzo R G 1999 Patterned polymer growth on silicon surfaces using microcontact printing and surface initiated polymerization. *Appl. Phys. Lett.* **75**, 4201–3
Jian L, Desta Y M, Goettert J 2001 Multilevel microstructures and mold inserts fabricated with planar and oblique X-ray lithography of SU-8 negative photoresist. *Proc. SPIE* **4557**, 69–76
Jian L, Desta Y M, Goettert J, Bednarzik M, Loechel B, Yoonyoung J, Aigeldinger G, Singh V, Ahrens G, Gruetzner G, Ruhmann R, Degen R 2003 SU-8 based deep X-ray lithography/LIGA. *Micromach. Microfabricat. Process Technol. VIII, Proc. SPIE* **4979**, 394–401
Jo B H, Moorthy J, Beebe D J 2000 Polymer microfluidic valves, membranes and coatings. *Proc. Micro Total Analysis System,* Kluwer, Dordrecht, The Netherlands, pp. 335–8
Joon N L, Clem P G, Nuzzo R G, Payne D A 1995 Patterning of dielectric oxide thin layers by microcontact printing of self-assembled monolayers. *J. Mater. Res.* **10**, 2996–9
Kancharla V, Chen S C 2002 Fabrication of biodegradable polymeric micro-devices using laser micromachining. *Biomed. Microdev.* **4**, 105–9
Kataoka K, Miyazaki H, Bunya M, Okano T, Sakurai Y 1998 Totally synthetic polymer gels responding to external glucose concentration: Their preparation and application to on–off regulation of insulin release. *J. Am. Chem. Soc.* **120**, 12694–5
Kesari P, Xu T, Boland T 2005 Layer-by-layer printing of cells and its application to tissue engineering. *Mater. Res. Soc. Symp. Proc.* **845**, AA4.5.1–6
King K R, Wang C, Kaazenpur M, Vacanti J P, Borenstein J T 2004 Biodegradable microfluidics. *Adv. Mater.* **16**, 2007–12
Kumar A, Whitesides G M 1993 Features of gold having micrometer to centimeter dimensions can be formed through a combination of stamping with an elastomeric stamp and an alkanethiol 'ink' followed by chemical etching. *Appl. Phys. Lett.* **63**, 2002–4
Kumar A, Abbott N L, Kim E, Biebuyck H, Whitesides G M 1995 Patterned self-assembled monolayers and meso-scale phenomena. *Acc. Chem. Res.* **28**, 219–26
Lahann J, Klee D, Hocker H 1998 Chemical vapour deposition polymerization of substituted [2.2]paracyclophanes. *Macromol. Rapid Commun.* **19**, 441–4
Lahann J, Balcelis M, Hang L, Rodon T, Jensen K F, Langer R 2003 Reactive polymer coatings: A first step toward surface engineering of microfluidic devices. *Anal. Chem.* **75**, 2117–22
Langer R 1990 New methods of drug delivery. *Science* **249**, 1527–33
Larmer F, Schilp A 1996 Method of anisotropically etching silicon. *US Pat.* **5** 501 893
Larsson O, Ohman O, Billman A, Lundbladh L, Lindell C, Palmskog G 1997 Silicon based replication technology of 3D-microstructures by conventional CD-injection molding techniques. *Proc. Int. Conf. Solid-State Sensors and Actuators,* Chicago, IL, USA, pp 1415–18
Leclerca E, Furukawab K S, Miyatab F, Sakaic Y, Ushidab T, Fujiic T 2004 Fabrication of microstructures in photosensitive biodegradable polymers for tissue engineering applications. *Biomaterials* **25**, 4683–90
Lee K Y, LaBianca N, Zolgharnain S, Rishton S A, Gelorme J D, Shaw J M, Chang T H P 1995 Micromachining applications of a high resolution ultrathick photoresist. *J. Vac. Sci. Technol.* **B13**, 3012–16
Lemmo A V, Fisher J T, Geysen H M, Rose D J 1997 Characterization of an inkjet chemical microdispenser for combinatorial library synthesis. *Anal. Chem.* **69**, 543–51
Liang Z, Chiem N, Ocvirk G, Tang T, Fluri K, Harrison D J 1996 Microfabrication of a planar absorbance and fluorescence cell for integrated capillary electrophoresis devices. *Anal. Chem.* **68**, 1040–8
Lin A W, Wong C P 1992 Encapsulant for nonhermetic multichip packaging applications. *IEEE Trans. Components Packaging Manufact. Technol.* **15**, 510–18
Liu Y, Fanguy J C, Bledsoe J M, Henry C S 2000 Dynamic coating using polyelectrolyte multilayers for chemical control of electroosmotic flow in capillary electrophoresis microchips. *Anal. Chem.* **72**, 5939–44
Lorenz H, Despont M, Fahrni N, LaBianca N, Renaud P, Vettiger P 1997 SU-8: A low-cost negative resist for MEMS. *J. Micromech. Microeng.* **7**, 121–4
Lorenz H, Despont M, Fahrni N, Brugger J, Vettiger P, Renaud P 1998 High-aspect-ratio, ultra thick, negative-tone near-UV photoresist and its applications for MEMS. *Sens. Actuators A* **64**, 33–9
Lu Y, Chen S C 2004 Micro and nano-fabrication of biodegradable polymers for drug delivery. *Adv. Drug Deliv. Rev.* **56**, 1621–33
Luo Q, Mutlu S, Gianchandani Y B, Svec F, Frechet J M J 2003 Monolithic valves for microfluidic chips based on thermoresponsive polymer gels. *Electrophoresis* **24**, 3694–702
Madou M, Lee L J, Koelling K W, Daunert S, Lai S, Koh C G, Juang Y J, Yu L, Lu Y 2001 Design and fabrication of polymer microfluidic platforms for biomedical applications. *Proc. SPE 59th Annual Technical Conference,* Huston, TX, USA, pp. 2534–8
Makamba H, Hsieh Y, Sung W C, Chen S H 2005 Stable permanently hydrophilic protein-resistant thin-film coatings on poly(dimethylsiloxane) substrates by electrostatic self-assembly and chemical cross-linking. *Anal. Chem.* **77**, 3971–8
Man P F, Jones D K, Mastrangelo C H 1997 Microfluidic plastic capillaries on silicon substrates: A new inexpensive technology for bioanalysis chips. *Proc. IEEE Micro Electro Mechanical Systems (MEMS),* Nagoya, Japan, pp. 311–16
Manz A, Graber N, Widmer H M 1990 Miniaturized total chemical analysis systems: A novel concept for chemical sensing. *Sens. Actuators B* **1**, 244–8
Maruo S, Kawata S 1998 Two photon absorbed near infrared photopolymerization for 3D microfabrication. *J. MEMS* **7**, 411–15

Maruo S, Nakamura O, Kawata S 1997 3D microfabrication with two photon absorbed photopolymerization. *Opt. Lett.* **22**, 132–4

Matsuda T, Mizutani M, Arnold S C 2000 Molecular design of photocurable liquid biodegradable copolymers. Synthesis and photocuring characteristics. *Macromolecules* **33**, 795–800

Meitzner E F, Olin J A 1980 Polymerization processes and products therefrom. *US Pat.* **4** 224 415

Metz S, Holzer R, Renaud P 2001 Polyimide-based microfluidic devices. *Lab Chip* **1**, 29–34

Miyata T, Asami N, Uragami T 1999 A reversibly antigen-responsive hydrogel. *Nature* **399**, 766–9

Mutlu S, Yu C, Selvaganapathy P, Svec F, Mastrangelo C H, Frechet J M J 2002 Micromachined porous polymer for bubble free electroosmotic pump. *Proc. IEEE Conf. MEMS,* Las Vegas, NV, USA, pp. 19–23

Mutlu S, Yu C, Svec F, Mastrangelo C H, Frechet J M J, Gianchandani Y B 2003 A thermally responsive polymer microvalve without mechanical parts photopatterned in a parylene channel. *Proc. Conf. Solid State Sensors Actuators and Microsystems,* Boston, MA, USA, pp. 802–5

Nicholson W 1997 *The Chemistry of Polymers*, 2nd edn. Royal Society of Chemistry, Cambridge, UK

Noll W 1968 *Chemistry and Technology of Silicones*. Academic Press, New York

Olson R 1989 Parylene conformal coatings and their applications for electronics. *Proc. 19th Electrical Electronics Insulation Conference,* Electrical Electronics Insulation, Chicago, IL, USA, pp. 272–3

Papra A, Bernard D, Juncker N B, Larsen B M, Delamarche E 2001 Microfluidic networks made of poly(dimethylsiloxane), Si, and Au coated with polyethylene glycol for patterning proteins onto surfaces. *Langmuir,* **17**, 4090–5

Pelton R 2000 Temperature-sensitive aqueous microgels. *Adv. Colloid Interf. Sci.* **85**, 1–33

Peters E C, Svec F, Frechet J M J 1997a The preparation of large diameter molded porous polymer monoliths and the control of pore structure homogeneity. *Chem. Mater.* **9**, 1898–902

Peters E C, Svec F, Frechet J M J 1997b Thermally responsive rigid polymer monoliths. *Adv. Mater.* **9**, 630–3

Petro M, Svec F, Frechet J M J 1996 Immobilization of trypsin on molded macroporous poly(glycidyl methacrylate-co-ethylene dimethacrylate) rods and use of conjugates are bioreactors for affinity chromatography. *Biotechnol. Bioeng.* **49**, 355–63

Piotter V, Hanemann T, Ruprecht R, Thies A, Hausselt J 1997 New developments of process technologies for microfabrication. *Proc. SPIE* **3223**, 91–9

Piotter V, Bauer W, Benzler T, Emde A 2001 Injection molding of components for microsystems. *Microsyst. Technol.* **7**, 99–102

Qin D, Xia Y, Whitesides G M 1996 Rapid prototyping of complex structures with feature sizes larger than 20 m. *Adv. Mater.* **8**, 917–19

Qin D, Xia Y, Xu B, Yang H, Zhu C, Whitesides G M 1999 Fabrication of ordered two-dimensional arrays of micro and nanoparticles using patterned self-assembled monolayers as templates. *Adv. Mater.* **11**, 1433–7

Rich C A, Wise K D 1999 An 8-bit microflow controller using pneumatically actuated microvalves. *Proc. IEEE,* Micro Electro Mechanical Systems (MEMS), Orlando, FL, USA, pp. 130–4

Richardson R R, Miller J A, Reichert W M 1993 Polyimides as biomaterials: Preliminary biocompatibility testing. *Biomaterials* **14**, 627–35

Roman G T, Hlaus T, Bass K J, Seelhammer T G, Culbertson C T 2005 Sol–gel modified poly(dimethylsiloxane) microfluidic devices with high electroosmotic mobilities and hydrophilic channel wall characteristics. *Anal. Chem.* **77**, 1414–22

Ruprecht R, Hanemann T, Piotter V, Hausselt J 1995 Injection molding of LIGA and LIGA-similar microstructures using filled and unfilled thermo-plastics. *Proc. SPIE* **2639**, 146–57

Selbrede S C, Zucker M L 1997 Characterization of Parylene-N thin films for low-*k* VLSI applications. *Low-dielectric Constant Materials III*, Materials Society Symposium Proceedings, Santa Clara, CA: USA, Vol. 476, pp. 219–24

Selvaganapathy P, Carlen E T, Mastrangelo C H 2003 Recent progress in microfluidic devices for nucleic acid and antibody assays. *Proc. IEEE* **91**, 954–75

Sethu P, Mastrangelo C H 2004 Cast epoxy-based microfluidic systems and their application in biotechnology. *Sens. Actuators B Chem.* **98(2–3)**, 337–46

Shah R R, Merreceyes D, Husemann M, Rees I, Abbott N L, Hawker C J, Hedrick J L 2000 Using atom transfer radical polymerization to amplify monolayers of initiators patterned by microcontact printing into polymer brushes for pattern transfer. *Macromolecules* **33**, 597–605

Shiga T 1997 Deformation and viscoelastic behavior of polymer gels in electric fields. *Adv. Polym. Sci.* **134**, 131–63

St. John P M, Craighead H G 1996 Microcontact printing and pattern transfer using trichlorosilianes on oxide substrates. *Appl. Phys. Lett.* **68**, 1022–4

Sun H, Kawata S 2003 Two-photon laser precision microfabrication and its applications to micro–nano devices and systems. *J. Light Wave Technol.* **21**, 624–31

Svec F, Frechet J M J 1995a Temperature, a simple and efficient tool for the control of pore size in macroporous polymers. *Macromolecules* **28**, 7580–2

Svec F, Frechet J M J 1995b Molded rods of polymer for preparative separation of biological products. *Biotechnol. Bioeng.* **48**, 476–80

Svec F, Frechet J M J 1995c Modified poly(glycidyl methacrylate-co-ethylene dimethacrylate) continuous rod columns for preparative-scale ion-exchange chromatography of proteins. *J. Chromatogr.* **702**, 89–95

Tacito R D, Steinbruchel C 1996 Fine-line patterning of parylene-N by reactive ion etching for application as an interlayer dielectric. *J. Electrochem. Soc.* **143**, 1973–7

Takagi T, Nakajima N 1993 Photoforming applied to fine machining. *Proc. 4th Int. Symp. Micromachine and Human Science,* Nagoya, Japan, pp. 173–8

Tanaka T, Fillmore D J 1979 Kinetics of swelling of gels. *J. Chem. Phys.* **70**, 1214–18

Tanaka T, Fillmore D, Sun S T, Nishio I, Swislow G, Shah A 1980 Phase transitions in ionic gels. *Phys. Rev. Lett.* **45**, 1636–9

Tay F E H, van Kan J A, Watt F, Choong W O 2001 A novel micro-machining method for the fabrication of thick-film SU-8 embedded micro-channels. *J. Micromech. Microeng.* **11**, 27–32

Unger K (*ed.*) (1990) *Packings and Stationary Phases in Chromatographic Techniques* Dekker, New York

Unger M A, Chou H P, Thorsen T, Scherer A, Quake S R 2000 Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113–16

Viklund C, Svec F, Frechet J M J 1996 Monolithic molded porous materials with high flow characteristics for separations, catalysis, or solid phase chemistry: Control of porous properties during polymerization. *Chem. Mater.* **8**, 744–50

Vozzi G, Flaim C, Ahluwalia A, Bhatia S 2003 Fabrication of PLGA scaffolds using soft lithography and microsyringe deposition. *Biomaterials* **24**, 2533–40

Wallace DB 1989 A method of characteristics model of a drop on-demand ink-jet device using an integral method drop formation model. *ASME Publication 89-WA/FE-4*

Wang Q C, Svec F, Frechet J M J 1993 Macroporous polymer stationary phase rod as continuous separation medium for reverse phase chromatography. *Anal. Chem.* **65**, 2243–8

Wang Q C, Svec F, Frechet J M J 1995 Hydrophilization of porous polystyrene based continuous rod columns. *Anal. Chem.* **67**, 670–4

Wang D, Thomas S G, Wang K L, Xia Y, Whitesides G M 1997a Nanometer scale patterning and pattern transfer on amorphous Si, crystalline Si, and SiO_2 surfaces using self-assembled monolayers. *Appl. Phys. Lett.* **70**, 1593–5

Wang H Y, Foote R S, Jacobson S C, Schneibel J H, Ramsey J M 1997b Temperature bonding for microfabrication of chemical analysis devices. *Sens. Actuators B* **45**, 199–207

Wang X, Lin Q, Tai Y C 1999 Parylene micro check valve. *Proc. IEEE, Micro Electro Mechanical Systems (MEMS),* Orlando, FL, USA, pp. 177–82

Weber L, Ehrfeld W, Freimuth H, Lacher M, Lehr H, Pech B 1996 Micromolding–A powerful tool for the large scale production of precise microstructures. *Proc. SPIE* **2879**, 156–67

Webster J R, Mastrangelo C H 1997 Large-volume integrated capillary electrophoresis stage fabricated using micromachining of plastics on silicon substrates. *Proc. 1997 Int. Conf. Solid-State Sensors and Actuators (Tranducers '97),* Chicago, IL, USA, pp. 503–6

Webster J R, Burke D T, Burns M A, Mastrangelo C H 1998 An inexpensive plastic technology for microfabricated capillary electrophoresis chips. *Proc. Tas '98 Workshop,* Banff, Canada, pp. 249–52

Wong R, Hair M L, Croucher M D 1988 Sterically stabilized polymer colloids and their use as ink-jet inks. *J. Imaging Technol.* **14**, 129–31

Wong W S, Ready S, Matusiak R, White S D, Lu J P, Ho J, Street R A 2002 Amorphous silicon thin-film transistors and arrays fabricated by jet printing. *Appl. Phys. Lett.* **80**, 610–12

Xia G M, Whitesides G M 1998 Soft lithography. *Angew. Chem. Int. Ed.* **37**, 550–75

Xia Y, Zhao X M, Whitesides G M 1996 Pattern transfer: Self-assembled monolayers as ultrathin resists. *Microelectron. Eng.* **32**, 255–68

Xiao D, Zhang H, Wirth M 2002 Chemical modification of the surface of poly(dimethylsiloxane) by atom-transfer radical polymerization of acrylamide. *Langmuir* **18**, 9971–6

Yang X, Yang J M, Wang X Q, Meng E, Tai Y C, Ho C M 1998 Micromachined membrane particle filters. *Proc. IEEE Micro Electro Mechanical Systems (MEMS),* Heidelberg, Germany, pp. 137–42

Yu C, Svec F, Frechet J M J 2000 Towards stationary phases for chromatography on a microchip: Molded porous polymer monoliths prepared in capillaries by photo-initiated *in-situ* polymerization as separation media for electro-chromatography. *Electrophoresis* **21**, 120–7

Yu C, Mutlu S, Selvaganapathy P, Mastrangelo C H, Svec F, Frechet J M J 2003 Flow control valves for analytical microfluidic chips without mechanical parts based on thermally responsive monolithic polymers. *Anal. Chem.* **75**, 1958–61

Yuan H G, Kalfas G, Ray W H 1991 Suspension polymerization. *Macromol. Sci. Rev. Macromol. Chem. Phys.* **C31**, 215–19

Zhang X, Jiang X N, Sun C 1998 Microstereolithography for MEMS. *Proc. IEEE Conf. MEMS,* Heidelberg, Germany, pp. 3–9

Zhang X, Jiang X N, Sun C 1999 Microstereolithography of polymeric and ceramic microstructures. *Sens. Actuators A* **77**, 149–56

Biography

Ponnambalam Ravi Selvaganapathy received his BS degree in Chemical and Electrochemical Engineering from the Central Electrochemical Research Institute (CECRI), Karaikudi, India, in 1998. He received his MS and Ph.D. degrees in Electrical Engineering from the University of Michigan, Ann Arbor, MI, in 2001 and 2002, respectively. During his tenure at the University of Michigan, he worked on a variety of projects related to MEMS and bioMEMS processes and device development including electrochemical detection in capillary electrophoresis systems, electroosmotic pumping systems, prevention of bubble formation in electrically active microfluidic systems, thermal phase change polymers, development of microfabrication techniques for porous polymers in microsystems, development of a contact-free technique for pH measurement using the Kelvin probe, and investigation of microplasmas as a fluorescence excitation source for DNA. He then worked as a Postdoctoral Research Scientist with Sandia National Laboratories, Livermore, CA, until June 2004, in the area of integrated microfluidic systems. He is presently an Assistant Professor of Mechanical and Biomedical Engineering at McMaster University, Canada. His current research interests include developing active microfluidic components, materials, and fabrication technologies for a number of applications including, medical diagnostics, drug discovery, drug delivery, automated sampling and sensing, and microtools for biology.

1.05 Biocompatibility of Microsystems

Jürgen Rühe and Markus Biesalski, Department of Microsystems Engineering – IMTEK, University of Freiburg, Freiburg, Germany

1.05.1 Introduction

Recent progress in the field of microengineering has opened new horizons for the development of new types of medical devices (Goodsell 2004, Lee and Lee 2004, Santini *et al.* 1999, Staples *et al.* 2006, Voldman *et al.* 1999, Willner and Willner 2001). These devices monitor the well-being of a person for extended periods of time and detect any critical situations such as unusually elevated or reduced level of blood sugar, unhealthy high or low blood pressure, or an unexpected immune reaction. They respond to any severe problem they encounter either by communicating the problem to the patient or by directly alarming an emergency response unit over cellular phone. Smart sensor/microdevice couples will intervene in critical situations or in situations that require a planned intervention, directly through the controlled release of a drug (Santini *et al.* 1999, Staples *et al.* 2006). A little further in the future microdevices will allow the restoration of bodily functions after such functions have been impaired or even destroyed through an accident or by an illness. One of the expectations in this area is that a patient who has been paralyzed through an accident or even lost a limb completely can regain function of the impaired site, so that it behaves similarly to how it did before the accident. In this scenario the patient will just think about what he/she wants to do, and a microsystem that is coupled to the appropriate neuronal cells bridges the destroyed nerve area and excites the necessary muscles or finely moves a prosthesis (Clark 2006, Margalit *et al.* 2002, Rodriguez *et al.* 2000). In other cases, in a patient who has turned blind because of problems related to the eye itself, for example, due to retinal function loss, vision can be restored by directly feeding electrical signals from a small micro camera into the nerve connecting the dysfunctional eye and the brain. While the latter two examples might be a vision for the future and to date only initial steps have been made in this direction, other devices such as cochlear implants (Clark 2006), which improve the hearing of patients who have turned deaf but still have a functional hearing nerve, are already widely in use and have been implanted in a large number of patients (**Figure 1**).

Concepts for the construction of bacterial-scale, remote control submarines comprising nanocomponents, like the one featured in the 1966 science fiction movie, *Fantastic Voyage*, or, more recently with the same topic, in the movie *Innerspace* (1987), have been proposed. These submarines are injected into the

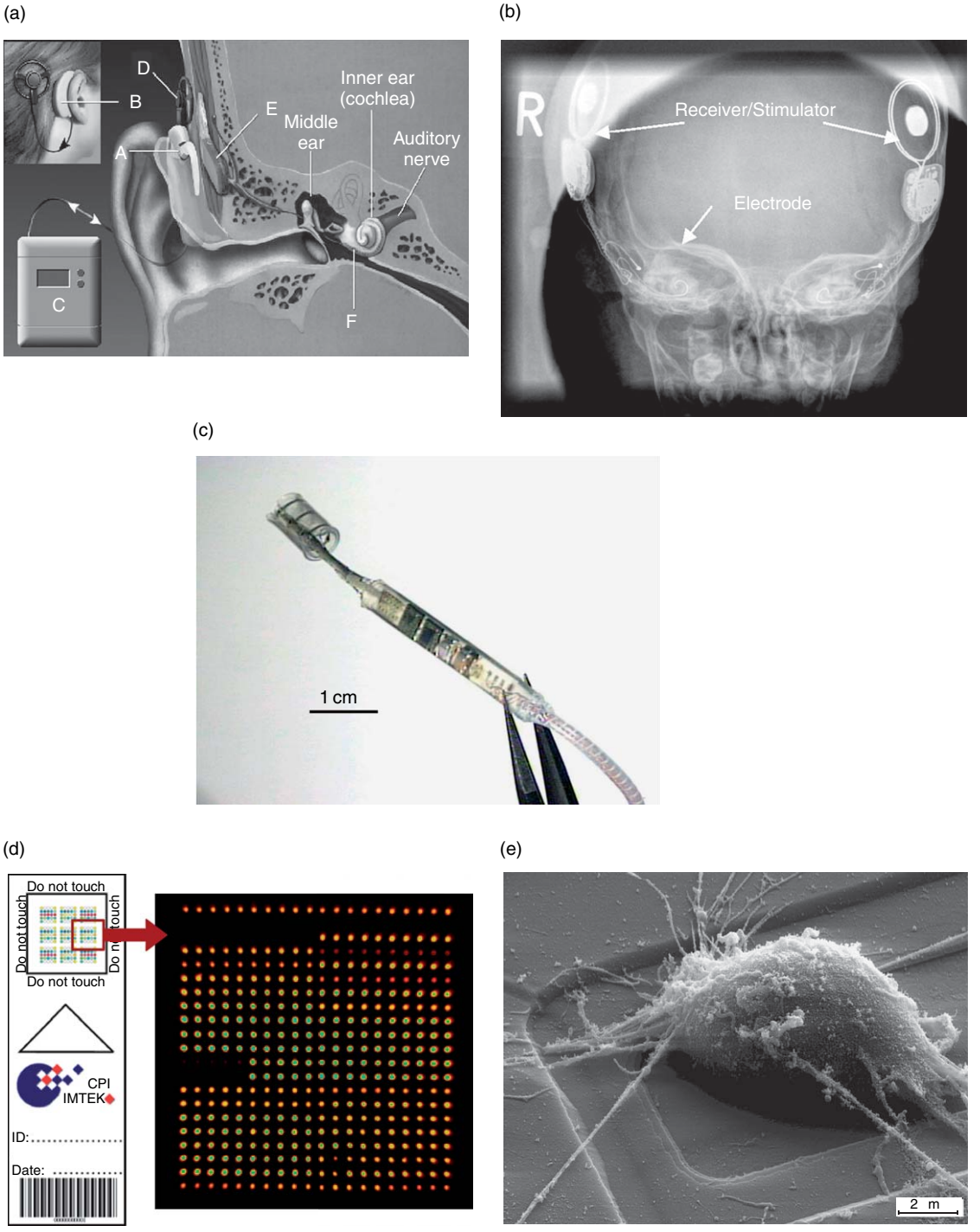


Figure 1 Examples and schematic illustrations of microdevices for medical applications: (a) The structure of the ear and a diagram of the multiple-channel cochlear implant. The components are as given: A, microphone; B, behind-the-ear speech processor; C, body-worn speech processor; D, transmitting aerial; E, receiver stimulator; F, electrode array. (Source: Clark G M 2006 The multiple-channel cochlear implant: The interface between sound and the central nervous system for hearing, speech, and language in deaf people-a personal perspective. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 791 [Review].) (b) X-ray micrograph of skull after bilateral implantation of a cochlea implant. (Picture courtesy: Dr. Aschendorff, Implant Center Freiburg, University Hospital Freiburg, Freiburg, Germany.) (c) Hybrid cuff electrode for neuronal stimulation (Picture courtesy: Prof. Stieglitz, University of Freiburg). (d) Photograph and fluorescence image of a DNA chip for massive parallel readout of DNA *ex corpore*. (e) Scanning electron microscopic image of a mouse neuron attached to the surface of a field-effect transistor (Picture courtesy: Prof. Stieglitz, University of Freiburg). (Figures reprinted in part with permission.)

bloodstream of patients and are transported to areas in the body that require examination and treatment. While such ideas belong to the realm of science fiction movies and will continue to do so for many years to come, great progress has undoubtedly been made in several diverse areas of microdevice-driven or microdevice-enhanced applications: Microdevices are already or will in the foreseeable future be employed in a large spectrum of biomedical applications such as minimal invasive diagnosis, monitoring and therapy, drug dosing, and implants, as well as in biochemical diagnostics. Some devices that have been developed and are currently in use are as follows:

- Chip-based bioanalytical devices (DNA, protein, cell chips) that allow high-throughput screening procedures and promise the development of low-cost diagnostic tools (Beier and Hoheisel 1999, Falconet *et al.* 2006, Ramsay 1998, Southern 1996, Zhu and Snyder 2003)
- Bioelectronics including active neurostimulation for the rehabilitation of patients who are suffering from paralysis (Egert *et al.* 1998, Fromherz *et al.* 1991, Scholl *et al.* 2000)
- Implantable sensors (wireless, remote) for readout and monitoring of important physiological parameters (Jain 2003, Kohli-Seth and Orpello 2003, Staples *et al.* 2006)
- Implantable drug dosing systems (Anderson 1994, Cleland *et al.* 2001, Santini *et al.* 1999)
- Smart catheters (Hefti *et al.* 1998, Singer *et al.* 1998)

This list is by no means exhaustive and is meant more as a general illustration of the potential of microdevices. It would be a considerable task to review in depth even the state of the art in one of these areas. For example the generation of biochips has progressed into a large field of scientific research with thousands of publications or patents worldwide. Bioelectronics has similarly developed into a large research area. An in-depth discussion of microdevices in biomedical applications is clearly beyond the goals of this chapter. Here we focus exclusively on the interactions between microdevices and biosystems such as the animal or human body.

If physical parameters of a patient such as blood pressure are measured outside of the body (*ex corpore*), the choice of materials involved in the generation of the microsystem is usually not critical. However, the situation changes drastically if the microsystem is brought into direct contact with the biosystem and the medical devices are implanted into the body of a patient. Arterial blood pressure is most accurately measured invasively by placing a cannula into a blood vessel and connecting it to an electronic pressure transducer. This invasive technique is regularly employed in human and veterinary intensive care medicine, in anesthesiology, and for research purposes, but it can be associated with complications such as thrombosis, infection, and bleeding. With respect to the latter, the choice of material becomes very critical. In the case of implanted devices, the interactions of the microdevice with the biological systems play a crucial role in its performance. During such intimate contact it becomes very important how the body responds to the implanted microsystem and vice versa. These mutual interactions are frequently summarized under the term biocompatibility. In many respects microdevices behave in a very similar way as macroscopic biomaterials, which have been extensively discussed in the literature. However, it should be noted that once a critical size scale is reached, the biosystem directly responds to the structure size. For example Whitesides *et al.* have shown that cells respond strongly to the size of the pattern that is generated at a substrate surface. In their experiments they patterned self-assembled monolayers (SAMs) through μ-contact printing (Kumar *et al.* 1995, Mrksich and Whitesides 1996). Whereas the overall chemistry was the same in all cases, the cells thrived very well on surfaces of one size scale while losing adhesion or even dying on surfaces where the same pattern was generated, but with a different size scale (Mrksich and Whitesides 1996).

A second example, which has already caused several thousand of deaths and forced the restoration or demolition of thousands of buildings worldwide, is the case of asbestos. Asbestos is a silicate mineral, which has wonderful physical and chemical properties and as such has been used as a widespread construction material. While in general silicates in other form and shapes are not toxic, the unfavorable aspect ratio of asbestos fibers has been identified as one of the major reasons for the development of lung cancer (Becklake 1976, Mossman 1989).

1.05.2 Biocompatibility and the Host Response

A useful definition of the term biocompatibility, which is widely used in the biomaterials community today, was given by Williams some twenty years ago (Williams 1986). He defined: "Biocompatibility is the ability of a material to perform with an appropriate

host response in a specific application." This definition is now widely accepted as it describes several important aspects of biocompatibility: Biocompatibility is not an absolute property of a material but must always be seen in the context of an application. While a material might perform very well in one application and one would call the biocompatibility excellent, it might utterly fail in another context. One of the reasons is that some microdevices such as a smart catheter are in contact with the patient for only hours or a few days at most, whereas other devices may be in place and perform their functions for many years, sometimes for the rest of the life of the patient. While a slow leaching out of some adverse/slightly toxic compound might be acceptable for short-term exposure, it might be completely detrimental in a long-term use of the same material. In addition, the long desired lifetime of the implant imposes much stricter restrictions on how critical wound healing and chronic responses become for the evaluation of the biocompatibility of the material.

Furthermore, the definition of biocompatibility given above, expressively points out that all biomaterials cause a host response of some kind. This is very important insofar as most materials cause a nonspecific and rather stereotypic response of the body against the foreign material. Such a response can occur at different levels: for example, a blood–material interaction, an acute inflammation, a foreign body reaction, formation of a fibrous capsule, or a chronic inflammation. An important question is whether the host response against the foreign material is acceptable – in other words, whether it works as intended and presents no harm to the patient, which could not be tolerated (appropriate host response). In the end, the key question is whether the benefits caused by the microdevice outweigh the adverse reactions.

If one looks more closely at the response of the host against the implant, one has to take different levels of host response into consideration:

(1) local response
(2) regional response
(3) global/systemic response

An unfavorable response at one of these levels might render the use of the whole microdevice not suitable. The first-level (local) response describes interactions of the microdevice directly with contacting blood or cells. Here the microdevice might lead to blood coagulation or trigger a direct immune response. The regional response describes how surrounding tissue that is not in direct contact, but close to it, responds to the biomaterial. A global/systemic response of the body can occur if toxic or allergy-inducing compounds are transported through blood or lymphatic vessels and cause reactions of the body against the implanted device sometimes even far away from the actual site of implantation. An example that is widely known and even frequently discussed in public is the so-called metal allergy (Grawkodger 1993, Landsdown 1995), where metal ions are transported through tissue or blood and cause an allergic reaction. While metal ions *per se* do not challenge the immune system of a person as they are way too small to do so, they can form metalloprotein complexes that can cause a strong allergic response locally and also further downstream, at locations where the metals are transported, and are being deposited. In some locations they can accumulate and cause a reaction despite their low concentration in the bloodstream (Landsdown 1995).

The term normal wound healing is commonly used to define a complex cascade of reactions that set in after an injury and that consist of events such as acute inflammation, formation of granulation tissue, and eventual new tissue formation. In brief, the normal healing typically starts with the flooding of blood from injured blood vessels into the injured area. The blood protein fibrinogen is converted into fibrin that promotes the adhesion and activation of platelets, which leads to the formation of a blood clot which closes the wound. Upon the release of chemical signaling molecules (e.g., cytokines), more cells infiltrate the injured area, among them are cells from the immune system (e.g., neutrophils of white blood cells). Macrophages resulting from differentiated monocytes remove foreign materials, bacteria, and dead cells. Upon stimulation, other cells, such as endothelial cells, and fibroblasts are then responsible for building up new blood vessels and extracellular matrix (ECM), thus generating a highly vascularized tissue (granulation tissue). Multiple factors such as the extent and location of the injury, and the loss of basement membrane structures govern the degree of restoration of normal tissue. In most cases the granulation tissue is remodeled into a scar tissue (fibrosis).

If a synthetic material is implanted into a living organism, this process is again accompanied by an injury and/or by a loss of tissue, regardless of the exact details of the medical device. However, now the organism responds in a different way to that outlined above, which is commonly known as foreign body reaction. In case of the presence of a biomaterial, the wound healing response is largely affected by the blood–material interactions, the formation of a

provisional matrix surrounding the implanted material, the extent and degree of cellular necrosis, as well as the extent and degree of the inflammatory response. For a detailed description of the host response of organisms to implanted synthetic materials, the reader can refer to several reviews by Anderson (Anderson 1993, 2001, Anderson *et al.* 1996). Here we will only summarize in brief the important events that take place (**Figure 2**) once a synthetic material (e.g., a prosthesis or a medical device) is implanted into a living organism.

Immediately after implantation, blood proteins, which are among the first components to arrive, adsorb at the surface of the material, typically in a nonspecific manner. Upon adsorption, many proteins change their conformation, i.e., they denature at the surface of the biomaterial. Note that nonspecific interaction of proteins with the biomaterial's interface, including the denaturation of proteins, does not occur in normal wound healing conditions. First, cells (e.g., neutrophils and platelets) approach and adhere to the biomaterial; subsequently, they sense the abnormal protein interface and react in a cascade of reactions in which cytokines signal other cells to infiltrate the area of the implant as well as signal for the differentiation of monocytes into macrophages that start attacking the foreign body (acute inflammatory response). As a biomaterial implant is most likely much larger than the recruited macrophages, chronic inflammation results in the fusion of macrophages into foreign body giant cells that are often present during the lifetime of an artificial implant. The final stage of this foreign body response, typically reached after 2–3 weeks, consists of the formation of a 50- to 200-μm-thick capsule that surrounds the biomaterial implant (Ratner and Bryant 2004). The fibrotic capsule consists of an avascular collagenous tissue, with the ultimate goal to cut off

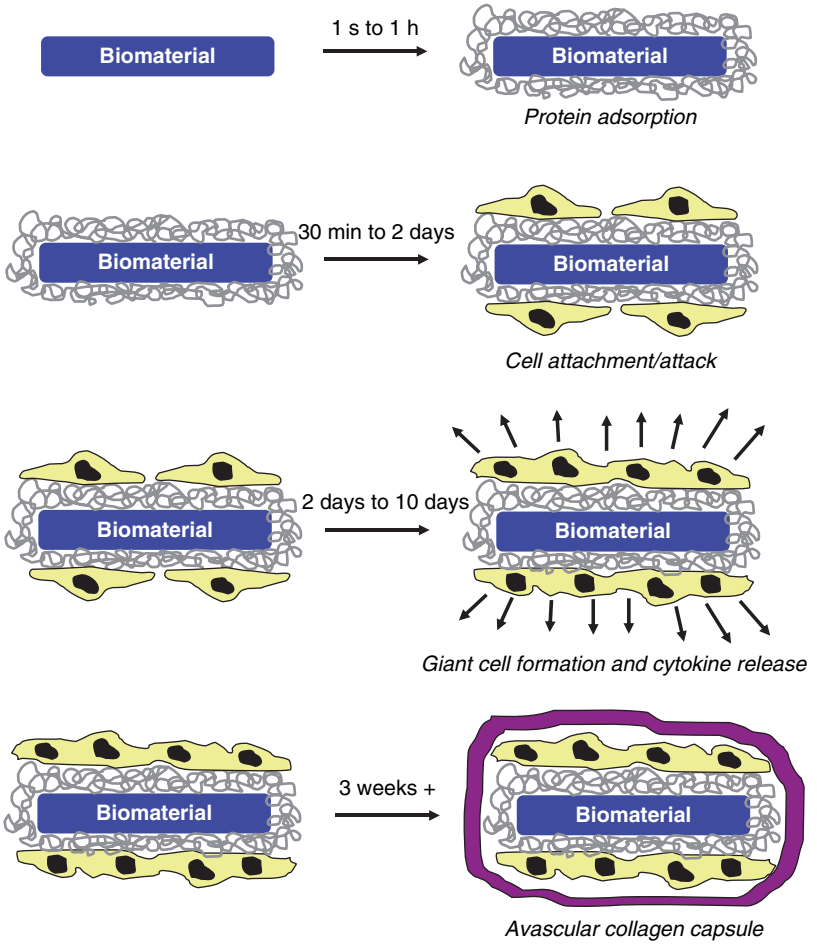


Figure 2 The foreign body reaction as illustrated here is the normal reaction by higher organisms to an implanted synthetic material (e.g., microdevice).

any support (e.g., nutritional supplements) and to wall-off the foreign body. Sometimes, chronic inflammation reduces after the formation of this capsule, and, in principle, the organism accepts the presence of the implant. However, in some cases strong inflammatory responses are still upregulated after weeks, or months, or even years of implantation, and the function of the injured tissue is significantly altered (Anderson *et al.* 1996).

Local responses determine the compatibility of a microsystem in contact with a biosystem. These responses in turn depend on a large spectrum of parameters. Some are directly related to the organism into which the microsystem is implanted, as not all information obtained from one species can be transferred directly to another species. The host response might further be influenced by the age, the gender, and the state of health of the patient and the tissue type and location.

Heart valve implants of porcine origin provide an example of the influence of the age of the patient on biocompatibility (Dahm *et al.* 2004, Schoen and Levy 1999). These materials can be implanted and behave rather well in the human body. They heal well and are next to carbon fiber-reinforced polymers, one of the standard materials for heart valve replacement. However, they have some tendency to become calcified, which causes strong wear to the valve and limits the lifetime of such a device in an adult person to about 10–15 years. As during this time the performance of the valve is very good, one would call the biocompatibility of the material satisfactory. However, when such a material is implanted into children, who have a much higher calcium level in the blood, calcification would occur within months, followed by mechanical destruction and failure. Thus as far as children are concerned, one would call the biocompatibility of the material low.

Some parameters controlling the biocompatibility are directly related to the device itself. Size and shape and mechanical properties such as fracture resistance and elasticity are some examples. Other parameters that are directly related to the surface properties of the implant are:

- Surface chemistry
- Surface roughness
- Surface charges
- Chemical stability/what kind of degradation products are set free
- Porosity
- Bleeding out of chemicals

Interactions of the microsystem and the host are not a one-way street in that the body reacts to the presence of the microsystem. The biosystem also represents an environment, which can sometimes be rather aggressive and highly challenging for microdevices. All biological systems consist of an aqueous solution, which contains a rather high concentration of ions and a wide spectrum of organic molecules. Thus degradation can and will to some extent occur and might change the properties of the microdevice significantly. Another problem that arises from the host response is a possible encapsulation of the microdevice. As already described in more detail above, a fibrous capsule starts to form some days after implantation. The formation of a fibrous capsule, however, may strongly interfere with the functionality of the implant (Brash and Horbett 1987, Ratner and Bryant 2004). If a sensor is implanted into the body, a reliable and accurate measurement of the analyte concentration requires the unrestricted diffusion of the analyte to the sensor at all times. If for example a glucose sensor is implanted into the body to measure the glucose concentration, it is essential that the glucose level at the sensor surface is closely correlated to that in the bloodstream. It is obvious that a barrier, such as the fibrotic collagenous bag surrounding the implanted device after 2–3 weeks, will influence the monitoring of the glucose concentration, as it resembles a barrier to the diffusion of the glucose molecules (**Figure 3**).

Why is the interaction of biological cells and biomolecules with polymers or other artificial materials for biomedical applications so complicated? One complication stems from the fact that biological environments are from a chemical point of view

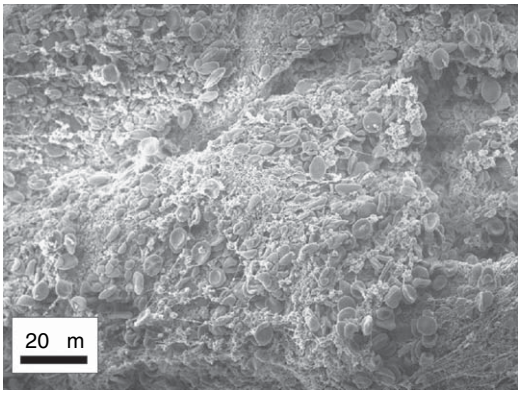


Figure 3 Scanning electron microscopic image of a fibrin network and thrombocytes on the surface of an artificial heart valve. (Picture courtesy: Dr. A. Schlitt, University Hospital, Mainz, Germany.)

rather complex mixtures of a wide spectrum of different compounds. Proteins, polysaccharides, lipids, and low molecular weight electrolytes are all simultaneously present in the biological systems and have all more or less strong interactions with the materials. However, the biocompatibility of implant materials and the success achieved in generating bioanalytical devices such as DNA or protein chips depend critically on the interaction of the surfaces in question with their respective environments.

These complex interactions lead frequently to the formation of more or less thick biofilms at the surface of the implanted material, which alter the surface strongly. As this process is mostly not desired, it is often called surface fouling. The formation of biofilms at the surface of the substrate is a complex process, which starts when the first proteins become adsorbed at the surface. This triggers a whole cascade of signaling events and leads to a strong adhesion of cells. Therefore controlling the adsorption of proteins to the surfaces of the material involved is one of the crucial factors in controlling the adhesion of cells and thus in enhancing the performance of the microsystem. This fact has guided a great quest for the synthesis of biomaterials with controlled surface properties. Before we review some important techniques for surface modification, we will briefly discuss some aspects of adhesion of living cells to surfaces.

1.05.3 Adhesion of Living Cells to Surfaces

For some devices the formation of a fibrotic capsule does not interfere with the functionality of the microsystem. For most applications, however, the encapsulation of the implanted device is very detrimental to its performance. In this case it is important to facilitate the restoration of a healthy tissue surrounding the implant.

In order to achieve this, conditions that allow living cells to attach to the biomaterial's interface, as well as the subsequent buildup of an intact (nondenatured), extra-cellular matrix (ECM), by the attached cells *in vivo*, need to be implemented. In order to design materials that provide the capability to resist nonspecific protein adsorption and to allow the directed attachment of living cells to these materials, it is necessary to understand the basic principles that underlay the adhesion of cells to other cells and to the ECM.

Our understanding of molecular mechanisms that underlay these phenomena has improved tremendously in the past decades. In almost all cases the cell–cell adhesion and cell–ECM adhesion are dominated by specific interactions formed between dedicated adhesion molecules found as transmembrane receptors embedded in the plasma membrane of cells. Five major classes of such receptors, so-called cell adhesion molecules (CAMs), have been identified (Alberts *et al.* 2002, Baszkin and Norde 2000, Lodish *et al.* 2000). Among them are immunoglobulins and cadherins, which interact with their counterparts situated in the neighboring cell membrane and form cell–cell contacts. Selectins bind to glycosylated mucins, and vice versa, and to some growth factors. Integrins are the most important class of CAMs that mainly bind to ECM proteins and interact with immunoglobulins too (**Figure 4**).

Integrins are heterodimers, consisting of two noncovalently linked subunits, called alpha and beta (Alberts *et al.* 2002, van der Flier and Sonnenberg 2001, Humphries 1990, Hynes 2002). Taken together, there are 24 different integrins known to date (Alberts *et al.* 2002, Baszkin and Norde 2000). Most integrins are expressed on a variety of cells, and most cells in turn express several integrins, enabling them to bind to several matrix molecules (Humphries 1990). Most important here being fibronectin and other ECM proteins such as collagens, laminin, and vitronectin.

The process of cell–matrix adhesion consists of a cascade of different overlapping events and is initiated by the interaction of integrin receptors that bind to small peptide ligands in the respective ECM proteins (Alberts *et al.* 2002). One well-known peptide ligand is the minimal cell recognition sequence RGD (R = Arginine, G = Glycine, and D = Aspartic acid) (Giancotti 2000, Giancotti and Ruoslahti 1999, Pierschbacher and Ruoslahti 1984a, b, Ruoslahti 1996, Ruoslahti and Pierschbacher 1987), which is found in many ECM proteins including fibronectin.

As a result of the ligand–receptor interactions, the cell begins to flatten (i.e., spreads) on the surface. Inside the cell, the same ligand–receptor interaction triggers the organization of actin filaments, which are often called stress fibers. Finally, integrin molecules, which cluster together in the membrane to so-called *focal adhesion points*, are attached to both the peptide ligand outside of the cell and the cytoskeleton inside (Geiger and Bershadsky 2001, Pande 2000, Petit and Thiery 2000, Zamir and Geiger 2001). It is important to recognize that the interaction of the peptide

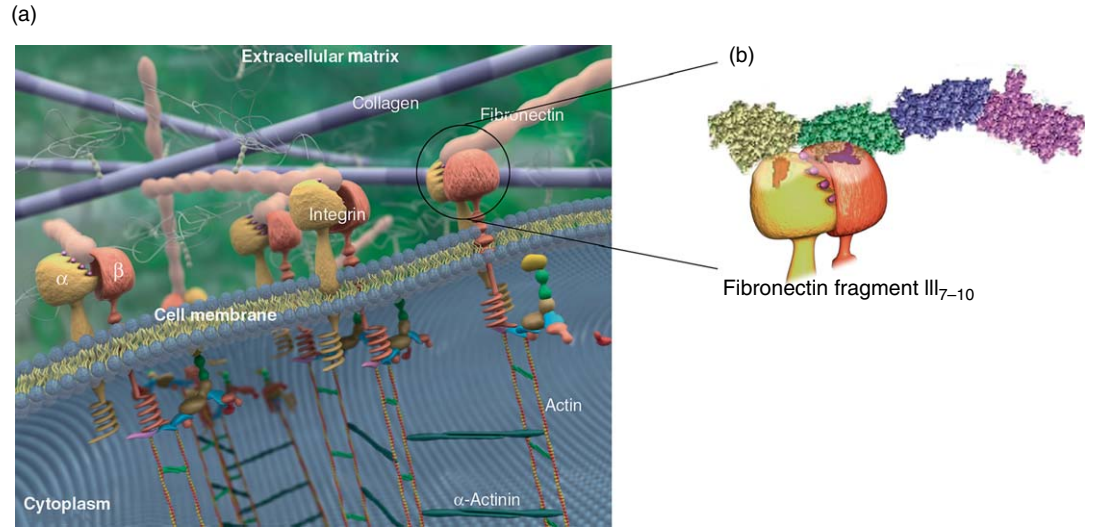


Figure 4 (a) Integrins are composed of two subunits, α and β, and they hold a cell in place by attaching at one end to proteins of the extracellular matrix (ECM) (and in specific to the cell-binding domain of the protein), like fibronectin, and at the other end to the cytoskeleton, the structural framework of the cell. They connect to the cytoskeleton through a highly organized aggregate of proteins, such as actin filaments and α-actinin. (b) Integrins bind to the adhesion domains of the ECM proteins by recognizing specific sites of that domain. A model for the recognition of the fibronectin G**RGD**S site and PHSRN synergy site by an integrin is shown. The four repeats of the fibronectin fragment $III_{7–10}$ are shown in different colors: purple for III_7 to yellow for III_{10}. The synergy site PHSRN in III_9 repeat is marked in blue and the G**RGD**S in III_{10} in red. The minimal amino acid sequence RGD in fibronectin and other cell adhesion proteins is the most important recognition site for about half of all known integrins. The synergy site is also required for high-affinity binding. (Source: Tirrell M, Kokkoli E, Biesalski M 2002 The role of surface science in bioengineered materials. *Surf. Sci.* **500**, 61.)

ligands with integrin receptors not only ensures the structural integrity of living cells but also signals a number of events inside the cell that have an influence on the metabolism, the differentiation, the proliferation and the immune response (Albeda and Buck 1990, Ruoslahti and Pierschbacher 1987, Travis 1993).

As the interaction of cells with microdevice surfaces is quite complex, a very precise tailoring of the surface properties of the involved materials is required. Thus various strategies have been developed to generate surfaces, which have tailor-made chemical composition and which allow the incorporation of functional groups.

1.05.4 Surface Modification of Microdevices

1.05.4.1 General Remarks

Various strategies have been developed in the past to obtain materials for microsystems applications with biocompatible properties. To obtain such a material (biomaterial), one can follow basically two different strategies. One is to generate a material that has all the desired bulk properties and still interacts in a satisfactory way with its surrounding environment. Such an approach leads to the ideal material but requires usually the screening of huge numbers of different materials, as frequently either one or the other requirement is not completely met and the material does not perform as desired. An alternative strategy for generating a biocompatible material is to optimize the bulk properties of the materials as far as possible, regardless of the surface properties and then modify the material's surface according to the requirements of the biological system to which it is exposed. Such a surface modification can be achieved through a chemical treatment process or through the deposition of a biocompatible coating.

The latter strategy has the advantage that the bulk and surface properties of the material become decoupled, and the use of a material is at first approximation independent from its surface properties. If, for example, a compound exhibits very good properties in contact with blood or tissue but has rather weak mechanical properties, it can still be applied as a coating material in mechanically challenging situations. On the other side, a material that has excellent bulk properties but has a strong

thrombogenic potential might still be useful when it is appropriately coated as long as the coating controls all the interactions with the blood.

It is well known that coatings influence the surface properties of a material strongly. Even layers that are just a few angstroms thick can alter the surface properties of a material in such a way that the chemical nature of the underlying material becomes completely hidden and the surrounding environment does not feel the presence of the material at all. The interaction of the whole system with the surrounding environment is then solely governed by the coating (stealth effect). In addition, such biofunctional coatings are by no means very cost effective, as the modification with appropriate proteins requires only little amount of these otherwise very expensive materials.

For the generation of biocompatible coatings on microdevices, a number of techniques are employed, e.g.:

- Adsorption of proteins/polymers
- Plasma treatment
- Formation of SAMs
- Chemical attachment (grafting) of polymers

1.05.4.2 Coatings Based on Physically Attached Polymers and Proteins

The simplest way to coat a substrate with a (biocompatible) coating is to deposit a biocompatible material directly on the surface of the substrate. Examples that are very popular in the literature are coatings containing poly(ethylene glycol) (PEG), with poly(lysine), or one of the serum albumins such as bovine serum albumin (BSA) or human serum albumin (HSA). The deposition of the material is most simple and can be achieved by dip coating, spin coating, spraying, casting, or through adsorption from a dilute solution of the macromolecules.

However, it should be noted that all chains, which are attached to the surface through physical interactions, are held at the surface through rather weak forces. While this is a favorable situation in cases when only a temporary coating is desired, it is in other cases problematic as the coating will slowly bleed off the surface and the properties of the coating decay. Under unfavorable conditions the films can be totally dissolved. This dissolution cannot be avoided as the concentration of the coating material in the contacting solution is zero. Accordingly, the surface concentration in equilibrium will also be zero and only a slow kinetics will prevent the complete disintegration of the film.

In many cases, even more problematic for the integrity of the surface coating is the displacement of the involved molecules by molecules present in the contacting environment. Especially if they show stronger interactions with the surface than the constituents of the coating. Molecules from the environment can compete efficiently for sites at the surface with polymer molecules and displace them from the surface. This aspect is especially important for production and storage of the devices. A somewhat frequently encountered problem is that the amount of material that can be deposited depends on the history of the sample, i.e., on the conditions to which it had been exposed prior to the coating step.

One way to avoid dissolution of the coatings material is to crosslink it, so that an infinite network is formed. In the biomaterials community, frequently the terms gel or hydrogel are used instead of polymer network. A gel is by definition a polymer network that is swollen by solvent and a hydrogel a polymer network that is swollen by water. 'Infinite' in this context means that the molecule forming the network is of the same size as the object itself. Networks can, by definition, not dissolve but only swell. If they swell, they will show at first a very strong buckling of the surface. What then happens, depends on the strength of interaction of the material and the substrate. If the interaction is weak, the strong mechanical forces caused by the osmotic pressure of the swelling network will lead to a complete or partial delamination of the network from the surface. If the adhesion forces are strong, just strong crumpling of the gel will result.

When inert surfaces that do not swell in the biological environment are desired, a deposition from the vapor phase can be envisaged. Examples are parylene deposition or deposition of siloxane molecules from the gas phase or the initiation of a crosslinking polymerization in the gas phase. As long as no creeping of solution under the network occurs in such systems, which would lead to delamination of the whole network from the surface, the coatings are very robust and stable.

Especially in biological environments, competing adsorbates can win out against physically deposited polymer chains/biomacromolecules and completely displace them from the surface. This is the case if the interactions of the contaminants with the surface are much stronger than those of the coated material with the same substrate. Under these circumstances

the polymer molecules of the coating will no longer stay in direct contact with the substrate but will be located on top of a thin layer consisting of the molecules that had displaced them. Displacement of already adsorbed polymer chains is a particularly serious issue if polar substrates are employed and are exposed to polar contaminants, a situation most frequently encountered in biological environments. As already discussed, the contacting liquids are in such systems from a chemical point of view complex and contain many molecules that show a strong interaction with surfaces. For example as far as proteins are concerned, if two proteins have a similar affinity to the surface (similar enthalpy of adsorption), the larger molecules will always displace the smaller ones for entropic reasons (entropy of mixing of the molecules with the aqueous solution). However, if short-term improvements of the biocompatibility of a device are required, physical deposition of polymers or biomolecules such as proteins will be the method of choice.

1.05.4.3 Plasma Treatment of Surfaces

Plasma, corona, or glow discharge processes are frequently used for the modification of surfaces (Inagaki 1996). In such processes the substrate is brought between two electrodes in a gas atmosphere at a reduced pressure. If now a voltage is applied, a cold plasma is formed inside the chamber, consisting of electrons, ions, and radical ions and neutral radicals originating from the gas molecules or atoms. In addition, visible and UV light are emitted. The UV light has a considerable energy and can induce further photochemical reactions. The highly reactive species formed, react with each other, with the wall of the reactor, and with the substrate. Typically, the gas in the plasma chamber is not static but is continuously flown through the chamber.

As far as the substrate is concerned, depending on the conditions during the plasma process, especially on the nature and concentration of the reactive species in the plasma and the energy input, either surface etching or the deposition of plasma polymers can be the predominant reaction. Through etching with, for example, argon or water plasma, reactive sites can be introduced into the surface and the surface can be cleaned from contaminants sticking on it. A surface that is otherwise chemically rather inert and not easily functionalized can be provided with a functional group through a short activation etch. After the introduction of the functional groups, the reactive species formed in the plasma process react with oxygen, once the sample is taken out of the plasma chamber to ambient conditions (**Figure 5**). As a word of caution, however, it should be mentioned that such a strategy for hydrophilization of an otherwise totally inert surface is only rather short lived. It is inherent to such techniques that with time the sample ages and more and more of the functional groups, that had been successfully created during the surface reaction, are lost. This is especially the case if the substrate is exposed to higher temperatures. This effect, commonly called hydrophobic recovery, is partly caused by a contamination of the surface with airborne organic molecules originating from the environment, but the most important aspect is a diffusion of the functional groups away from the surface into slightly deeper regions of the sample.

The driving force for this surface reconstitution is the tendency of any surface to minimize its surface free energy. Although the diffusion coefficient of polymer segments that are well below the glass transition temperature is rather slow, it should be noted that diffusions over very short distances, i.e., only a couple of angstroms, are required to cover all functional groups with low surface energy materials (Andrade *et al.* 1985). In addition, there is also evidence that surfaces are regions of enhanced mobility of polymer segments (Prucker *et al.* 1998). However, if the surface is directly modified after the plasma process, a simple functionalization of otherwise very inert materials can be achieved.

Under appropriate conditions during plasma deposition of organic compounds, layers of highly crosslinked polymers (plasmapolymer) are formed from fragments of the monomers that are present in the plasma. Appropriate conditions refers here to the amount of energy imparted into the gas molecules, which depends both on the energy input and the flow conditions. As monomers, any organic compound that has a sufficiently high vapor pressure under the reaction conditions can be used, even saturated alkanes such as hexane. The reaction mechanism for buildup of the plasmapolymer layers is rather complicated and is based on the fragmentation of the monomers in the gas phase into mono- and biradicals that eventually recombine to form the crosslinked plasmapolymer layer (Yashuda 1981). But even after deposition of the polymer layer, the presence of ions, free electrons, and high-energy irradiation can alter the chemical structure of the material already deposited. Thus the structure of the polymer layer has often no visible connection to the monomer used, and it is not possible to generate structurally well-

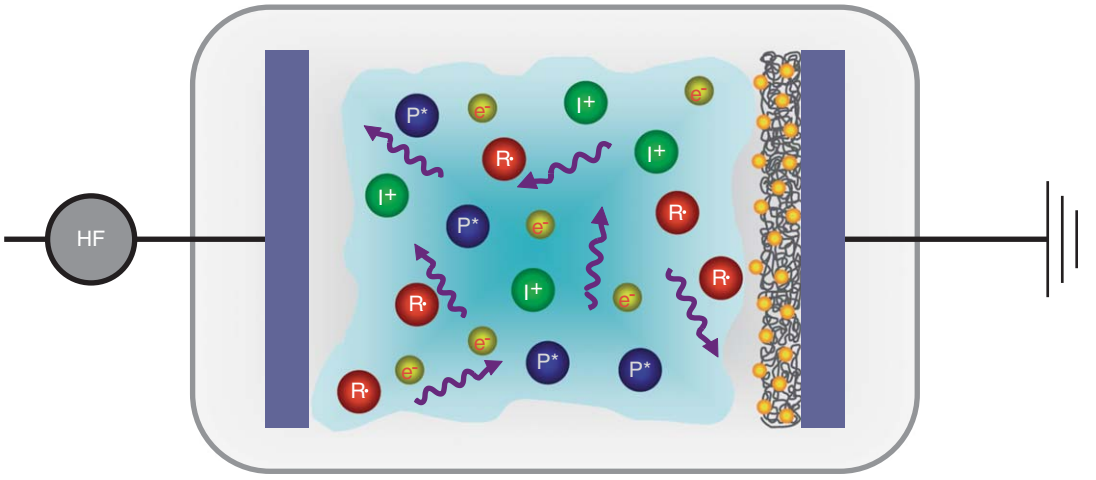


Figure 5 Schematic illustration of a plasma polymerization process.

defined functional monolayers. However, it is possible to deposit in a very efficient manner inert layers consisting of fluorinated units or polymers that resemble more or less glassy carbon. Thus, plasma processes are a very simple and cheap way to generate inert surfaces. If appropriate reaction conditions (low energy input, pulsed plasma) are applied, some monomers can also be used to generate layers containing reactive groups. Such layers can be used in postplasma reactions for the buildup of functionalized surfaces. Examples are plasmapolymer layers containing amine, carboxylic acid, or epoxide groups.

1.05.4.4 Self-Assembled Monolayers

An alternative that can overcome some of the problems of the processes described in the previous section, is to covalently attach the molecules of the coating to the surface of the substrate to be coated. In such cases removal of the coating is only possible at the expense of breaking one or several of the chemical bonds that connect them to the surface. As this requires very high amounts of energy, such a process occurs only under rather drastic conditions. An additional attractive feature of all situations, where molecules are chemically attached to the substrate, is that chemical surface reactions in general are independent from the finer details of the geometry of the substrate, which is a situation, especially favorable for microsystems. Thus even substrates with a rather odd topography such as sharp needles, cavities that are connected to the outside only through a tiny hole, or the inside of narrow microfluidic channel can be coated. This holds true irrespective of the finer details of the immobilization reaction.

Apart from studies where a blocking of the surface with alkyl or fluoroalkyl groups is desired, most silanes employed in microsystems engineering are α, ω-functionalized. One end of the molecule carries a reactive site, which causes the immobilization reaction. The other end carries functional groups such as amino, active ester, or epoxy groups, which can be used for the binding of complex moieties involved in improving the biocompatibility of the surface (**Figure 6**).

Reactive groups that allow to couple the molecules to the surface most frequently consist of silane, thiol, disulfide, or phosphonate groups. Silanes are mostly used in the form of alkoxy or chloro silane compounds and are used to modify oxidic surfaces. When di- and trifunctional silanes are used, self-condensation can occur, which leads to the formation of thicker layers, which represents in a way a siloxane network attached to the surface. Such layers are,

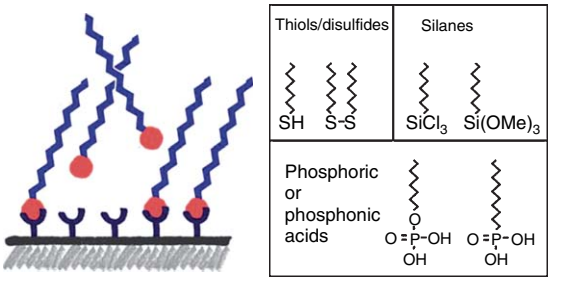


Figure 6 Schematic illustration of the self-assembly process and examples of anchor groups used for the modification of surfaces with self-assembled monolayers (SAMs) of organic molecules.

strictly speaking, not monolayers. This difference is important for a detailed understanding of the surface chemistry; however, in practical terms, this difference is more semantic than of real practical value as far as qualitative results are concerned. If, for example, a glass surface is to be made hydrophobic using a fluorinated silane, the difference between a clean monolayer and a thin surface-attached network of these compounds is marginal. However, if quantitative results are required, for example, when DNA molecules are immobilized on an SiO_2 surface and then used for a further quantitative analysis, the network formation process leads to difficulties in process and quality control of the so-generated biochips (Southern 1996). Silane monolayers are for this reason a little more difficult to form. If clean monolayers are desired, the immobilization reaction must occur in the strict absence of water. However, silane monolayers on SiO_2 surfaces are very stable and can be removed from the surface only when the substrate starts to dissolve, which is the ultimate stability limit for surface coatings anyway. One of the most popular ways in the bio-oriented community is the attachment of functional groups via γ-aminopropylsilanes immobilized to oxidic surfaces (e.g., glass-substrates). This is the case, despite that it has been known for many years that the (basic) amino groups exhibit a strong interaction with the acidic silanol groups on the surface. If, however, some of the molecules are bound only through acid–base interactions instead of formation of a siloxane bond, this will lead to a slow bleeding of the silane compound. In addition, anyone who has experience with amines, knows that alkyl amines have a strong tendency to photo-oxidation, which transforms the amine group into nitroxides and other oxidation products, so that the shelf lifetime for such monolayers is rather limited. Aminosilane monolayers are popular as they allow a very simple additional functionalization. This is especially important as the reactive groups in silanes do not tolerate many functional groups. For example, hydroxy or acid groups are not compatible with reactive silanes.

The sulfur compounds mentioned above (thiols and disulfides) can be used to modify metal, especially noble metal surfaces. Particularly gold substrates have been very popular with these compounds. Monolayers on gold have been a very common subject of study, because they are most easily prepared. Dipping the gold substrates into the thiol or disulfide solution followed by a quick wash with pure solvent is sufficient for the generation of high-quality and in some cases even 2D crystalline monolayers (**Figure 7**).

Thiol and disulfide groups have the additional great advantage that they tolerate a wide spectrum of functional groups, and the buildup of complex surface architectures is greatly facilitated. As the substrate is a noble metal, electrical or electrochemical measurements are also possible as long as the oxidation potential is kept in a moderate range. At electrochemical potentials higher than -1.1 V, the sulfur-containing monolayers are completely removed from the surface (Widrig *et al.* 1991, Yang *et al.* 1997).

On the downside, the binding energy of the compounds to gold surfaces is not so very high. Thermal desorption studies place the bond strength somewhere between a real chemical, say carbon–carbon bond and a hydrogen bond, so that under some circumstances thermal energy is sufficient to remove the molecules (Delamarche *et al.* 1994). In addition, thiols and disulfides have a strong tendency to oxidize under ambient conditions with the formation of sulfites and ultimately sulfates. The latter, however, have a much weaker interaction with the substrate surface, so that the oxidation leads in the end to a bleeding of the compound and a strong reduction in the quality of the monolayer.

The surface immobilization reactions, with very few exceptions, are based on condensation reactions. One of the few exceptions is the immobilization of alkanes on SiH-terminated silicon surfaces. Despite the huge number of publications on thiol monolayers, the mechanism of the attachment of the compounds is not entirely clear. The present model discusses a gold oxide (Widrig *et al.* 1991). The surface reactions occur more or less spontaneously when surface and reagent solution or vapor are brought into contact and stop once a monolayer is formed as the reactive

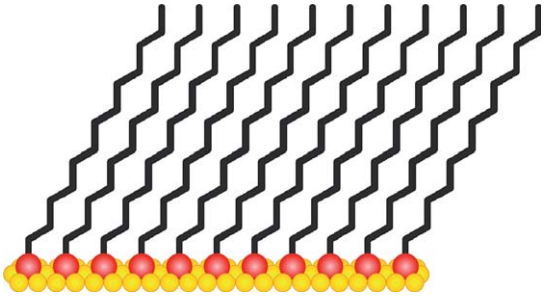


Figure 7 Schematic illustration of a self-assembled monolayer consisting of an *n*-alkanethiol ($CH_3(CH_2)_nSH$) attached to a gold surface.

sites at the surface are consumed. Thus the term self-assembled monolayer is commonly used. SAMs clearly represent the most important technique for surface modification at least as far as oxide materials such as glass, silicon, alumina, titanium dioxide, and similar compounds or metals (especially gold) are concerned. A variety of other compounds such as nitrides contain also a thin layer of oxide at the surface. Such surface modification processes are technically very important (priming of surfaces) and many thousand research papers have been published on the generation, analysis, and application of such layers.

1.05.4.5 Immobilization of Polymers

For the immobilization of polymers to substrate surfaces, essentially four different strategies can be employed:

- Chemisorption of polymers (grafting to)
- Chain attachment through immobilized monomers (grafting through)
- Surface-initiated polymerization at the surface *in situ* (grafting from)
- Chemical attachment and simultaneous network formation

In the first process, specific polymer molecules that carry functional groups which allow the surface attachment are required. In the next two processes, the surface-attached chains are formed at the surface *in situ*, i.e., the substrate is exposed to conditions in which a polymerization reaction is carried out, i.e., the presence of monomer is required (**Figure 8**).

1.05.4.5.1 Chemisorption

One important strategy for the attachment of polymers to a biomaterial's surface is the chemisorption of an appropriately functionalized polymer (Ben Ouada *et al.* 1988, Bridger and Vincent 1980, Krenkler *et al.* 1953, Laible and Hamann 1980, Tsubokawa *et al.* 1989, 1990). In such a case the polymer either contains a functional moiety that can react with an appropriate site in a SAM or carries an anchor group either as an end group or in a side chain. This can then react with appropriate sites at the substrate surface. In either case a monolayer of polymer molecules is obtained, which is chemically bound to the surface. This strategy is in principle closely related to the formation of SAMs described above. In agreement with what has been said before, such endfunctionalized polymers are available; the attachment of the polymers is from a chemical point of view rather simple. This condition sounds rather trivial, but it is not so. The synthesis of polymer chains with reactive end groups is restricted to certain classes of polymers, mostly polymers that can be obtained by polycondensation. In case of polymerization processes occurring as chain processes, the reactive group is typically protected during the polymerization reaction. Only in very few cases a polymerization reaction can be carried out in the direct presence of reactive groups. In most cases either an initiator or a monomer containing the protected moiety is employed during the polymer generation. After the chain formation is complete, the protection group is cleaved off in a polymer-analogous reaction so that the reactive site is retrieved. While a number of polymers having functional groups in the side chain are available, not quite as many that contain a reactive side at one of the chain ends are available.

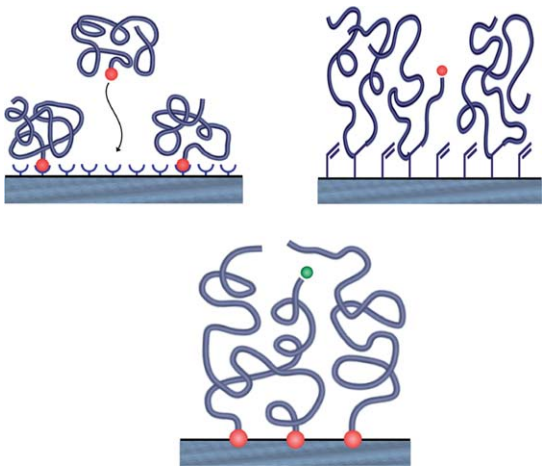


Figure 8 *Top left*: Schematic illustration of different processes used for the attachment of polymers to surfaces – Grafting-to process. *Top right*: Grafting via incorporation of surface-bound monomeric units. *Bottom*: Grafting from or surface-initiated polymerization.

While the attachment of terminally functionalized polymers to the surface leads to layers where one group is connected to the surface, side chain attachment usually leads to multiple attachment points and both very tight binding and a rather flat conformation of the polymer chains. In the latter case the functional groups either on the same chain or on different molecules compete for the reactive sites on the surface. Accordingly, the amount of polymer that can be immobilized depends strongly on the reaction conditions, especially on the concentration of the

polymer in solution. If the concentration is rather low and the rate of the surface attachment reaction is high, the molecule will after initial contact rapidly bind at many different sites to the surface. Accordingly, the molecule obtains a rather flat conformation and uses up many surface sites. If the formation of the chemical bond is irreversible under the reaction conditions, the overall number of molecules that can be attached remains rather low and only relatively thin film can be attached. At high polymer concentration, however, many different molecules are simultaneously present at the surface and compete in a very efficient way for surface sites, so that the overall graft density becomes higher and the thicker monolayers are obtained.

One aspect that is inherent to grafting-to processes and of which one has to be aware is an intrinsic limitation of the film thickness and accordingly the number of functional groups per surface area that can be obtained by such an approach (**Figure 9**). Films generated by chemisorption from solution are limited to (dry) thicknesses of typically 1 to 5 nm. This limitation has both kinetic and thermodynamic origins. With increasing coverage of the surface with attached chains, the polymer concentration at the interface quickly becomes larger than the concentration of the polymers in solution. Additional chains, which are to become attached to the surface, have to diffuse against this concentration gradient that ever increases with increasing graft density of the attached polymer. This diffusion slows down the immobilization reaction further and further as the reaction proceeds. Thus the rate of the attachment reaction levels off rather quickly (exponentially!) and another polymer is linked to the substrate only at an extremely slow rate due to this kinetic hindrance.

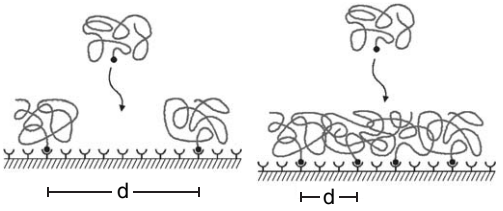


Figure 9 *Left*: Schematic illustration of the grafting-to process. Chains that are to be attached to the surface can easily reach the surface at low grafting densities. *Right*: The attachment process comes virtually to an end as soon as the surface is significantly covered with polymers, as the already attached chains form a kinetic barrier against which incoming polymer chains have to diffuse to reach the surface.

Indeed, it has been shown both theoretically and experimentally that once the surface-attached coils overlap, the attachment of other polymer molecules takes place only on a logarithmic time scale and very quickly time frames of thousands or even millions of years would be required to add a few more nanometers of polymer to the layer. Accordingly, as far as practical reaction times are concerned, films generated by this technique are intrinsically limited to the film thickness. Even if this kinetic limitation is somehow circumvented (e.g., by resorting to high polymer concentrations, which de-emphasize the importance of the diffusion of polymer molecules), the attachment of chains to a strongly covered surface becomes unfavorable also for thermodynamic reasons. At high grafting densities the surface-attached polymer chains are in a rather stretched conformation due to the presence of strong segment–segment interactions as will be discussed in greater detail later. A chain that now attaches to the surface has to change from a coil conformation in solution to a stretched (brush-like) conformation at the surface. The entropy loss during this process, however, is only compensated by the establishment of one chemical bond, namely, the one connecting the polymer to the surface. The higher the graft density of the chains at the surface, the stronger will be the entropy penalty that rapidly precludes the attachment of more chains.

An intrinsic limit to what layer thickness can be reached sounds a bit negative, but it also has its positive side. As the layer thickness is given more or less by the surface concentration of the segments of the already attached chains, the chemical peculiarities of the system do not matter. Thus a very robust system that gives the same result under many different conditions is obtained. As long as the desired film thickness conforms to this regime, i.e., is between 0 and about 5 nm, such a system is from a practical point of view great as the same film thickness is always obtained.

A new way of covalently attaching polymers, which has been described recently, is through a photochemical reaction. Here photoactive moieties are self-assembled at the surface of the substrate. During irradiation of the immobilized molecules with UV light, a photochemical reaction sets in, which leads to a covalent binding of polymers or other macromolecules (DNA, proteins, and sugars) at these surfaces. The binding is achieved through a radical coupling reaction. This reaction can occur with almost all organic molecules and is followed by a very simple and versatile immobilization process.

1.05.4.5.2 Immobilization using surface-attached monomers

Another straightforward technique for the attachment of polymers to surface, which allows to generate a large variety of functional surfaces, is to add a monomer-modified substrate to a solution in which a polymerization reaction is carried out (Dmitrenko *et al.* 1990, Hashimoto *et al.* 1982a, b, Kopf *et al.* 1996, Trachenko *et al.* 1986, Zajac and Chakrabarti 1995). For the performance of the surface-attachment reaction, the monomer units are initially self-assembled at the surface. During the polymerization reaction, which occurs in the contacting solution, statistically every once in a while a surface-attached monomer is incorporated into a growing polymer chain. However, as soon as one surface-attached monomer is incorporated, the growing polymer is becoming permanently attached to the surface. If several monomer units on the surface are incorporated, this will not change the general picture, but the conformation of the polymer molecule will change (**Figure 10**).

As far as the polymerization mechanism is concerned, the surface-attachment reaction is a two-step process: During polymerization a macroradical first attacks a monomer immobilized on the surface (grafting-to reaction). In the second step more monomer units are added, so that the chain grows further (grafting-from reaction). However, careful studies of the polymerization mechanism have shown that the grafting-to step represents the bottleneck of the reaction and thus limits the polymer immobilization.

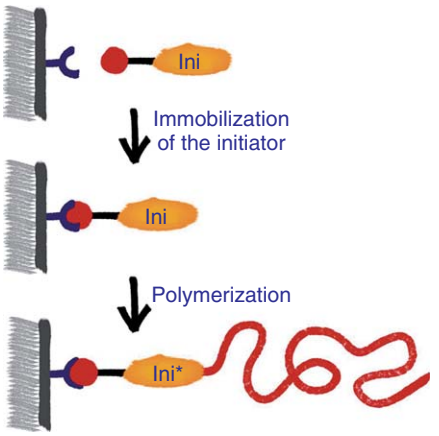


Figure 10 Synthetic strategy for the generation of polymer brushes via surface-initiated polymerization. An initiator molecule is deposited on a surface by means of self-assembly process via the reaction of an anchor group to suitable surface functional sites. Subsequently, chains are grown from the surface using the immobilized initiation sites.

Accordingly, very similar layers are obtained by using immobilized monomers as through the chemisorption of preformed chains.

However, if this poses no problem, it is one of the simplest techniques to generate surface-attached layers, especially as no polymer functionalized with an anchor group needs to be synthesized.

1.05.4.5.3 Growing polymer chains at surfaces

Interesting systems with rather unusual molecular architecture of the polymer chains, and accordingly unusual physical properties quite different to those in films described above, can be obtained if the polymer chains are grown at the surface of the substrate *in situ* (Advincula *et al.* 2004, Pyun *et al.* 2003, Rühe and Knoll 2002, Zhao and Brittain 2000). For the synthesis of such systems the initiator species is self-assembled at the surface of the substrate either directly or using a functional monolayer and attaching the initiator in a surface-analogous reaction. The initiator monolayer is then used to kick off the growth of polymer chains from the surface. Depending on the initiator graft density and the reaction conditions, dense or less dense lawns of long or short polymer chains can be grown. Here the term dense refers to the graft density of the polymer chains and is not related to the material density. Depending on the chemical nature of the initiator, the polymerization reaction can be started by applying heat or light or by carrying out a chemical reaction such as a redox process. As the reaction proceeds, more and more monomer units are added until either the chain growth process is terminated or a transfer reaction to monomer or solvent occurs. Such transfer reactions lead only to the formation of free polymer that is washed away in an extraction step and that does not contribute to the film formation. Depending on the reaction mechanism and on the rate of chain growth versus the rate of chain initiation, all polymer chains are either grown simultaneously or at different periods of time (**Figure 11**).

This way, monomolecular layers of polymer molecules with film thicknesses of more than 2000 nm in the dry state have been obtained. In this case polymer molecules with number-average molecular weights of several 10^6 g mol^{-1} are attached at distances of anchor points of less than 3 nm. Whereas similar graft densities can be easily reached, such high molecular weights can be achieved only under certain favorable conditions and not with every monomer. The growth of polymers from the substrate surface is accordingly ideally suited if layers ranging from a few nanometers

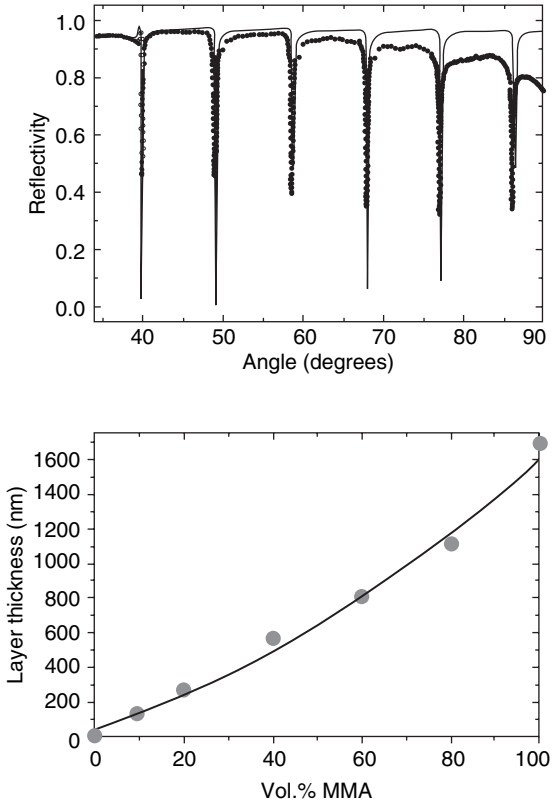


Figure 11 *Top*: Optical waveguide mode spectrum (symbols) obtained from a polymethyl methacrylate (PMMA) brush deposited on an evaporated SiO_2 layer. The solid line corresponds to a model-fit calculation based on a Fresnel formalism assuming a 2200-nm-thick polymer layer. The sample was prepared in neat MMA at 50 °C. *Bottom*: Dry thickness of a PMMA brush as a function of the monomer concentration used in the polymerization feed. Polymerizations were carried out at 60 °C for 18 h (solvent: toluene).

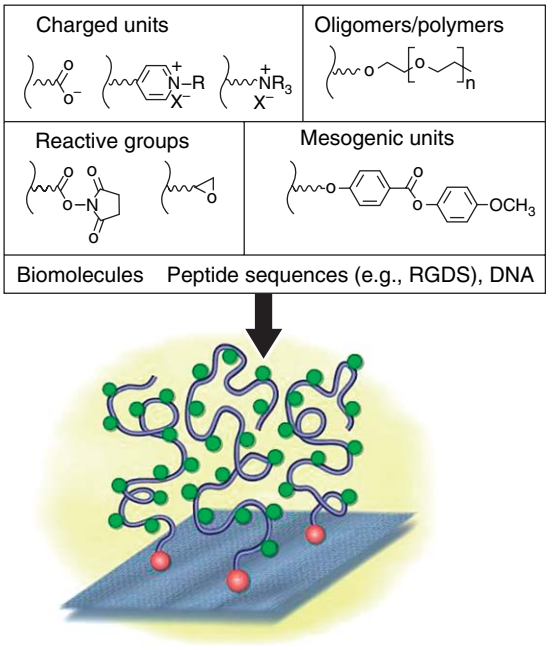


Figure 12 *Top*: Examples of functional groups incorporated into polymer brushes. *Bottom*: Schematic illustration of a functional polymer brush.

to a few hundred nanometers are desired. Especially to the high end of layer thicknesses, some limitations exist as neither the graft density nor the molecular weight of the surface-attached chains can be pushed beyond certain limits.

Polymer chains that are attached to a surface at a high graft density assume a brush-like conformation. The repulsion between polymer segments leads to a strong stretching of the polymer chains away from the surface. This chain stretching causes very unusual properties insofar as polymer brushes cannot be penetrated by other macromolecules in solution, exhibit strongly reduced friction, or show an autophobic wetting behavior (i.e., cannot be wetted even from chemically identical chains) (**Figure 12**).

Surface-initiated polymerization reactions work for any polymer that can be obtained by a chain growth reaction such as free and controlled radical polymerization, carbocationic polymerization, anionic polymerization, and ring-opening metathesis polymerization. As pointed out earlier, the growth of the polymer chains can be achieved on arbitrarily shaped surfaces (planar, curved, and irregular surfaces, and nanoparticles) and is ideally suited for growing chains on complex microstructures. Depending a little on the details of the chosen reaction mechanism, chain growth processes allow for the generation of polymers from a wide spectrum of different monomers.

1.05.4.5.4 Chemisorption and simultaneous network formation

All of the three strategies described above for the covalent attachment of polymer monolayers to microdevice surfaces can also be used for the covalent attachment of polymer networks/gels (**Figure 13**). In polymer networks all the chains form one gigantic molecule with infinite molecular weight. Through network formation, layers with infinite thickness can be obtained, the only limit being the availability of crosslinkable material. To form these networks, either the polymer chains have to be modified with reactive units, which allow a reaction of the polymer molecules with each other, or bi-, tri-, or multifunctional monomers have to be used in

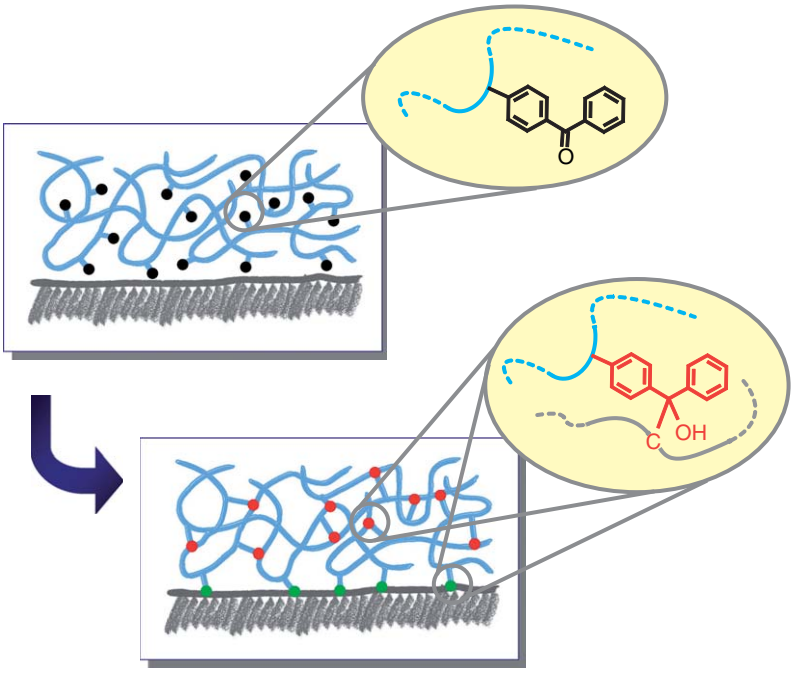


Figure 13 Schematic illustration of the formation of surface-attached polymer networks (polymer gels) by photochemical crosslinking benzophenone moieties built in into polymer molecules; during formation, the network is simultaneously attached to the surface.

a polymerization reaction. To obtain a chosen finite thickness of the layer of the coating, the thickness of the layer containing the polymer or monomer that is to be crosslinked has to be confined. Examples of such reactions are the photochemical crosslinking of benzophenon group-containing polymers and the grafting-through reaction using trifunctional methacrylates (Prucker *et al.* 2003, Toomey *et al.* 2004).

1.05.4.5.5 Surfaces that resist the nonspecific adsorption of proteins

As nonspecific adsorption of proteins to microdevice surfaces is a very important factor for the functioning of these devices, a brief mention will be made about the generation of surfaces that resist nonspecific protein adsorption, regardless of how they were synthesized and whether the molecules were chemically or physically attached to the surfaces.

In order to develop surfaces that resist the nonspecific adsorption of proteins (i.e., nonfouling surfaces), a large number of physicochemical and biochemical studies have been carried out on the fundamentals and the applications of protein adsorption to the material's surface (Brash and Horbett 1987, Hern and Hubbell 1998, Horbett and Brash 1995, Jenney and Anderson 1999, Jeon *et al.* 1991, Kunz *et al.* 1999, Lehmann and Ruehe 1999, Sigal *et al.* 1998, Szleifer and Carignano 2000, Xu and Marchant 2000, Yang *et al.* 1999). Despite its large impact on the aforementioned local reactions of a biological host to an implanted biomaterial, and extensive research efforts, the details of the mechanism of nonspecific protein adsorption have not yet been completely understood. However, it is believed that parameters such as van der Waals interactions, electrostatic forces, and H-bonding govern unspecific adsorption (Ratner and Bryant 2004, Sigal *et al.* 1998, Szleifer and Carignano 2000).

Interesting examples of nonfouling surfaces are materials that have been surface-modified with PEG (Andrade *et al.* 1996, Harris 1992), polymer hydrogels (Andrade *et al.* 1996, Elbert and Hubbell 1996, Griffith 2000, Hern and Hubbell 1998, Horbett and Brash 1995, Hubbell 1999, Jenney and Anderson 1999, Jeon *et al.* 1991, Kunz *et al.* 1999, Peppas *et al.* 2000), polymer brushes (Lehmann and Ruehe 1999, Szleifer and Carignano 2000, Xu and Marchant 2000, Yang *et al.* 1999), or more complex architectures (Shen *et al.* 2002, Siegers *et al.* 2004). PEG is a water-swellable, nontoxic, and biocompatible polymer, and its use in biology and medicine has been widely reviewed in detail in the literature (Amiji and

Park 1993, Andrade *et al.* 1996, Jenney and Anderson 1999, Llanos and Sefton 1993). A different name for essentially the same polymer, which is more common in the polymer science community, is poly(oxyethylene) (PEO). Various schemes have been developed to attach PEG molecules to surfaces, through SAMs, as block copolymers, as comb-polymers, and so on.

Although PEG is widely employed to coat surfaces of biomedical devices and is the material of choice for many applications, it is also known to exhibit thermal instability (Gordienok *et al.* 1986) and rapid auto-oxidation either through O_2 in the presence of transition metal ions or enzymatically *in vivo* (Andrade and Hlady 1986). Thus the quest for stable, nonfouling surfaces that resist the nonspecific adhesion of protein to surfaces is still an ongoing battle.

1.05.5 Engineered Biomaterials Directing the Adhesion of Living Cells

One approach to address the challenge of designing biomaterials that specifically interact with living cells is the coating of a material with whole proteins, such as laminin, collagen, or fibronectin. Although all these proteins carry ligands that promote the attachment of cells, there are some limitations that these strategies have in common. For example, proteins need to be isolated from organisms and purified. Proteins may induce unwanted immune responses, and proteins are often subject to enzymatic degradation, which limits in particular their long-term applications. Finally, protein attachment to a biomaterial's surface may induce conformational changes in the protein itself, which elicits again unwanted foreign body reactions.

Since the discovery of small peptide ligands found in the cell adhesion-promoting ECM proteins, scientists have started to manipulate cellular behavior by engineering these small peptide ligands into a variety of biomaterials, e.g., by directly grafting these ligands onto solid surfaces or by immobilizing them into/onto polymeric materials and surface coatings. A detailed overview of the different strategies that have been followed to date has been recently published by Tirrell *et al.* (2002) and Hersel *et al.* (2003).

In principle, there are several important fundamental parameters that determine the specificity and binding affinity of synthetically formed biomimetic surfaces. An important parameter is the conformation of the peptide ligand, which is presented at the surface. For example, human melanoma cells attach and spread on surfaces presenting the RGD ligand in a looped conformation (similar to the naturally occurring conformation in fibronectin) in a concentration-dependent manner (Pakalns *et al.* 1999). In contrast the same cell type spreads indiscriminately on carboxyl-coupled RGD and does not spread on amino-coupled RGD surfaces (Pakalns *et al.* 1999).

Another very important parameter is the surface density of RGD peptides, i.e., the surface concentration of the bioactive ligand (Arnold *et al.* 2004, Fittkau *et al.* 2005, Hern and Hubbell 1998, Irvine *et al.* 2001, Maheshwari *et al.* 2000, Massia and Hubbell 1991, Pakalns *et al.* 1999). A surface concentration of 1×10^{-15} mol cm^{-2} of GRGDY ligand, which was covalently grafted to the surface of otherwise poorly adhesive glass substrate, was sufficient to promote fibroblast cell spreading. However, focal contact formation was only observed at concentrations of at least 10×10^{-15} mol cm^{-2} and higher (Massia and Hubbell 1991). Hence, a minimum RGD density of 10×10^{-15} mol cm^{-2}, corresponding to a lateral (stochastic) spacing of about 140 nm between peptide ligands, is necessary to render the cells attached in healthy conditions.

Another parameter that influences the extent of cell adhesion is ligand (peptide) clustering at the nanoscale level (Maheshwari *et al.* 2000). Griffith and coworkers applied a number of star-like peptide-modified polymer molecules to surfaces, thereby forming a surface coating that on the one side resists protein adsorption and on the other side promotes the ligand-directed adhesion of living cells. By varying the number of peptide ligands per star molecule, and by mixing these bioactive peptide-polymer materials with nonmodified polymer stars, the authors were able to address different situations, where the bioactive ligand is presented either in clusters (more than one peptide ligand per polymer molecule) or homogenously on the surface (Maheshwari *et al.* 2000). The clustering of the YGRGD ligand significantly reduced the average ligand density required to support fibroblast cell migration, whereas nonclustered ligands supported cell attachment but did not promote full spreading (Maheshwari *et al.* 2000). Cell behavior also can be influenced by creating a biofunctional surface in which the accessibility of the ligand is used as the control parameter. Therefore, selective masking of a peptide ligand by PEG chains of varying length on a biomimetic surface is another new method of controlling the surface bioactivity (Dori *et al.* 2000).

Finally, the use of peptide synergy sites copresented with the bioactive ligand can have an influence on the extent of cell attachment to artificial surface-modified materials (Mardillowitch and Kokkoli 2004, Mardillowitch *et al.* 2006).

As one can infer from the data available on cell adhesion on peptide-modified artificial surfaces, the impact of many of these parameters on the cellular response is still not very well understood, and quantitative results reported so far vary significantly. For example, as outlined above, the minimum concentration of peptide ligand for cell adhesion and spreading varies over a very broad range from about femtomoles per centimeter square (Massia and Hubbell 1991) to almost nanomoles per centimeter square (Dori *et al.* 2000, Maheshwari *et al.* 2000), depending on the cell lines used and the design of the artificial matrix.

With respect to the latter, both top-down (e.g., the postsynthetic modification of materials with peptides) and bottom-up strategies (e.g., the self-assembly of preformed peptides and peptide-hybrid molecules) for various devices or template surfaces have been the locus of ongoing interest. Significant efforts have been devoted to research on peptide-modified polymers (Hersel *et al.* 2003). However, in many of these cases, the exact peptide surface distribution and the amount of peptides that finally interact with the integrin receptors have not been quantitatively proven yet. Recently, Arnold *et al.* (2004) have used peptide-modified gold nanoparticles adsorbed to solid substrates in order to investigate the function behind molecular arrangement of single-integrin receptors in cell adhesion. Because of the size of the gold clusters used ($D < 8$ nm), the authors claim that only one integrin binds to each of the particles. The authors propose a universal length scale of 58—73 nm, i.e., distance of individual integrins, as optimum for integrin clustering and activation (Arnold *et al.* 2004).

Another way to present the peptide ligands on artificial surfaces that possess a very high amount of control of the ligand has been introduced by the design of so-called peptide-amphiphiles (Baronas-Lowell *et al.* 2004a, b, Berndt *et al.* 1995, Biesalski *et al.* 2005, 2006, Dori *et al.* 2000, Fields *et al.* 1998, Malkar *et al.* 2003, Mardillowitch and Kokkoli 2004, Mardillowitch *et al.* 2006, Pakalns *et al.* 1999). Peptide-amphiphiles can be synthesized by attaching distinct peptide motifs onto a synthetic amphiphile using solid-phase organic chemistry, and subsequently, these molecular building blocks can be self-assembled into various planar and 3D architectures (monolayers, bilayers, micelles, and vesicles) with excellent control over peptide presentation, which allows the detailed analysis of parameters such as the ligand conformation (Pakalns *et al.* 1999), the ligand accessibility (Dori *et al.* 2000), or the order and conformation of the applied amino acid sequence (Baronas-Lowell *et al.* 2004a, b, Malkar *et al.* 2003, Pakalns *et al.* 1999) on the cellular behavior on such model surfaces. Moreover, in a recent study, Fields and coworkers have shown that a mixed peptide-amphiphile surface can be used to promote endothelial cell adhesion while inhibiting the attachment of other cell types that are commonly found in inflammatory responses (Baronas-Lowell *et al.* 2004a, b). Such bioactive bilayer films offer advantages as model systems; however, the potential use of such layers as surface coatings for biomedical applications is limited because of the poor stability of such architectures in air and organic solvents. A possible way to overcome such problems has recently been introduced by the preparation of so-called polymerizable peptide-amphiphiles, where a covalent bond is formed between adjacent amphiphiles in order to stabilize both spherical (i.e., vesicles) (Biesalski *et al.* 2005) and planar bioactive architectures on solid supports (Biesalski *et al.* 2006) (**Figure 14**).

Although the most common cell-binding domain that has been used is the RGD sequence, other non-RGD-containing cell-binding domains exist (Ruoslahti 1996), such as the peptides YIGSR and IKVAV in laminin (Vukicevic *et al.* 1990), REDV and LDV in fibronectin (Mould *et al.* 1991), DGEA in collagen I, and various heparin-binding domains (Dalton *et al.* 1995). In one example it was shown that a more complete cell attachment, spreading, focal contact formation, and organized cytoskeletal assembly were obtained by providing the cell with both the RGD ligand and heparin-binding domains of fibronectin (Dalton *et al.* 1995) or bone sialoprotein (Healy *et al.* 1999, Rezania and Healy 1999).

1.05.6 Conclusions

In the coming years, microdevices will become an important addition to medical technology and revolutionize some areas of medicine completely. This revolution will occur along several avenues. Microengineered devices will create new diagnostic

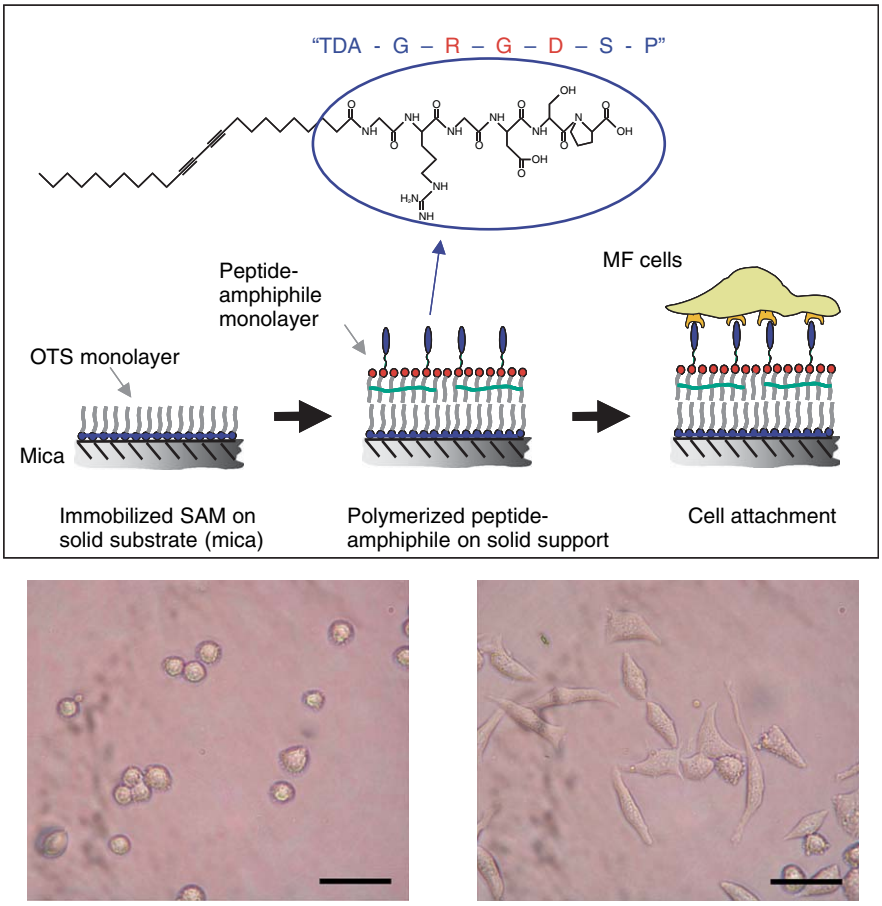


Figure 14 *Top*: Schematic description of a strategy for the buildup of biomimetic surfaces that direct cell adhesion on solid surfaces using self-assembled, polymerized peptide-amphiphiles. The structure of the bioactive polymerizable amphiphile (TDA-GRGDSP) carrying the bioactive RGD peptide sequence is schematically shown as inset. *Bottom left*: Mouse fibroblast (MF) cells adhere but do not spread at surfaces that expose no bioactive ligand. *Bottom right*: Cells adhere and spread in a concentration-dependent manner on surfaces that expose the ligand. SAM, self-assembled monolayer. Scale bar, 100 μm. (Source: Biesalski M, Knaebel A, Tu R, Tirrell M 2006 Cell adhesion on a polymerized peptide-amphiphile monolayer. *Biomaterials* **27**, 1259; reproduced in part with permission.)

tools and improve the long-term monitoring of patients. They will improve medical diagnosis and help avoid the occurrence of critical situations that could cause the patient harm or even death. In some areas it can even be envisioned that microdevices allow to merge diagnostics and therapy and create a new field for which the term theranostics has already been coined. Microdevices will allow to restore body functions that have been impaired or destroyed by acute or chronic illness or through an injury.

However, the intimate contact between the microdevice and the body of the patient also puts very stringent requirements on the biocompatibility of the devices as the contact of the devices with the body might occur for prolonged periods of time. The response of the biosystem against the implanted device might cause harm to the patient or, vice versa, the biological environment to which the microdevice is permanently exposed to might damage or even destroy important functions of the microdevice. Thus the surface engineering of the materials involved in the construction of the microdevices will in the future become more and more important, as surfaces with engineered properties allow to control the interactions of the device with the biosystem.

The creation of tailor-made surfaces of microdevices allows to improve the biocompatibility of such devices significantly. The biocompatibility of the device, however, is critical to the success of the

whole microdevice development project. As long as no appropriate material is found, the whole project has to halt regardless of success in other areas.

In the past biocompatibility was to a great extent related to empirical studies only and various materials had to be screened to find an appropriate one. For such studies the device was exposed to biological environments of increasing complexity and the response was studied. This rather unsatisfactory situation of a more or less educated guessing is currently giving way to a deeper understanding of the interactions of a biosystem and a microdevice. On the one hand, surface analytical tools have improved a great deal and made a more in-depth understanding of complex interfaces of biomaterials with biosystems possible. On the other hand, a number of strategies have been developed, which allow a long-term modification of the surface chemistry of the device, so that adverse reactions can be avoided and specific benign responses can be induced. As a consequence, device surfaces become more and more tailored toward a specific application and mimic a biological system more closely. However, before the goal of truly biocompatible microdevices is reached, very significant research efforts are required, as still many aspects of the interaction of biomaterials and biological environments are poorly understood. Accordingly, any R&D project geared toward biomedical applications of microsystems must tackle this challenge.

References

Advincula R C, Brittain W J, Caster K C, Rühe J 2004 (eds.) *Polymer Brushes*. Wiley-VCH, Weinheim

Albelda S M, Buck C A 1990 Integrins and other cell-adhesion molecules. *FASEB J.* **4**, 2868

Alberts B, Bray D, Lewis J, Roberts K, Watson J D 2002 *Molecular Biology of the Cell*. Garland Publishing, New York

Amiji M, Park K 1993 Surface modification of polymeric biomaterials with poly(ethylene oxide), albumin, and heparin for reduced thrombogenicity. *J. Biomater. Sci. Polym. Ed.* **4**, 217

Anderson J M 1993 Cardiovascular device retrieval and evaluation. *Cardiovasc. Pathol.* **2**, 33–41

Anderson J M 1994 In-vivo biocompatibility of implantable delivery systems and biomaterials. *Eur. J. Pharm. Biopharm.* **40**, 1

Anderson J M 2001 Biological responses to materials. *Annu. Rev. Mater. Res.* **31**, 81–110

Anderson J M *et al.* 1996 Inflammation, wound healing and the foreign-body response. In: Ratner B D *et al.* (eds.) *Biomaterials – An Introduction to Materials in Medicine, 2nd edn,* Academic Press, San Diego, CA, p. 296

Andrade J D, Hlady V 1986 Protein adsorption and materials biocompatibility - A tutorial review and suggested hypotheses. *Adv. Polym. Sci.* **79**, 1

Andrade J D, Gregonis D E, Smith L M 1985 In: Andrade J D (ed.) *Surface and Interfacial Effects of Biomedical Polymers.* Plenum Press, New York, Vol. 1, Chap. 2

Andrade J D, Hlady V, Jeon S I 1996 Poly(ethylene oxide) and protein resistance - Principles, problems, and possibilities. *Adv. Chem. Ser.* **248**, 51

Arnold M, Cavalcanti-Adam E A, Glass R, Blümmel J, Eck W, Kantlehner M, Kessler H, Spatz J P 2004 Activation of integrin function by nanopatterned adhesive interfaces. *ChemPhysChem* **5**, 383

Baronas-Lowell D, Lauer-Fields J L, Borgia J A, Sferrazza G F, Al-Ghoul M, Minond D, Fields G B 2004a Differential modulation of human melanoma cell metalloproteinase expression by alpha(2)beta(1) integrin and CD44 triple-helical ligands derived from type IV collagen. *J. Biol. Chem.* **279**, 43503

Baronas-Lowell D, Lauer-Fields J L, Fields G B 2004b Induction of endothelial cell activation by a triple helical alpha(2)beta(1) integrin ligand, derived from type I collagen alpha 1(I)496–507. *J. Biol. Chem.* **279**, 952

Baszkin A, Norde W 2000 *Physical Chemistry of Biological Interfaces*. Marcel Dekker Inc., New York

Becklake M R 1976 Asbestos-related diseases of lung and other organs – their epidemiology and implications for clinical practice. *Am. Rev. Resp. Disease* **114(1)**, 187

Beier M, Hoheisel J D 1999 Versatile derivatisation of solid support media for covalent bonding on DNA-microchips. *Nucleic Acid Res.* **27**, 1970

Ben Ouada H, Hommel H, Legrand A P, Balard H, Papirer E 1988 Organization of the layers of polyethylene oxide grafted with different densities on silica. *J. Colloid Interface Sci.* **122**, 441

Berndt P, Fields G B, Tirrell M 1995 Synthetic lipidation of peptides and aminoacids – monolayer structure and properties. *J. Am. Chem. Soc.* **117**, 9515

Biesalski M, Tu R, Tirrell M 2005 Polymerized vesicles carrying molecular recognition sites. *Langmuir* **21**, 5663

Biesalski M, Knaebel A, Tu R, Tirrell M 2006 Cell adhesion on a polymerized peptide-amphiphile monolayer. *Biomaterials* **27**, 1259

Brash J L, Horbett T A 1987 (eds.) *Proteins at Interfaces: Physicochemical and Biochemical Studies,* Advanced Symposium Series 343. American Chemical Society, Washington, DC

Bridger K, Vincent B 1980 The terminal grafting of poly(ethylene oxide) chains to silica surfaces. *Eur. Polym. J.* **16**, 1017

Clark G M 2006 The multiple-channel cochlear implant: The interface between sound and the central nervous system for hearing, speech, and language in deaf people – A personal perspective. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 791 [Review]

Cleland J L, Daugherty A, Mrsny R 2001 Emerging protein delivery methods. *Curr. Opin. Biotechnol.* **12**, 212

Dahm M, Ruhe J, Berchthold B, Prufer D, Prucker O, Chang B J, Wallrath A, Oelert H 2004 Ultrathin polymer monolayers for promotion of cell growth on bioprosthetic materials – Evolution of a new concept to improve long term performance of biologic heart valves. *Biomed. Mater. Eng.* **14**, 419

Dalton B A, McFarland C D, Underwood P A, Steele J G 1995 Role of the heparin-binding domain of fibronectin in attachment and spreading of human bone-derived cells. *J. Cell Sci.* **108**, 2083

Delamarche E, Michel B, Kang H 1994 Thermal stability of self-assembled monolayers. *Langmuir* **10**, 4103

Dmitrenko A V, Shadrina N E, Ivanchev S S, Ulinskaya N N, Volkov A M 1990 Polymer inorganic selective adsorbents for gas-chromatography produced by graft-polymerization. *J. Chromatogr.* **520**, 21

Dori Y, Bianco-Peled H, Satija S K, Fields G B, McCarthy J B, Tirrell M 2000 Ligand accessibility as means to control cell response to bioactive bilayer membranes. *J. Biomed. Mater. Res.* **50**, 75

Egert U, Schlosshauer B, Fennrich S, Nisch W, Fejtl M, Knott T, Muller T, Hammerle H 1998 A novel organotypic long-term culture of the rat hippocampus on substrate-integrated multielectrode arrays. *Brain Res. Brain Res. Protoc.* **2**, 229

Elbert D L, Hubbell J A 1996 Surface treatments of polymers for biocompatibility. *Annu. Rev. Mater. Sci.* **26**, 365

Falconet D, Csucs G, Grandin H M, Textor M 2006 Surface engineering approaches to micropattern surfaces for cell-based assays. *Biomaterials* **27**, 3044

Fields G B, Lauer J L, Dori Y, Forns P, Yu Y C, Tirrell M 1998 Proteinlike molecular architecture: Biomaterial applications for inducing cellular receptor binding and signal transduction. *Biopolymers* **47**, 143

Fittkau M H, Zilla P, Bezuidenhout D, Lutolf M P, Human P, Hubbell J A, Davies N 2005 The selective modulation of endothelial cell mobility on RGD peptide containing surfaces by YIGSR peptides. *Biomaterials* **26**, 167

van der Flier A, Sonnenberg A 2001 Function and interactions of integrins. *Cell Tissue Res.* **305**, 285

Fromherz P, Offenhäuser A, Vetter T, Weis J 1991 A neuron-silicon junction – a Retzius cell of the leech on an insulated-gate field-effect transistor. *Science* **252(5010)**, 1290

Geiger B, Bershadsky A 2001 Assembly and mechanosensory function of focal contacts. *Curr. Opin. Cell Biol.* **13**, 584

Giancotti F G 2000 Complexity and specificity of integrin signaling. *Nat. Cell Biol.* **2**, E13

Giancotti F G, Ruoslahti E 1999 Transduction – Integrin signaling. *Science* **285**, 1028

Goodsell D S 2004 *Bionanotechnology – Lessons from Nature*. Wiley-Liss, Hoboken, NJ

Gordienok N I, Freidin B G, Proskurina L S 1986 Autooxidation of diethylene glycol. *Russ. J. Appl. Chem.* **59**, 1441

Gawkrodger D J 1993 Nickel sensitivity and the implantation of orthopedic prostheses. *Contact Dermatitis* **28**, 257

Griffith L G 2000 Polymeric biomaterials. *Acta Mater.* **48**, 263

Harris J M (ed.) (1992) *Poly(Ethylene Glycol) Chemistry.* Plenum Press, New York

Hashimoto K, Fujisawa T, Kobayashi M, Yosomiya R 1982a Graft copolymerization of glass-fibers and its application. *J. Appl. Polym. Sci.* **27**, 4529

Hashimoto K, Fujisawa T, Kobayashi M, Yosomiya R 1982b Graft copolymerization of glass-fibers and its application. *J. Macromol. Sci. Chem.* **A18**, 173

Healy K E, Rezania A, Stile R A 1999 Designing biomaterials to direct biological responses. *Ann. NY Acad. Sci.* **875**, 24

Hefti J L, Epitaux M, Glauser D, Frankhauser H 1998 Robotic three-dimensional positioning of a stimulation electrode in the brain. *Comput. Aided Surg.* **3**, 1

Hern D L, Hubbell J A 1998 Incorporation of adhesion peptides into nonadhesive hydrogels useful for tissue resurfacing. *J. Biomed. Mater. Res.* **39**, 266

Hersel U, Dahmen C, Kessler H 2003 RGD modified polymers: biomaterials for stimulated cell adhesion and beyond. *Biomaterials* **24**, 4385

Horbett T A, Brash J L 1995 (eds.) *Proteins at Interfaces II: Fundamentals and Application*, Advanced Symposium Series 602.American Chemical Society, Washington DC

Hubbell J A 1999 Bioactive biomaterials. *Curr. Opin. Biotechnol.* **10**, 123

Humphries M J 1990 The molecular basis and specificity of integrin ligand interactions. *J. Cell Sci.* **97**, 585

Hynes R O 2002 Integrins: Bidirectional, allosteric signaling machines. *Cell* **110**, 673

Inagaki N 1996 *Plasma Surface Modification and Plasma Polymerization*. Technomic Publishing, Lancaster

Irvine D J, Ruzette A V G, Mayes A M, Griffith L G 2001 Nanoscale clustering of RGD peptides at surfaces using comb polymers. 2. Surface segregation of comb polymers in polylactide. *Biomacromolecules* **2**, 545

Jenney C R, Anderson J M 1999 Effects of surface-coupled polyethylene oxide on human macrophage adhesion and foreign body giant cell formation *in vitro*. *J. Biomed. Mater. Res.* **44**, 206

Jeon S I, Lee J H, Andrade J D, De Gennes P G 1991 Protein surface interactions in the presence of polyethylene oxide 1. Simplified theory. *J. Colloid Interface Sci.* **142**, 149

Kohli-Seth R, Orpello J M 2003 The future of bed-side monitoring. *Crit. Care Clin.* **16**, 557

Kopf A, Baschnagel J, Wittmer J, Binder K 1996 On the adsorption process in polymer brushes: A Monte Carlo study. *Macromolecules* **29**, 1433

Krenkler K P, Laible R, Hamann K 1953 Polyreactions on pigment surfaces 7. Reactions of polymers with chlorosilane end groups on silicon dioxide surfaces. *Angew. Makromol. Chem.* **53**, 101

Kumar A, Abbott N L, Kim E 1995 Patterned self-assembled monolayers and mesoscale phenomena. *Acc. Chem. Res.* **28**, 219

Kunz R, Anders C, Heinrich L, Gersonde K 1999 Investigation into the mechanism of bacterial adhesion to hydrogel-coated surfaces. *J. Mater. Sci. Mater. Med.* **10**, 649

Laible R, Hamann K 1980 Formation of chemically bound polymer layers on oxide surfaces and their role in colloidal stability. *Adv. Colloid Interface Sci.* **13**, 65

Lansdown A B G 1995 Physiological and toxicological changes in the skin resulting from the action and interaction of metal ions. *Crit. Rev. Toxicol.* **25**, 397

Lee S J, Lee S Y 2004 Microarrays of peptides elevated on the protein layer for efficient protein kinase assay. *Appl. Microbiol. Biotechnol.* **64**, 289

Lehmann T, Ruehe J 1999 Polyethyloxazoline monolayers for polymer supported biomembrane models. *Macromol. Symp.* **142**, 1

Llanos G R, Sefton M V 1993 Does polyethylene oxide possess a low thrombogenicity. *J. Biomater. Sci. Polym. Ed.* **4**, 381

Lodish H, Berk A, Zipursky S L, Matsudaira P, Baltimore D, Darnell J 2000 *Molecular Cell Biology*, 4th edn. Freeman, New York

Maheshwari G, Brown G, Lauffenburger D A, Wells A, Griffith L G 2000 Cell adhesion and motility depend on nanoscale RGD clustering. *J. Cell Sci.* **113**, 1677

Malkar N B, Lauer-Fields J L, Juska D, Fields G B 2003 Convenient synthesis of glycosylated hydroxylysine derivatives for use in solid-phase peptide synthesis. *Biomacromolecules* **4**, 518

Mardillovitch A, Kokkoli E 2004 Biomimetic peptide-amphiphiles for functional biomaterials: The role of GRGDSP and PHSRN. *Biomacromolecules* **5**, 950

Mardillovitch A, Craig J A, McCammon M Q, Garg A, Kokkoli E 2006 Design of a novel fibronectin-mimetic peptide-amphiphile for functionalized biomaterials. *Langmuir* **22**, 3259

Margalit E, Maia M, Weiland J D, Greenberg R J, Fuji G Y, Torres G, Piyathaisiere D V, O'Hearn T M, Liu W T, Lazzi G, Dagnielie G, Scribner D A, de Juan E, Humayun M S 2002 Retinal prosthesis for the blind. *Surv. Ophthalmol.* **47**, 335

Massia S P, Hubbell J A 1991 An RGD spacing of 440 nm is sufficient for integrin α, β-3-mediated fibroblast spreading and 140 nm for focal contact and stress fiber formation. *J. Cell Biol.* **114**, 1089

Mossman B T, Gee J B L 1989 Asbestos-related diseases. *New Engl. J. Med.* **320(26)**, 1721

Mould A P, Komoriya A, Yamada K M, Humphries M J 1991 The CS5 petide is a 2nd site in the IIICS region of fibronectin recognized by the integrin α-4-β-1-inhibition of α-4-β-1 function by RGD peptide homologs. *J. Biol. Chem.* **266**, 3579

Mrksich M, Whitesides G M 1996 Using self-assembled monolayers to understand the interactions of man-made surfaces with proteins and cells. *Annu. Rev. Biophys. Biomol. Struct.* **25**, 55

Pakalns T, Haverstick K L, Fields G B, McCarthy J B, Mooradian D L, Tirrell M 1999 Cellular recognition of synthetic peptide amphiphiles in self-assembled monolayer films. *Biomaterials* **20**, 2265

Pande G 2000 The role of membrane lipids in regulation of integrin functions. *Curr. Opin. Cell Biol.* **12**, 569

Peppas N A, Huang Y, Torres-Lugo M, Ward J H, Zhang J 2000 Physicochemical, foundations and structural design of hydrogels in medicine and biology. *Annu. Rev. Biomed. Eng.* **2**, 9

Petit V, Thiery J P 2000 Focal adhesions: structure and dynamics. *Biology of the Cell* **92**, 477

Pierschbacher M D, Ruoslahti E 1984a Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. *Nature* **309**, 30

Pierschbacher M D, Ruoslahti E 1984b Variants of the cell recognition site of fibronectin that retain attachment-promoting activity. *Proc. Natl. Acad. Sci. USA* **81**, 5985

Prucker O, Christian S, Bock H, Rühe J, Frank C W, Knoll W 1998 On the glass transition in ultrathin polymer films of different molecular architecture. *Macromol. Chem. Phys.* **199**, 1435

Prucker O, Stöhr T, Murata H, Toomey R, van Os M T, Förch R, Knoll W, Rühe J 2003 Novel pathways towards tailor-made surfaces. *Eur. Coatings J.* **40**, 30

Pyun J, Kowalewski T, Matyjaszewski K 2003 Synthesis of polymer brushes using atom transfer radical polymerization. *Macromol. Rapid Commun.* **24**, 1043

Ramsay G 1998 DNA chips: State-of-the-art. *Nat. Biotechnol.* **16**, 40

Ratner B D, Bryant S J 2004 Biomaterials: Where we have been and where we are going. *Annu. Rev. Biomed. Eng.* **6**, 41–75

Rezania A, Healy K E 1999 Biomimetic peptide surfaces that regulate adhesion, spreading, cytoskeletal organization, and mineralization of the matrix deposited by osteoblast-like cells. *Biotechnol. Prog.* **15**, 19

Rodriguez F J, Ceballos D, Schuttler M, Valero A, Valderrama E, Stieglitz T, Navarro X 2000 Polyimide cuff electrodes for peripheral nerve stimulation. *J. Neurosci. Methods* **98**, 105

Rühe J, Knoll W 2002 Functional polymer brushes. *J. Macromol. Sci. Polym. Rev.* **C42**, 91

Ruoslahti E 1996 RGD and other recognition sequences for integrins. *Annu. Rev. Cell Dev. Biol.* **12**, 697

Ruoslahti E, Pierschbacher M D 1987 New perspectives in cell-adhesion – RGD and integrins. *Science* **238**, 491

Santini J T, Cima M J, Langer R 1999 A controlled-release microchip. *Nature* **397**, 335

Schoen F J, Levy R J 1999 Tissue heart valves: Current challenges and future research perspectives. *Biomed. J. Mater. Res.* **47**, 439

Scholl M, Sprossler C, Denyer M, Krause M, Nakajima K, Maelicke A, Knoll W, Offenhausser A 2000 Tissue heart valves: Current challenges and future research perspectives. *J. Neurosci. Methods* **104**, 65

Shen M C, Martinson L, Wagner M S, Castner D G, Ratner B D, Horbett T A 2002 PEO-like plasma polymerized tetraglyme surface interactions with leukocytes and proteins: *in vitro* and *in vivo* studies. *J. Biomater. Sci. Polym. Ed.* **13**, 367

Siegers C, Biesalski M, Haag R 2004 Self-assembled monolayers of dendritic polyglycerol derivatives on gold that resist the adsorption of proteins. *Chem. Eur. J.* **10**, 2831

Sigal G B, Mrksich M, Whitesides G M 1998 Effect of surface wettability on the adsorption of proteins and detergents. *J. Am. Chem. Soc.* **120**, 3464–73

Singer I, Barold S S, Camm A J 1998 *Nonpharmacological Therapy of Arrhythmias for 21st Century: The State of the Art*. Futura Publishing, New York

Southern E M 1996 DNA chips: Analyzing sequence by hybridization to oligonucleotides on a large scale. *Trends Genet.* **12**, 110

Staples M, Daniel K, Cima M G, Langer R 2006 Application of micro- and nano-electromechanical devices to drug delivery. *Pharm. Res.* **23**, 847

Szleifer I, Carignano M A 2000 Tethered polymer layers: phase transitions and reduction of protein adsorption. *Macromol. Rapid Commun.* **21**, 423

Tirrell M, Kokkoli E, Biesalski M 2002 The role of surface science in bioengineered materials. *Surf. Sci.* **500**, 61

Toomey R, Freidank D, Rühe J 2004 Swelling behavior of thin, surface-attached polymer networks. *Macromolecules* **37**, 882

Travis J 1993 Frontiers in biotechnology – biotech gets a grip on cell-adhesion. *Science* **260**, 906

Tsubokawa N, Kuroda A, Sone Y 1989 Grafting onto carbon-black by the reaction of reactive carbon-black having epoxide groups with several polymers. *J. Polym. Sci.* **A27**, 1701

Tsubokawa N, Hosoya M, Yanadori K, Sone Y 1990 Grafting onto carbon-black – reaction of functional groups on carbon-black with acyl chloride-capped polymers. *J. Macromol. Sci. Chem.* **A27**, 445

Voldman J, Gray M L, Schmidt M A 1999 Microfabrication in biology and medicine. *Annu. Rev. Biomed. Eng.* **1**, 401

Vukicevic S, Luyten F P, Kleinman H K, Reddi A H 1990 Differentiation of canacular cell processes in bone-cells by basement-membrane matrix components – regulation by discrete domains of laminin. *Cell* **63**, 437

Widrig C A, Chung C, Porter M D 1991 The electrochemical desorption of N-alkanthiol monolayers from polycrystalline Au and Ag electrodes. *J. Electroanal. Chem.* **310**, 335

Williams D 2003 Year-end perspectives: New technologies. *Med. Device Technol.* **14**, 10

Williams D F 1986 Definitions in biomaterials. *Proc. Consensus Conf. Eur. Soc. Biomater.*, London, Vol. 4.

Willner I, Willner B 2001 Biomaterials integrated with electronic elements: en route to bioelectronics. *Trends Biotechnol.* **19**, 222

Xu Z, Marchant R E 2000 Adsorption of plasma proteins on polyethylene oxide-modified lipid bilayers studied by total internal reflection fluorescence. *Biomaterials* **21**, 1075

Yang D F, Wilde C P, Morin M 1997 Studies of the electrochemical removal and efficient re-formation of a monolayer of hexadecanethiol self-assembled at an Au(111) single crystal in aqueous solutions. *Langmuir* **13**, 243

Yang Z H, Galloway J A, Yu H U 1999 Protein interactions with poly(ethylene glycol) self-assembled monolayers on glass substrates: Diffusion and adsorption. *Langmuir* **15**, 8405

Yasuda H 1981 Glow-discharge polymerization. *J. Polym. Sci. Macromol. Rev.* **16**, 199

Zajac R, Chakrabarti A 1995 Irreversible polymer adsorption from semidilute and moderately dense solutions. *Phys. Rev. E* **52**, 6536

Zamir E, Geiger B 2001 Components of cell-matrix adhesions. *J. Cell Sci.* **114**, 3583

Zhao B, Brittain W J 2000 Polymer brushes: surface-immobilized macromolecules. *Prog. Polym. Sci.* **25**, 677

Zhu H, Snyder M 2003 Protein chip technology. *Curr. Opin. Chem. Biol.* **7**, 55

Biographies

Jürgen Rühe studied chemistry at the Universities of Münster and Mainz, where he received his Ph.D. degree in 1989 for work carried out at the Max-Planck-Institute for Polymer Research. After a postdoctoral stay at the IBM Research Laboratories in San Jose, California, he returned to Germany and in 1991 started to work at the University of Bayreuth as a 'Liebig'-fellow of the VCI and a fellow of the German Research Council (DFG) in the area of polymers at interfaces. In 1995 he completed his Habilitation and rejoined the MPI of Polymer Research where he held an associate professor position. In October 1999 he accepted a full professor position at the Department for Microsystems Engineering (IMTEK) of the University of Freiburg, where he heads the Laboratory for Chemistry and Physics of Interfaces. He has been visiting scientist at UCSB, Georgia Tech, RIKEN (Japan) and Cambridge University (UK). In 1999 he received the chemistry award of the Academy of Sciences, Göttingen, and in 2001 the Max-Buchner Award of the DECHEMA society. His research interests are focused on the development of new methods for the generation of tailor-made surfaces and micro- and nanostructured thin films, as well as towards the application of such surfaces in various fields of microsystems engineering, especially for chip-based bioanalytical devices and biomedical technologies.

Markus Biesalski studied chemistry at the University of Mainz. In 1999 he received his Ph.D. from the University, Mainz, working at the MPI for Polymer Research. From 2000–2002 he was a postdoctoral fellow in Chemical Engineering and Materials, at the University of California, Santa Barbara. In 2002 he returned to Germany as an Assistant Professor, working at the Department for Microsystems Engineering (IMTEK, University of Freiburg). In July 2002 he received a young investigator award of the German Research Council (DFG), and since then he is heading an 'Emmy-Noether Research Group' at the IMTEK. Since 2004 he is a member of the 'Eliteförderprogramm' of the Landesstiftung Baden-Württemberg, Germany. His research interests focus on the design of molecularly defined peptide-polymer hybrid materials for the use in technological and biomedical applications, and as biophysical model systems.

1.06 Harsh Environment Materials

Erhard Kohn, University of Ulm, Ulm, Germany

1.06.1 Harsh and Hostile Environments

Harsh and hostile conditions are found in various environments and may be firstly associated with high temperature or aggresive media. But they are also found in extreme conditions like high pressure and shock cycling, during exposure to high-energy radiation, or while monitoring chemical reactions.

Of course, the material used in these environments needs to withstand these conditions and the materials associated with applications in harsh environment are ceramics, refractory metals, and highly stable oxides often in heterogeneous configurations. Transducer functions have been realized by a combination of materials like ceramics serving as passive, highly insulating substrate and refractive metals for the electrically active elements. In microsystems, the choice are semiconductors for two reasons: firstly, the modulation of electrons and holes over many decades results in high efficiencies and secondly, the monolithic integration with a system allows direct conditioning of sensors signals enabling smart sensors, or allows direct integration of driver circuits. Therefore, for 30 years microsystems have been designed around semiconductors, namely Si. Again, mainly two reasons have made Si-based microelectromechanical systems (MEMS) successful. Firstly, it is a fact that the processing technology of microsystems can be fully compatible with the complementary metal oxide semiconductor (CMOS) technology, to a large extent the MEMS technology being an extension of the microelectronics technology into the third dimension. Secondly, there is the availability of large surface areas with high quality of the materials. Thus, Si-based MEMS are found in automotive electromechanical applications, in printing with the electorthermal or piezoelectric ink-jet, in mobile communication as radiofrequency (RF) MEMS, and in biochemistry as biochip. Nevertheless, Si is limited by its materials properties. At high temperature above approximately 400°C, its high intrinsic carrier concentration will prevent it from using its semiconducting properties (Colinge 1998). It will also loose its elasticity, eventually giving way to plastic deformation. In liquids it may corrode or be etched like in hot KOH. Under high-energy radiation particle bombardment, the crystal properties may be degraded irreversibly.

Applications, which cannot be used with Si-based devices, are found in many different areas, however, are always confined to very specific cases. They are therefore to date not developed for a mass market. However, this may change in the future, when high-energy processes need to be controlled more and more efficiently. An example is sensing in combustion and jet engines in excess of 300°C, or monitoring and cleaning of exhaust gases at even higher temperatures. Here high-temperature stability is required as well as chemical inertness. Electrodes can be used for water disinfection, without the need of hazardous chemicals. In wastewater treatment highly corrosion-resistant electrodes, which can trace heavy metals or oxidize poisonous molecules, are preferred. In high-energy physics experiments, nuclear power plants, and space missions, radiation-resistant materials with a long life to damage are needed, which can be used for high-energy particle and neutron detection. This short summary tells already that various different materials properties matter, besides the semiconducting characteristics. It should also be mentioned that in a semiconductor with a bandgap of 5.5 eV like diamond the intrinsic carrier concentration at 1000°C is approximately that of Si at room temperature (RT). High-temperature operation of semiconductors will therefore most likely not be limited by their ideal electronic properties.

Nevertheless, the ability of transistors and diodes to tolerate harsh environments, like a high temperature, is a strong indication for the usefulness of the semiconductor materials system in question. Although the bandgap will determine the ultimate temperature limit of its electronic properties, the chemical stability will be the limit to degradation. The materials discussed here are therefore wide bandgap semiconductors with high chemical stability, sometimes called

hard or ceramic-like semiconductors and will center around SiC, GaN, AlN, and diamond, where especially GaN-based devices are heterostructures in a materials matrix of (In, Al, Ga)N. Table 1 lists the highest temperature, at which devices realized in these materials systems have been operated.

On the one hand, this chapter will concentrate on the transducer properties of such highly stable semiconductor materials and their supplementary materials like contact metals and dielectrics for passivation, namely refractory metals and metal oxide dielectrics. On the other hand, all four semiconductors are still rather restricted in their use as electronic material, especially lacking the possibility of an effective CMOS-like configuration. Thus, attempts for systems integration and packaging will also be considered.

For microsystems, wafer processing and bulk as well as surface machining technologies need to be developed, borrowing heavily from the Si-MEMS fabrication routines. Thus, all major device structures developed in SiC, GaN, AlN, and diamond have also been realized on Si substrates.

1.06.2 Materials Requirements

Because sensing is in general a very heterogeneous field, in the following, various materials aspects are discussed in comparison to Si and other alternative materials.

1.06.2.1 Overview of General Materials Properties

An overview of the electromechanical properties of materials suitable for MEMS technologies in harsh environment is given in Table 2.

In microsystems the materials are either bulk substrates or thin films. As bulk substrates, they are mainly single crystal and the ideal material parameters may apply. As thin films, they are often deposited onto foreign substrates, which are highly mismatched. These are therefore often polycrystalline, nanocrystalline, or amorphous. Then, the materials properties may deviate essentially from the ideal case. In harsh and extreme environments, however, the material should display close to the ideal properties, and amorphous films often do not meet this requirement. Therefore, they will be discussed only occasionally. Films used in microsystems must be smooth enough to allow lithography with micrometer size resolution. If not grown single crystalline, they must therefore either be polished or grown nanocrystalline.

SiC, on the one hand, has been used mainly because of its high mechanical strength and high-temperature stability and is the only material, which has already been implemented in a number of commercial high-temperature sensors. Diamond, on the other hand, possesses many exceptional and good properties, but is difficult to process and the development has only led to few device applications up to now, mainly in particle detection and electrochemistry.

GaN-based heterostructures have been used mainly as deep ultraviolet (DUV) and radiation detectors. But it has for a long time lacked a substrate, which can be bulk machined. Only recently with GaN-on-Si first results have become available concerning piezoelectric applications. Semiconducting AlN is still in the early stage of its materials development, and microsystems based on the semiconducting properties of AlN have not yet been reported.

1.06.2.2 Mechanical Properties

SiC and diamond are both materials with a large Young's modulus (YM). Therefore, high forces need to be applied for deflection of a beam or membrane; the limit being the fracture strength. The ratio of YM

Table 1 Maximum temperature of operation of wide bandgap semiconductor devices

Si	MOSFET	400°C	Colinge (1998)
GaAs	HEMT	500°C	Schmid *et al.* (1998)
SiC	MOSFET	650°C	Palmour *et al.* (1991)
SiC	HBT	580°C	Perez *et al.* (2003)
SiC	MISiCFET	1000°C	Spetz *et al.* (2001)
AlGaN/GaN	HEMT	750°C	Daumiller *et al.* (1999)
InAlN/GaN	HEMT	1000°C	Medjdoub *et al.* (2006)
AlN	Buffer isolation	1000°C	Neuberger *et al.* (2007)
Diamond	Diode	1050°C	Zimmermann *et al.* (2005a)

HBT, heterojunction bipolar transistor; HEMT, high electron mobility transistor; MISiCFET, SiC MOSFET; MOSFET, metal oxide semiconductor field effect transistor.

Table 2 Overview of microelectromechanical (MEMS) materials properties

	Density (g cm^{-3})	*Young's modulus (GPa)*	*Hardness (GPa)*	*Thermal conductivity (W mK^{-1})*	*CTE (ppm K^{-1})*
Si	2.33	170	13	130	2.6
Diamond	3.52	1143	90	2000	1
NCD	3.52	800–1000	~80	100–400	1
AlN	3.26	331	11.8	175	4.6
GaN	6.15	181	10.8	130	3.6–5.6
(3c)SiC	3.21	300–700	~30	360	3.8
(6h)SiC	3.21		~30	490	4–6
Sapphire	*3.98*	*345*	*40*	*46*	*5.8*

	Resistivity (intrinsic) (Ω cm)	*Breakdown field (10^8 V m^{-1})*	*Relative dielelectric constant*	*Bandgap (eV)*	*Phase stability (°C)*
Si	10^4	0.5	12	1.1	1412, melting
Diamond	10^{12}–10^{16}	10	5.7	5.4	750 in atm. 2200 in H_2
NCD	10^8	10	5.7	5.4	750 (in atmosphere)
AlN	$>10^{14}$	1.8	8.9	6.2	>2200
GaN	$>10^8$	5	8.9	3.2	650, atmosphere[a]
(3c)SiC	10^2–10^6	1	9.7	2.4	3070, Sublimation in
(6h)SiC	$>10^8$	3–5	9.7	2.9	inert atm.
Sapphire	$>10^{14}$	0.48	9.4	9.9	–

[a]GaN decomposition temperature in atmosphere. If capped by AlGaN or InAlN, higher temperatures are possible (Pisch A *et al.* 1998).
NCD, nanocrystalline diamond.
Source: Internet libraries: www.webelements.com, www.memsnet.org/material/environmentalchemistry.com/Yogi/periodic/, www.ioffe.rssi.ru/SVA/NSM/Semicond.

and fracture strength is therefore an important figure of merit.

For SiC, a YM of 300–700 GPa is obtained depending on its quality (Fu *et al.* 2005).

For GaN epitaxial films, values of the YM between 250 and 350 GPa have been determined by nanoindentation (Jian *et al.* 2003).

For diamond nanocrystalline films, the YM is mostly between 800 and 1000 GPa, and the fracture strength is between 2 and 4.5 GPa (Espinosa *et al.* 2003, Hernandez-Guillen *et al.* 2005). Thus, while a near ideal stiffness can be obtained, the fracture strength is less than half of the single crystal value (**Table 2**). On the other hand, no plastic range and no hysteresis are observed, when bending cantilevers up to fracture. Plastic deformation is only seen with very high forces inflicted by nanoindentors used to determine the hardness. The high elasticity combined with the high fracture strength may be illustrated in **Figure 1**.

Another important feature is the thermal stability of the mechanical properties. For SiC ceramics the YM has been measured up to 1000°C. It is nearly constant up to approximately 600°C, slowly decreasing for higher temperatures (Wolfenden 1997). With respect to diamond, no change has been observed in direct measurements up to 600°C (Werner *et al.* 1997a), and changes are also not observed indirectly up to at least 650°C, inferred from a constant threshold

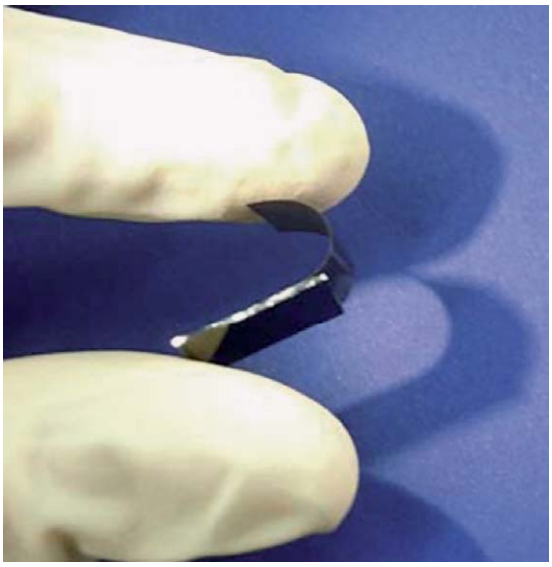


Figure 1 Bending of a freestanding diamond cantilever of approximately 20-μm thickness.

voltage of electrostatically driven cantilever switches (Adamschik *et al.* 2000a) (see also Section 1.06.6.2). In contrast, at this temperature, deviations from Hook's law can already be expected for Si.

A third aspect is the thermal expansion or contraction between growth temperature and the temperature of operation. This may be especially important for membrane structures on Si or SiC, which are operated within a large temperature range or for thermal–electric (bimetal) and piezoelectric actuator stacks. As the coefficient of thermal expansion of all materials is not constant for the temperature range in question, membranes may experience compressive as well as tensile stresses in certain temperature ranges (which on the other hand may be balanced by built-in stresses in the film). **Figure 2** shows the difference in thermal stress between RT and 1200°C for Si, SiC, AlN, and diamond materials combinations. Growth temperature for diamond chemical vapor deposition (CVD) film deposition may be between 600°C and 900°C; that of CVD 3C–SiC between 1100°C and 1300°C as discussed below. The stress between diamond and Si is less than 1 GPa and below the fracture strength of both the materials. Thus, thick crack-free diamond/Si stacks can be realized. However, thick diamond films on Si often lead to a concave wafer bowing and a high dislocation density in the Si substrate, which is mostly caused by a vertical stress profile built into the film during growth (von Kaenel *et al.* 1997).

1.06.2.3 Thermal Properties

Thermal conductivity and heat capacity are important for the dissipation of thermal losses generated during operation under heavy load. In the case of the bubble ink-jet, the thermal properties may determine the overdrive capability or frequency limit of operation. In case of a bimetal actuator, it may limit the speed of recovery.

Two aspects matter: firstly, the direct heat transport through the microstructure and, secondly, the heat extraction through the substrate. Therefore, both the thin-film properties and the bulk properties are important. In general, the thermal conductivity will be reduced with smaller grain size, and nanocrystalline materials have to be evaluated carefully with respect to their granular structure. In **Figure 3** the thermal conductivity is plotted for ideal, single-crystal diamond, AlN, GaN, SiC, and Si, and compared with Cu. Because the thermal conductivity of Si is lower than that of SiC and diamond, but is still the substrate in most cases, it is important to design the thickness of the top SiC or diamond layers according to the heat spreading requirements.

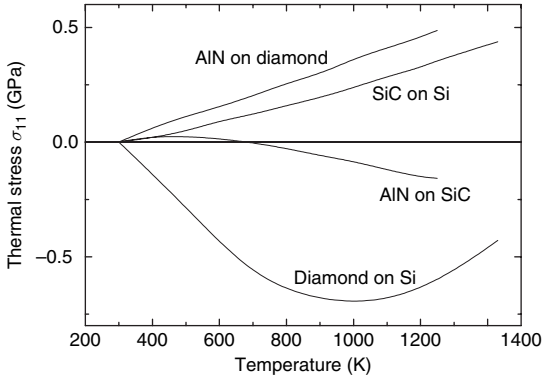


Figure 2 Thermal stress developed for diamond on Si, AlN on diamond, SiC on Si, and AlN on SiC, which may develop between growth at high temperature and temperature of operation. (Source: Aleksov A, Kubovic M, Kasu M, Schmid P, Grobe D, Ertl S, Schreck M, Stritzker B, Kohn E 2004. Diamond-based electronics for RF-applications. *Diamond Relat. Mater.* **13**, 233–40; Slack G A, Bartram S F 1975 Thermal expansion of some diamond like cystals. *J. Appl. Phys.* **46**, 89–98; Okada Y, Tokumaru Y 1984 Precise determination of lattice parameters and thermal expansion coefficient of silicon between 300 and 1500 K. *J. Appl. Phys.* **56**, 314–20.)

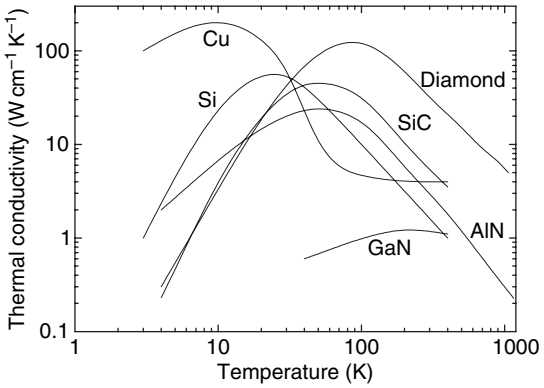


Figure 3 Thermal conductivity of SiC, GaN, AlN, and diamond in comparison to Si and Cu. (Source: Siechel E K, Pankove J I 1975 Thermal conductivity of GaN. *J. Chem. Solids* **38**, 330; Slack G A, Tanzilli R A, Pohl R O, Vandersande J W 1987 The intrinsic thermal conductivity of AlN. *J. Chem. Solids* **48**, 641–47; Glassbrenner C J, Slack G A 1964 Thermal conductivity of silicon and germanium from 3°K to the melting point. *Phys. Rev.* **134**, A1058–69; Slack G A 1964 Thermal conductivity of pure and impure silicon, silicon carbide, and diamond. *J. Appl. Phys.* **38**, 3460–6.)

From **Figure 3** it may be seen that diamond is most effective for a heavy duty operation around RT, but the difference to Cu becomes less important for an operation in a hot environment. Aspects of the chemical high-temperature stability are discussed in Section 1.06.2.4.

1.06.2.4 Chemical Properties

The chemical stability may be considered in three different ways, the chemical stability of the bulk material with respect to phase stability and decomposition, the chemical stability of the surface with respect to corrosion and etching, and the ability to form reliable interfaces, for example, for adhesion of molecules.

1.06.2.4.1 Silicon carbide

SiC is an exceptional stable material subliming at approximately 1800°C. It can only be etched in KOH at around 600°C and is thus stable at the etching conditions for the Si substrate (in KOH at approximately 70°C). Dry etching can be performed in F-based plasma chemistries as also used for Si, however, with lower etch rate. The surface is oxidized to SiO_2 (also forming CO) and the passivation behavior is similar to that of Si, however, and is stable up to high temperatures. The electrochemical stability is higher than that of Si, but it can be etched electrochemically in HF (Mikami *et al.* 2005).

1.06.2.4.2 GaN, GaN heterostructure, and AlN

GaN dissociates in an atmosphere at approximately 650°C, forming Ga oxide (mainly Ga_2O_3) and N_2 (Pisch *et al.* 1998), and Ga_2O_3 itself is stable up to a high temperature above 800°C (Fleischer and Meixner 1992). The stability is still improved, when it is capped with AlGaN and InAlN. Indeed, AlGaN/GaN high electron mobility transistors (HEMTs) have been operated up to approximately 750°C and InAlN/GaN HEMTs up to 1000°C (in vacuum) without loosing their spontaneous polarization and piezopolarization (Daumiller *et al.* 1999, Medjdoub *et al.* 2006). GaN can have two surfaces, Ga face and N face, with a highly stable Ga face and a rather reactive N face, respectively. Usually, device films are grown with a Ga face. In this case, in atmosphere, the top layer is terminated with oxygen.

AlN is usually deposited as an amorphous or nanocrystalline film. It can be etched in KOH, depending on the crystalline quality. It is oxidized above 800°C with a higher stability of the Al face than the N face (Gu *et al.* 2005). The etching of GaN-based films is discussed in Section 1.06.5.3.

1.06.2.4.3 Diamond

Diamond is the metastable form of carbon with respect to graphite and its phase stability must be limited. Nevertheless, the barrier for the phase transition is very high and the survival of the diamond phase has been documented at 2200°C in H_2 atmosphere (Kumar 1996). In an inert atmosphere graphitization is observed above 1500°C; however, with respect to nanodiamond powder, the process already takes place above 800°C (Chen *et al.* 1999). Electronic device operation has been demonstrated at 1050°C (Zimmermann *et al.* 2005a). In the liquid, it is highly corrosion resistant and cannot be etched in any wet solution. It can, however, be etched in an oxygen plasma and in an oxygen atmosphere above approximately 750°C. Dry etching in an oxygen plasma is therefore the common way to pattern diamond thin films (Ding *et al.* 2005).

1.06.2.5 Radiation Detector Properties

The field of wide bandgap semiconductors is the solar blind detection of UV radiation and the detection of high-energy particles. For high-energy photon and particle detection, radiation hard semiconductors are needed to replace current Si detectors, which possess a limited lifetime under high-energy fluences (Bruzzi *et al.* 2002). The wavelength cutoff is determined by the bandgap and reaches from 2.6 to 3.3 eV for SiC (depending on the polytype), 3.5 eV for hexagonal GaN, 5.45 eV for diamond, and 6.2 eV for AlN. SiC and diamond are indirect, whereas GaN and AlN direct bandgap semiconductors. The bandgap of AlGaN can be tuned between that of GaN and AlN depending on the composition and indeed AlGaN/GaN heterostructures are the mostly used ones for UV detectors. However, GaN does not possess a natural substrate and growth on Al_2O_3 or SiC and results in a high mismatch and a high dislocation density in the GaN buffer layer. AlGaN with high Al content is difficult to growth with low oxygen content and defect density.

On the other hand, SiC is a mature substrate technology developed for high-temperature electronics. Thus, active layers on hexagonal SiC substrates have become available with low defect density and

SiC starts to compete with Si detectors in the field of X-ray and particle detection (Bruzzi *et al.* 2002).

The basis for diamond detectors have mostly been polycrystalline CVD films with large grain size. Recently, freestanding single-crystal CVD films with a background impurity concentration of only $10^{13}\,cm^{-3}$, an associated electron mobility of $4600\,cm^2\,V^{-1}\,s^{-1}$, a hole mobility of $3800\,cm^2\,V^{-1}\,s^{-1}$, and photo carrier lifetime in the microsecond range have become available (Isberg *et al.* 2002). Therefore, diamond is an alternative for DUV detection below the 225-nm wavelength. However, the substrate size is still limited to approximately 1 cm^2. An overview can be found in Bergonzo *et al.* (2004).

In general, the materials must posses a large charge collection efficiency, boiling down to a high mobility–lifetime ($\mu\tau$) product and collection distance ($\mu F\tau$). In addition, deep level charging and related drift need to be controlled, which involves priming in some cases. Finally, the atomic displacement energy as measured for radiation damage should be high. **Table 3** shows a comparison of some key parameters.

1.06.2.6 Electrothermal Properties

The electrothermal properties may be used in sensor as well as actuator elements. In sensing, the temperature-dependent change in resistance is used in thermistors. The heat generated in a resistor may on the other hand be used in spot heaters like in ink-jets or in a bimetal type of actuator. They make use of the actuation of deep dopants, the barrier characteristics of grain boundaries in polycrystalline materials, and the change of carrier mobility with temperature, mostly in a highly disordered amorphous system. In SiC, sputtered SiC has been used, most likely taking advantage of a disorder and an activation effect (Dulloo *et al.* 1999). In diamond, boron is a deep acceptor and its thermal activation can be used.

SiC and diamond have also been employed as heaters. Here a temperature-independent behavior is desired. But many doping levels in wide bandgap semiconductors are deep. For diamond no shallow doping exists. However, at high concentrations, the activations energy can be essentially reduced by miniband formation (Borst and Weis 1996, Werner *et al.* 1997b). Such resistive heaters have been used in microhotplates and spot heaters. For operation under high thermal load or in aggressive media also, the chemical corrosion resistance, the resistance to cavity damage, and the resistance to fatigue due to thermal cycling are important. To be used in electrothermal actuators, based on the bimetal effect, both SiC (Jiang *et al.* 2006) and diamond (Schmid *et al.* 2003) have been used as materials with low coefficient of expansion combined with a metal of high coefficient of expansion (**Table 4**). Such pairs are diamond/Ni or SiC/NiCr. Here, either the semiconductor or metal can be chosen as a resistive heater element.

1.06.2.7 Electrochemical Properties

The heart of all electrochemical devices are their electrodes. Two different types of electrodes are used. Firstly this is the metal oxide semiconductor (MOS) diode with a chemically active surface, mostly hydroxyl groups on metal oxides like TiO_2. Here potential differences are detected via field effect. These may be related to work function differences generated by redox reactions at the surface active sites or surface charging due to the attachment of charged molecules. This method is mostly used in biochemical analysis with MOS chemical field effect

Table 3 Materials parameters relevant for particle detection – the Wigner energy is the energy needed to create a site vacancy in the periodic lattice

	Material		
	Si	*4H SiC*	*Diamond*
Bandgap (eV)	1.12	3.3	5.45
Breakdown field (V cm^{-1})	3×10^5	4×10^6	1×10^7
Saturated velocity (cm s^{-1})	0.8×10^7	2.0×10^7	2.7×10^7
Mininimum carrier lifetime (s)	2.5×10^{-3}	5×10^{-7}	$\sim 10^{-5}$
e–h creation energy (eV)	3.6	8.4	13
Wigner energy (eV)	13–20	25	43

Source: Bruzzi *et al.* 2002 *Nucl. Instrum. Meth. Phys. Res. A* **485**, 172–7.

Table 4 Electrothermal properties of selected materials

	NCD	*3C SiC*	*GaN*	*Ni*	*Cu*
Coefficient of thermal expansion, α (ppm K^{-1})	~1	3.8	3.2–5.6	13.4	16.5
Elasticity modulus, E (GPa)	700–1000	307	181	210	130
Thermal conductivity, σ_{th} (W (m K)$^{-1}$)	100–400	360	130	91	400
Heat capacity, c (J (g K)$^{-1}$)	–	0.69	0.49	0.44	0.385
$E\alpha$ (kPa K^{-1})	700–1000	1170	580–1070	2815	2145

NCD, nanocrystalline diamond.

transistors (ChemFETs) or ion-sensitive field effect transistors (ISFETs) (Bergovel 2003) within a rather narrow range of pH.

Secondly this is the conductive electrode in which a DC current is passed and a redox reaction can be sustained. This is the case in a battery, where the electrode material itself also takes part in the reaction. Another case are inert electrodes supporting electron transfer between the liquid and the solid without etching or corrosion, the classical example being Pt. They are mostly used in electroanalytical chemistry but also in electrooxidation like in water disinfection. An important aspect of such electrodes is the water dissociation reaction. Water is dissociated into H_2 and O_2 at the cathode and anode, respectively, with a dissociation potential of $\Delta V = 1.23\,V$ (potential window for water dissociation). In the reaction, electrons or holes may need to be transferred, with high density available in a metal. In a semiconductor, three possibilities exist for charge transfer: electron transfer via conduction band, hole transport via valence band, and charge transfer via surface states. The energy/potential scale is illustrated in **Figure 4**, where the bandgap lineup of wide bandgap semiconductors is plotted with reference to the potential of the standard hydrogen electrode (SHE) potential (for diamond, see Angus *et al.* 1999). The electron affinity of diamond is positive for an oxygen-terminated surface and negative for hydrogen termination. The electron affinity of AlN is still under debate, depending on the doping level and status of surface oxidation.

However, to conduct current, the semiconductor surface should not be depleted and the semiconductor should therefore be quasi-metallically doped, allowing tunneling across the interface (from the semiconductor side). Thus, the dissociation reaction may still be suppressed, if the transfer states cause band bending preventing tunneling, if the density of active sites is low, and if the surface is shielded from the electrolyte by adsorbed species. However, within

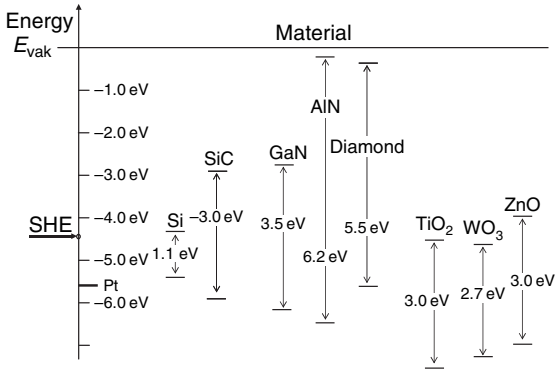


Figure 4 Approximate band structure lineup for various semiconductors, metal oxides, and Pt. The metal oxide materials are common in metal oxide semiconductor field effect transistors (MOSFETs), and are therefore not used in amperometric applications. SHE, standard hydrogen electrode.

this potential window, the electrode surface may still be active with respect to other redox reactions. At such an electrode with wide potential window, species can react at high overpotentials, which are otherwise stable in water. This is the basis of electrooxidation of hazardous organics like phenol. Typical ranges of potential windows are shown in **Figure 5**. Clearly, GaN and diamond show the highest potential windows, and also the advantage of diamond with respect to glossy carbon becomes obvious. However, the stability of GaN is less than that of diamond (discussed in Section 1.06.4.1.4).

1.06.3 Materials Synthesis and Characteristics

Depending on their use, the synthesis of bulk material and thin films is widespread from sputtering of nanocrystalline films to the growth of high-quality single layers by metal organic chemical vapor deposition (MOCVD) and the synthesis of substrates by high-temperature processes like sublimation growth.

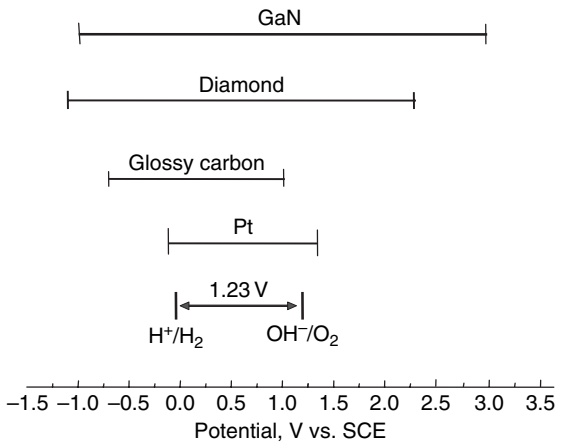


Figure 5 Potential window of water dissociation in H_2SO_4 (pH 1) for various electrode materials and the water reaction itself. The bars represent average values for currents in the microampere per square centimeter range. SCE: standard calomel electrode.

1.06.3.1 Silicon Carbide

SiC has been deposited as single crystalline, polycrystalline, and amorphous material. It has many possibilities of a stacking order of the SiC tetrahedrons and thus a high number of polytypes. The most commonly used are the cubic 3C–SiC phase and the H4 and H6 hexagonal phases. This variety is mainly reflected in the energy bandgap and the free carrier transport properties but not so much in the mechanical, thermal, and electrochemical properties. The material sublimes at around 1800°C and would (if capped) melt at 3800°C.

As amorphous material, it can be deposited at a low temperature by sputtering or ion beam deposition from a SiC target. These films are often Si rich. Films deposited by plasma-enhanced CVD (PECVD), electron cyclotron resonance CVD (ECR-CVD), atmospheric pressure CVD (APCVD), low-pressure CVD (LPCVD), and CVD under ultrahigh vacuum conditions (gas source molecular beam epitaxy (MBE)) mostly using CH_4 or C_3H_8 and SiH_4 as precursors, are amorphous, nanocrystalline, or polycrystalline, depending on the growth temperature, the grain size increasing with increasing growth temperature. On SiO_2 or Si_3N_4, the crystallites are randomly oriented but may develop some texture. The growth temperature is usually between 1100°C and 1300°C, but may be as low as 200°C (Colder *et al.* 2005). Such films are used for freestanding structures on sacrificial layers like poly-Si, SiO_2, and Si_3N_4 and are rarely compatible with CMOS postprocessing technologies (Cheng *et al.* 2002). Thin buried SiC films can be generated by high-dose ion implantation ($D \sim 10^{17}$ cm^{-2}) into the Si substrate (Serre *et al.* 1997).

Single-crystal hexagonal SiC wafers are used for bulk devices, as high-quality 4H, and 6H substrates are available for electronic and optoelectronic device fabrication (Powell and Rowland 2002). Etching is restricted to dry etching (see Section 1.06.5.2).

SiC grown on Si develops in the 3C–SiC polytype. It is therefore used in surface machined devices on Si. It is also available freestanding in wafer size, however, with some texture (available from Hoya Advanced Semiconductor Technologies). On Si it is grown by APCVD, LPCVD, and MOCVD when metal organic precursors are used at a temperature above 1300°C. The heteroepitaxial deposition on Si starts with the carboration of the Si surface and growth of a thin nucleation layer (Morales *et al.* 2003). Nevertheless, due to the high temperatures employed, it is difficult to maintain a smooth interface to the Si substrate, which becomes pitted. Due to the high mismatch, nucleation on the carbonized layer results in mosaic growth and oriented texture after outgrowth as illustrated in **Figure 6**.

In addition, the high deposition temperature causes high thermal stresses, leading to the formation of a high density of crystal defects. Migration of those defects into the Si substrate will affect its electronic, mechanical, and chemical stability. It can, for

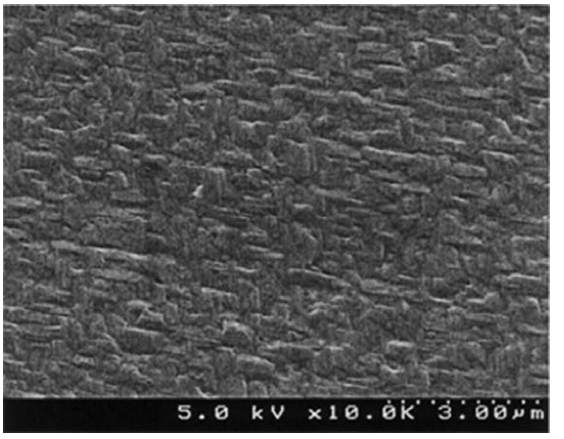


Figure 6 Surface topography of 3C–SiC layer grown by metal organic chemical vapor deposition (MOCVD) on (100)-oriented Si. (Source: Young D J, Du J, Zorman C A, Ko W H 2004 High-temperature single-crystal 3C–SiC capacitive pressure sensor. *IEEE Sens. J.* **4**, 464–70.)

example, affect the anisotropic Si wet etching characteristics essential for membrane etching. These effects are attenuated, when the film is grown on a silicon on insulator (SOI) substrate, where the top Si layer serves as a seed layer and the buried SiO_2 layer as a stress relaxation layer. The buried SiO_2 layer will also serve as a thermal barrier between the surface structure and the substrate, which is important for electrothermal actuators, and where heat loss into the substrate should be avoided. It is also generating electrical isolation between the SiC layer and the Si substrate. This is important for high-temperature application, where the Si substrate will become metallic in conduction. Layers grown on amorphous films like SiO_2 are nanocrystalline or polycrystalline. Thick films are therefore often polished.

The films can be n-type doped with N, P, and O, and p-type doped with Al or B. Nitrogen exhibits an activation energy of 50–58 meV, reduced to zero by miniband formation at approximately $3.6 \times 10^{19}\,\mathrm{cm}^{-3}$ (Eickhoff *et al.* 2004). The incorporation seems not to be influenced significantly by the crystalline quality. Oxygen is thought to be incorporated with an activation energy of approximately 200 meV. Both nitrogen and oxygen may be present, when films are grown on Si_3N_4 or SiO_2 and outdiffused into the SiC, causing residual conduction. Oxygen is thought to decorate defects and grain boundaries preferentially (Eickhoff *et al.* 2004). Thus, high residual conductivities are often observed in nanocrystalline layers.

The deposition by LPCVD is very similar to MOCVD or MBE. In the case of MBE, the deposition is mainly driven by the kinetics and deposition temperatures are lower, mainly in the range of 500–1000°C (Lee *et al.* 1997, Lim *et al.* 2004).

In contrast, most bulk single-crystal substrates have been developed for use in optoelectronics and high-power electronics, and are synthesized in hexagonal configuration (mostly 4H–SiC and 6H-SiC). Due to the high crystalline quality and low defect density (mainly micropipes and screw dislocations), films grown on these substrates can be semi-insulating with high thermal stability. They can be doped n-type and p-type as already mentioned above. However, SiC substrates are not easily etched (except with deep reactive ion etching (DRIE)) and they have, therefore, often been used in bulk devices like heater platforms, high-temperature gas sensors, and stress gauges including the entire wafer thickness as cantilever or membrane.

1.06.3.2 Gallium Nitride

Wurzite GaN grown with *a*-plane orientation contains a strong vertical charge dipole generated by the vertical N–Ga bond, which is not compensated by symmetry (**Figure 7**). This leads to a high bonded spontaneous polarization charge density at the surfaces of $\sigma_p = 3 \times 10^{13}\,\mathrm{cm}^{-2}$, and the direction of the field depending on the growth orientation. Normally obtained in MOCVD growth is the Ga face orientation, with a Ga-oxide terminated surface, when exposed to the atmosphere. Under specific conditions, the N face orientation can be obtained (Georgakilas *et al.* 2001). However, this surface is chemically less stable and therefore not frequently used.

It does not possess a natural substrate and bulk GaN is only used in proof-of-concept experiments. It is therefore mostly grown as film on foreign hexagonal substrates like Al_2O_3 and 6H–SiC. Several hundreds of micrometer thick quasi-substrates are grown by hydride vapor phase epitaxy (HVPE).

GaN is the most commonly used wide bandgap semiconductor material in optoelectronics and RF power electronics, and the materials synthesis technology has mainly been driven by these applications. The key has been a high concentration of (atomic) N radicals. Two CVD techniques are common, MOCVD and radiofrequency molecular beam epitaxy (RF-MBE). In the first case, metal organic precursors are used as a Ga source and NH_3 as a precursor for N^*, in the second case, N_2 is split in a low-energy RF plasma source. MOCVD is therefore performed mostly at around 1000°C close to the dissociation temperature of NH_3. Due to the high mismatch to

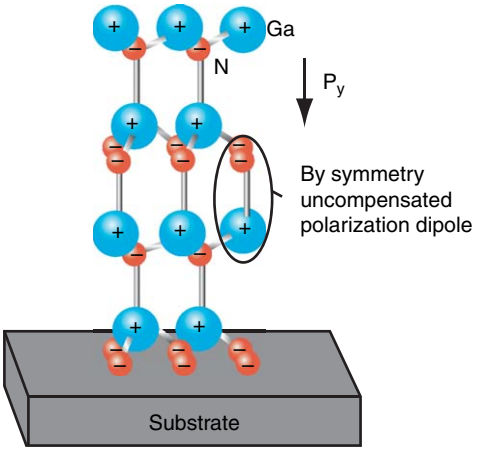


Figure 7 Origin of the spontaneous polarization in wurzite GaN.

the foreign substrate, direct epitaxial growth is not possible. Instead, a nucleation layer of low-temperature-deposited AlN (LT-AlN) with high amorphous content is used. Therefore, the first nanometers close to this nucleation layer are very defective and defects need to be outgrown subsequently.

In electronics and microsystems, the wurzite material phase can combine pyroelectric and semiconducting properties. Thus, diode and transistor piezoelectric sensor devices based on n-type as well as p-type GaN and AlGaN/GaN heterostructures have been investigated. In this case, three effects may contribute to the overall piezoresponse. First, this is the electroceramic effect, which is the variation of the surface charge dipole induced by the bonded polarization charge dipole. Second, the piezobulk effect caused by the band structure realignment and related charge redistribution in the bulk, and, third, in the presence of a stress gradient across the film the generation of free charge carrier in the bulk. The later effect has been called polarization doping and is observed in heterostructure FETs (Jena *et al.* 2002), but not well identified in piezo-MEMS structures yet.

Because this is an unusual effect, it may shortly be illustrated in **Figures 8** and **9**, showing the case of a doped GaN cantilever, which is bent by a force, acting on the tip. In **Figure 8**, the distribution of the bonded polarization charges is shown under stress, and in **Figure 9** the corresponding effect on the induced mobile charge distribution. The bending will induce a vertical stress gradient across the beam, generating a polarization gradient, which in turn will generate a distribution of free carriers (holes or electrons) across the beam. The corresponding donors or acceptors are represented by the stress-induced change in the bonded polarization charge distribution. Thus, a constant stress gradient will result in a constant free carrier profile. The effect has been first identified using a graded $Al_{(x)}Ga_{(1-x)}N$ layer with a vertical stress profile incorporated by growth. Thus, free carriers are generated without external doping. For a Ga face materials orientation, which is usually the case for MOCVD-grown layers on a LT-AlN buffer, upward bending of the cantilever will show p-type polarization doping in the bulk. Correspondingly, downward bending will result in n-type polarization doping, both generating bulk conduction paths in parallel to a piezoresistive channel on the surface.

Single-crystalline GaN films have been deposited onto a variety of substrates (Miskys and Kelly 2003), especially Al_2O_3, α-SiC, (111) and (100) Si, (111) GaAs (Kumagai *et al.* 2002), $LiNbO_3$ (Wu *et al.* 2002), and γ-$LiAlO_2$ (Takagaki *et al.* 2004a). Deposition on Al_2O_3 and SiC are common for optoelectronic and electronic applications. However, it is difficult to etch membrane structures to obtain freestanding cantilevers. Deposition on GaAs, $LiNbO_3$, and γ-$LiAlO_2$ shows rather poor quality and device grade material has not been obtained up to now. Nevertheless, it has been attempted to fabricate freestanding structures from these films. High-quality GaN layers can however be grown on (111)-oriented Si wafers. However, the large mismatch limits the thickness of the GaN layer to approximately 1 µm, before cracks appear. To circumvent this problem, graded AlGaN buffers (Kim *et al.* 2001) and AlN/GaN superlattices have been introduced (Feltin *et al.* 2001). Most successful was the incorporation of low-temperature-grown LT-AlN stress release layers into the GaN film, enabling the growth of thick GaN layers in excess of 6 µm (Krost and Dadgar 2002). Because AlN

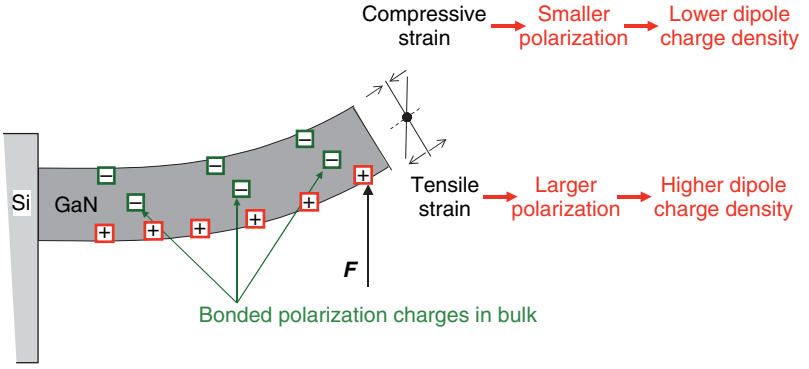


Figure 8 Change in bonded charge distribution (squares) while bending a cantilever upward, assuming that compressive strain will reduce the polarization. (Source: Neuburger M, Zimmermann T, Benkart P, Kunze M, Daumiller I, Dadgar A, Krost A, Kohn E 2004. GaN based piezo sesnsor. *62nd Device Research Conference (DRC)*, South Bend, IN, USA, Abstracts 45–6, June 2004.)

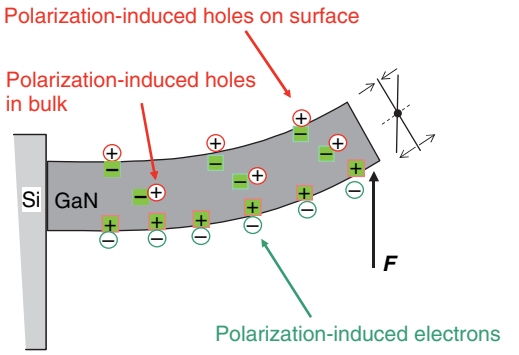


Figure 9 Compensation of bonded charge by the generation of a hole charge density (circles) in the bulk of the cantilever. This bulk effect has been named polarization doping.

contains a larger polarization field than GaN, polarization dipoles are introduced by this procedure. This is illustrated in **Figure 10**, which shows an atomic force microscope (AFM) potential scan across a GaN buffer on Si with two AlN stress release layers. It shows the growth sequence with the LT-AlN nucleation layer on the Si substrate and the strong dipoles of the AlN stress relief layer and an AIGaN/GaN FET channel, the dipole charge of which is compensated by the surface.

Most GaN structures are designed with strained AlGaN/GaN heterojunctions, making use of the piezopolarization at the interface, generating a 2D electron gas (2DEG) channel of high mobility. The various polarization charge contributions are illustrated in **Figure 11**.

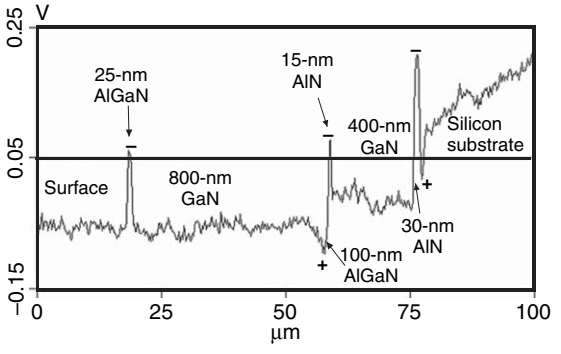


Figure 10 Measurement of polarization dipoles across a GaN heterostructure grown on Si, containing two AlN stress relief layers and an AlGaN/GaN high electron mobility transistor (HEMT) heterostructure. The measurement was performed with a potential contrast atomic force microscope (AFM) tip on a beveled cut. (Source: Dadgar A *et al.* 2003 Gallium-nitride-based devices on silicon. *Phys. Stat. Sol. (c)* **0**, 1940–9 with permission.)

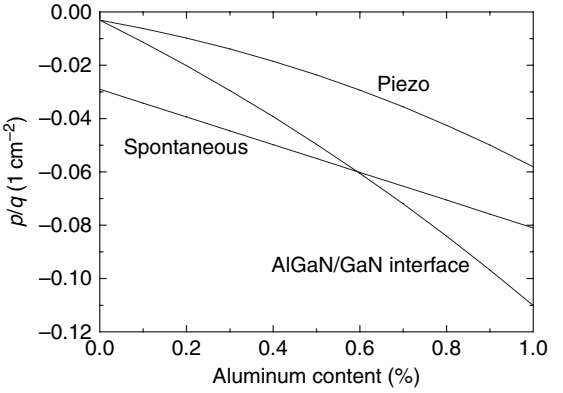


Figure 11 Contributions to the polarization discontinuity at the AlGaN/GaN interface, namely piezopolarization caused by the strained AlGaN layer and the spontaneous polarization of GaN.

This heterojunction is also the basis for FETs. Depletion mode as well as enhancement mode N channel FET structures have been realized using the interfacial 2DEG as channel, P channel devices are not attractive up to now, because of their low hole sheet charge density and mobility (Zimmermann *et al.* 2004). Nevertheless, front end signal processing circuits as well as driver circuits can be realized on-chip.

1.06.3.3 Aluminum Nitride

Amorphous and nanocrystalline AlN (mostly prepared from powder) is a traditional electroceramic with a high spontaneous polarization of $\sigma = 6.2 \times 10^{13}\,\mathrm{cm}^{-2}$. It has been used widely as a transducer material in piezoresistive sensors and in surface acoustic wave (SAW) devices on Si (Caliendo *et al.* 2003), SiC (Takagaki *et al.* 2004b), $LiNbO_3$ (Wu *et al.* 2002), GaN (Chen *et al.* 2003), diamond (Shikata and Nakahata 2004), and GaAs (Liaw and Hickernell 1994) among other substrates, and a large number of studies have been published in this field. Recently, it has also been proposed for piezoelectric actuation (Olivares *et al.* 2005). For applications in a harsh environment, the AlN film is deposited on SiC and (nano/polycrystalline) diamond. The high internal polarization field and piezostriction is used, in conjunction with its highly insulating characteristics. The semiconducting properties are not used. On the contrary, bulk conduction is considered detrimental, as it will short circuit the internal field and shield the surface. Therefore, the residual conduction may be compensated by oxygen addition. It is

deposited by a variety of techniques like powder sintering, DC or RF sputtering, ion beam deposition, and as crystalline film by MOCVD.

Bulk single-crystal AlN has become available recently, but has been designed as insulating substrate material for GaN-based power FETs (Hu *et al.* 2003) and light-emitting diodes (LEDs) (Gaska *et al.* 2002). It is highly resistive and temperature stable. **Figure 12** shows a temperature measurement of a sample grown by induction heated vapor phase transport (VPT) from an AlN powder source at 2300°C (Zhuang *et al.* 2006). In the experiment, the temperature was ramped up in steps of 100°C and the resitivity measured with probes at each temperature in vacuum. The AlN single crystals remained stable up to 1100°C. No change of the RT conductivity was observed after several cycles. At RT, the resistance through the sample is very high reaching into the teraohm range. It acts therefore as perfect dielectric. From the geometry as shown in the inset, it may be estimated to $8 \times 10^{10}\,\Omega\,\text{cm}$ at RT. The conductance changes by three orders of magnitude with temperature when heated to high temperatures. However, even at a temperature of 750°C the estimated specific resistivity of the samples is in the order of $1\,\Omega\,\text{cm}$ and comparable to the RT resistivity of semi-insulating GaAs. The increase of conductivity may be linked to impurities within the material, but may in part be counterbalanced by a decrease in mobility. Impurities commonly detected by secondary ion mass spectrometry (SIMS) in bulk AlN crystals grown by the VPT technique are oxygen and carbon (Slack and McNelly 1976).

Thin single-crystalline AlN films are commonly grown on GaN by MOCVD or RF-MBE. Because of the high mismatch, the critical thickness for the growth of a strained layer is only a few nanometers and such layers are only used for capping of the surface or intermediate smoothing and stress release layers. Thicker layers are therefore relaxed and contain a high density of defects. Nevertheless, the first LED structures could be fabricated, also indicating that n-type doping (by Si) as well as p-type doping (by Mg) are possible (Taniyasu *et al.* 2002, 2006).

The application in SAW devices is one of the few, where membrane etching or cantilever fabrication are not required. Cantilevers and membranes with AlN ceramic-like thin films (not yet containing active layers) have thus mainly been realized using Si substrates and conventional Si-MEMS processing technologies. Growth as single crystal semiconductor thin film is mainly performed on GaN buffer layers, the buffer being mostly deposited onto Al_2O_3 or hexagonal SiC. Both substrate materials are etched mainly using a DRIE process.

1.06.3.4 CVD Diamond

Diamond is the metastable cubic form of carbon with exceptional stability. The phase has been reported to survive in H_2 at 2200°C (Kumar *et al.* 1996). However, in an inert atmosphere, graphitization is observed in the range of 1500°C (Chen *et al.* 1999). In atmosphere, it starts to be etched by oxygen above approximately 750°C (Adamschik *et al.* 2000a). For high-temperature operation, it needs therefore passivation or operation in vacuum, and all high-temperature measurements above 500°C are therefore generally performed in vacuum. It does not possess a natural substrate of

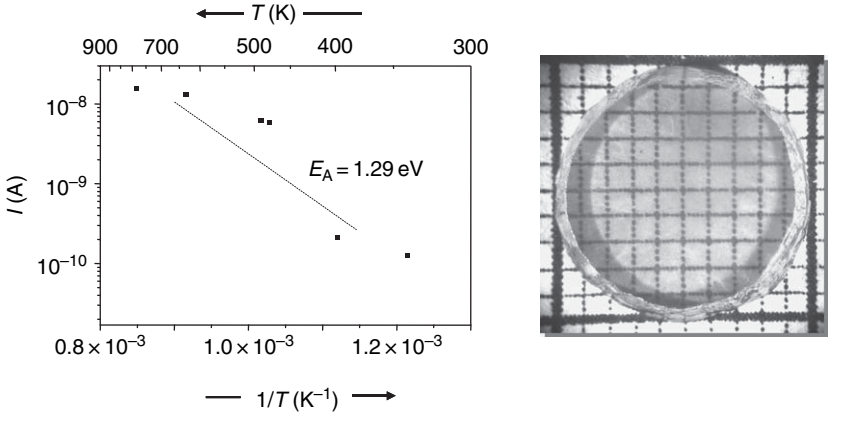


Figure 12 High-temperature conductivity measurement of single-crystal AlN up to 800°C. The data may be fitted with an average activation energy of 1.29 eV. The diameter of the sample was approximately 3 mm. (Source: Neuberger M, Aleksov A, Schlesser R, Kohn E 2007 *Electronics Lett.* **43**, 592–4 with permission.)

relevant size for a microsystems technology and is therefore generally deposited onto foreign substrates. With respect to microsystems, this is nearly exclusively (100)-oriented Si. Due to the high mismatch, the films are polycrystalline or heteroepitaxial. The exception is diamond on Ir (the Ir being deposited either on $SrTO_3$ or Si), which results in fully (100)-oriented epitaxial growth after a few nanometer thickness (Bauer *et al.* 2005). Unfortunately, this technique is still confined to small surface areas of approximately 1 cm^2.

The diamond phase can only be deposited by CVD. Two techniques are commonly employed, microwave plasma-enhanced CVD (MPCVD) and hot filament CVD (HFCVD), both techniques being used in the examples discussed below. An introduction to this subject may be found in Sevillano (1998) and Klages and Schäfer (1998). Both use a high concentration of H radicals by splitting H_2 in a high energetic microwave plasma (usually at 2.4 GHz) or by temperature while passing an array of hot wires (at approximately 2500°C). The precursor for diamond growth is usually CH_4, the growth conditions are chosen so that any graphitic phase is etched with a high rate. Thus, the H_2/CH_4 ratio is high (usually a H_2 concentration above 98% is used). The growth temperature is normally between approximately 600°C and 900°C. Deposition at high methane concentration or low plasma energy (like at 13.5 MHz) results in the deposition of amorphous carbon (a-C) or diamond-like carbon (DLC) with high graphitic content.

The deposition on Si can be separated into two phases: the seeding and nucleation phase and the outgrowth phase. Concerning seeding or nucleation, two directions are pursued: first, the seeding with nanocrystalline particles (Avigal and Hoffman 1999), sometimes embedded into an a-C film (Rotter 1999) and second the *in situ* formation of diamond nuclei at the Si surface (mostly also involving and interfacial SiC phase) by an electrical bias treatment (bias-enhanced nucleation (BEN)) (Janischowsky *et al.* 2003, Sheldon *et al.* 1994). While the deposited seed crystallites are randomly oriented, the nuclei formed by BEN can be oriented accordingly to the (100) orientation of the Si substrate. With both techniques a high density of nuclei or seeds ($>10^{11}$ cm^{-2}) can be generated, resulting in continuous films after a thickness of a few tens of nanometers.

Accordingly, two outgrowth techniques have been developed during the last decade: first, this has been the growth of highly oriented films (HOD) with a surface texture oriented with the Si substrate (Jiang *et al.* 1996, Kawarda 1998). The key was BEN of (100)-oriented crystallites with continuous outgrowth of large terraces, which start to coalesce into large single-crystalline areas. This seemed to be the near perfect case. However, during orientation-controlled outgrowth by α-parameter engineering (Wild *et al.* 1994), many smaller disoriented nuclei are overgrown and a vertical stress profile is developed, especially within the initial nucleation layer (von Kaenel *et al.* 1997). It has not been possible to eliminate this stress profile without the addition of chemical doping, which however influences the mechanical properties. Thus, HOD-coated Si wafers and freestanding cantilevers are generally bent upward (Kohn *et al.* 2001). **Figure 13** shows such a film. The second outgrowth mode is that of randomly oriented crystallites. Therefore the size of the individual crystallites is increased with film thickness, but can be limited, if growth conditions for a high nucleation rate are chosen. This is the basis for nanocrystalline films (NCD) and ultrananocrystalline films (UNCD).

For films in microsystems technologies, the surface needs to be smooth enough to allow high-resolution lithography and stress in the film needs to be controlled. Both points toward a nanocrystalline materials configuration. An overview is given in Sharda and Bhattacharyya (2004). With a grain size of 2–10 nm, this configuration is called ultra-nanocrystalline diamond (UNCD). It is realized with

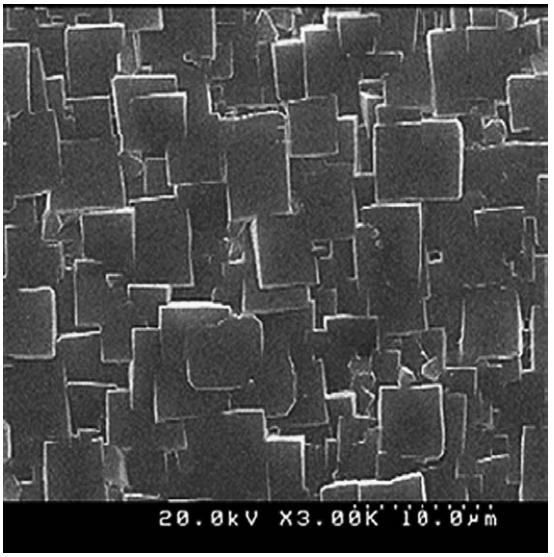


Figure 13 Scanning electron microscope (SEM) view of the surface of a highly oriented diamond surface, grown by microwave plasma-enhanced chemical vapor deposition (MPCVD) and bias-enhanced nucleation (BEN). The layer thickness is approximately 12 µm and the surface roughness is approximately 200 nm.

MPCVD with high Ar content of the plasma (Sekaric *et al.* 2002, Zhou *et al.* 1998). With a grain size between 10 and 100 nm, the material is labeled nanocrystalline diamond (NCD). It can contain individual nanosized grains or can contain thin vertical columns with a small visible grain size at the surface and nanometer smoothness (sometimes labeled 2D-NCD). Although these films contain already a continuous grain boundary network of graphitic bonds, these grain boundaries are still tight (generally less than 2 nm wide) and the films exhibit therefore many properties close to those of single-crystal material. Major differences may be observed with respect to the thermal conductivity (**Figure 14**) and electrochemical stability.

Undoped single-crystal and nanocrystalline films are generally highly resistive by deep level compensation with an RT resistivity up to the teraohm-centimeter range. Films can be doped p-type with boron during growth. Ion implantation is difficult to use, because the risk of graphite formation which is not removed during damage annealing. At concentrations below $10^{19}\,cm^{-3}$, carriers are not fully activated at RT due to a boron acceptor activation energy of 0.37 eV (Borst and Weis 1996). At high concentrations above $10^{20}\,cm^{-3}$, the activation energy vanishes and the acceptor fully activated at RT due to miniband formation with a resistivity in the milliohm-centimeter range (Werner *et al.* 1997b). These characteristics are well controlled for single-crystal material as shown in **Figure 15**. However, at high concentrations, the films become highly stressed as illustrated in **Figure 16**. The window for full activation and low stress is indeed quite small. On the other hand, highly boron-doped layers on undoped films will introduce a vertical stress profile in the stack, which can be used in stress engineering of the materials structure. Tunneling contacts to metals become possible for ohmic contact formation.

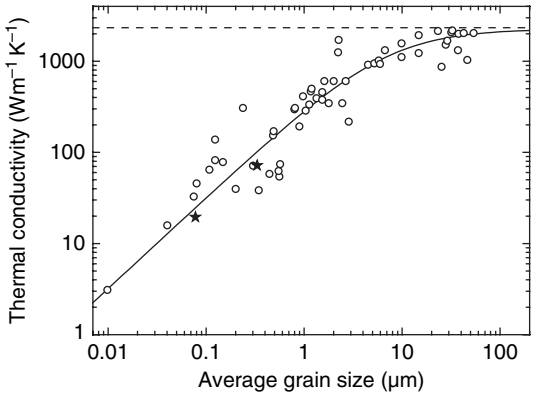


Figure 14 Thermal conductivity of nanocrystalline diamond (NCD) films of different grain size. Open circles are data from literature. Black stars are measurements of the lateral thermal conductivity of freestanding 3D-NCD cantilevers. Solid line: theory described in P, Schmidt, PhD thesis, University of Ulm, 2006. Data points: Verhoeven H, Flöter A, Reiß H, Zachai R 1997 Influence of the microstructure on the thermal properties of thin polycrystalline diamond films. *Appl. Phys. Lett.* **71**, 1329–31. Stars: P. Schmidt, PhD thesis, University of Ulm, 2006.

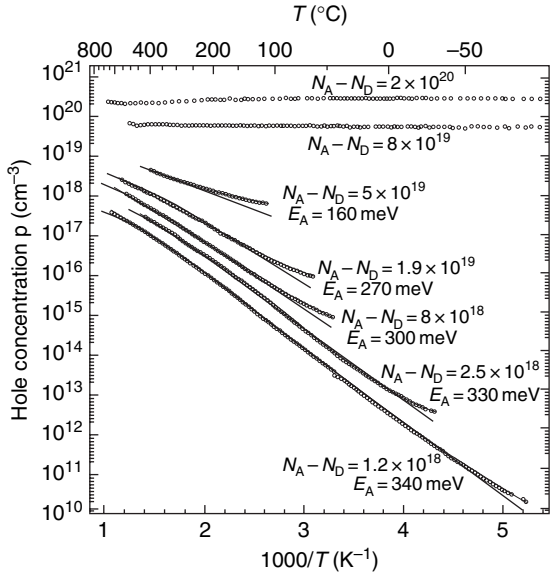


Figure 15 Concentration-dependent boron doping. (Source: Borst T H, Weist O 1996 Boron-doped homoepitaxial diamond layers: Fabrication, characterization, and electronic applications. *Phys. Stat. Sol. (a)* **154**, 423–44 with permission.)

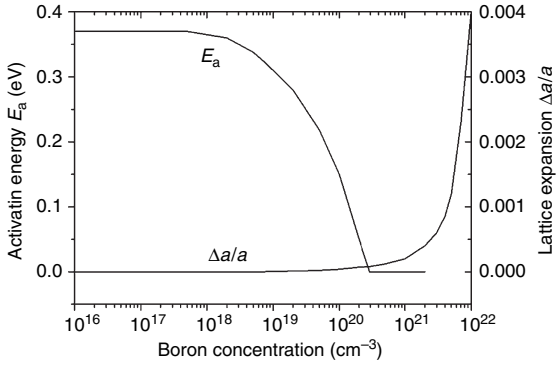


Figure 16 Mismatch induced by high concentration of boron.

Very similar data are obtained with polycrystalline and nanocrystalline material. **Figure 17** shows the incorporation of boron from the gas phase into the film with an indication of its activation. The Fermi level is not well defined for boron concentrations below $10^{18}\,\mathrm{cm}^{-3}$ and the layers may be fully compensated by donor-like defects mainly generated by the grain boundary network. Also, very high resistivity close to teraohm-centimeter can be obtained for undoped films as inferred from the characteristics of gigaohm seals in a patch clamp MEMS device (Kusterer *et al.* 2005). Complications are also expected at high doping levels ($N_A > 10^{21}\,\mathrm{cm}^{-3}$) with an inhomogeneous incorporation in the individual crystallites and the grain boundaries, and are indeed seen in electrochemistry (Becker and Jüttner 2003). At high concentrations above $10^{21}\,\mathrm{cm}^{-3}$, the film becomes highly stressed and increasingly brittle.

Donor doping is possible with nitrogen ($E_D = 1.7\,\mathrm{eV}$) (Collins 1989) and phosphorous ($E_D = 0.6\,\mathrm{eV}$), both deep impurities with negligible activation at RT (Kuizumi 2003). In nanocrystalline material, a high nitrogen concentration in the gas phase during growth leads to n-type conductivity with high RT activation (Bhattacharyya *et al.* 2001). This is mainly associated with grain boundary incorporation and grain boundary conductivity (Bhattacharyya 2004). Both p-type doping with boron and n-type doping with nitrogen have been used in NCD and UNCD films as active layers in MEMS structures.

Growth of UNCD layers has been performed at temperatures low enough to incorporate them into fully processed CMOS circuits (Xiao 2004). However, at temperatures below 400°C, the influence of graphitic bonds cannot be neglected anymore and the thermal conductivity, elasticity, and chemical properties need to be considered carefully. It may therefore be more feasible to consider the integration possibility with β-SiC FETs (Nagasawa *et al.* 2006) once the surface morphology of β-SiC is improved sufficiently. Growth of diamond on 100-oriented 3C–SiC has indeed lead to highly oriented diamond films with low spread in rotational angle (Kawarada 1998).

Stress engineering is an important aspect in the thin-film deposition technology for microsystems and relatively new to CVD diamond. In the past, most films have shown a vertical stress profile originating from the seeding and outgrowth techniques employed. Stress is especially developed during the first seeding phase, when the film coalesces and small crystallites are overgrown. Stress-free films are therefore obtained if growth starts on a high density of nucleation points, the grain size is limited to nanometer and the initial nucleation layer is only a few nanometers thick (Hernandez-Guillen *et al.* 2005). Due to the dissimilar thermal expansion of the diamond film and the Si substrate and the nonlinear behavior of this parameter, both tensile and compressive stress could be obtained by varying the deposition temperature. By specifically adjusting the growth parameter and adding chemical impurities a vertical stress profile can be artificially built in. Such a vertical stress profile has been deliberately used in bistable membrane and double anchored cantilever structures mentioned in Section 1.06.7.3 and discussed in Kusterer *et al.* (2006).

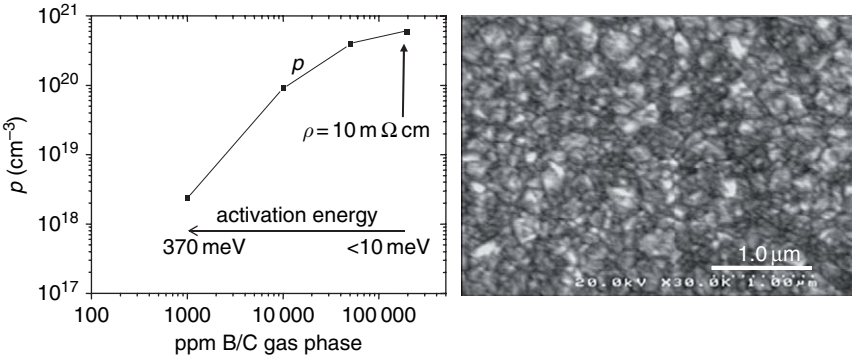


Figure 17 Boron doping of nanocrystalline diamond (NCD) films and indication of boron activation. Also shown is the morphology of the NCD film. (Source: Zimmermann T, Janischowsky K, Denisenko A, Kubovic M, Kohn E 2006 Nanocrystalline diamond pn structure grown by hot-filament CVD. *Diamond and Related Materials*, **15**, 203–5 with permission.)

1.06.4 Transducer Elements

As already listed in the first section, the materials are used in various fields of application like electromechanical, the electrothermal, and electrochemical fields. All rely on electrically active elements and thus doped layers.

1.06.4.1 Sensor Elements

A doped layer may serve in a number of sensor functions, the main ones being discussed in the following. To some extent, the parameters needed for high sensitivity may be conflicting, like a high gauge factor (GF) for a piezoresistor and a high-temperature coefficient of resistivity for a thermistor, when a piezosensor should be operated through a large temperature range.

1.06.4.1.1 Piezoresistors

Piezoresistors are mostly realized in conjunction with membranes, cantilevers, or double anchored beams. Therefore, undercut etching or substrate etching are an important part of the piezosensor technology. The etching properties are discussed in Section 1.06.5.3.

Piezoresistors make use of several effects. First, the geometrical effect found in all conductive materials leading to change in resistance due to a change in shape, when applying an external force. Second, the change in band structure of a semiconductor, when the lattice is deformed. The lattice deformation leads to change in carrier distribution in the longitudinal and transverse valleys and thereby to a change in average effective mass. Third, the piezopolarization effect needs to be added for polar materials like discussed in the case of GaN. In the diamond lattice (3C–SiC and diamond), large highly orientation-dependent piezoresistive coefficients are found both for n-type and p-type doping, translating into highly directional sensitivities at low concentrations.

One of the important figures of merit is the GF (or K-factor), which is defined as follows:

$$\mathrm{GF} = \frac{\Delta R/R}{\Delta L/L}$$

where $\Delta R/R$ is the change is resistance and $\Delta L/L$ is the strain ε.

1.06.4.1.1.(i) Silicon carbide Freestanding 3C–SiC structure are realized on sacrificial layers, by undercut etching of the Si substrate, by DRIE or wet etching of a Si substrate. The corresponding MEMS processes are discussed in Section 1.06.5.2.

In the case of 3C–SiC, the texture and accociated defects and grain boundaries limit the GF to approximately 32 (Okojie *et al.* 1998a, Ziermann *et al.* 1997). The doping active in the resistor is only in part shallow and fully activated at RT. In addition, active areas may be isolated from the substrate by pn junctions and leakage across the pn junction can generate bypass conduction at high temperature. Therefore, the GF is not constant for high-temperature applications. The most stable temperature performance is obtained with degenerate doping levels but also a reduced GF (Okojie *et al.* 1998b). Such piezoresistors have been used in pressure sensors up to 500°C (Okojie 2006).

Pressure sensors have also been realized on 6H–SiC substrates using piezoresistors. They have been developed and evaluated for high-temperature operation (Okojie *et al.* 1998a). To realize the membrane structures, a specific DRIE process for SiC substrates had been developed (Beheim and Evans 2006).

1.06.4.1.1.(ii) GaN, GaN heterostructures, and AlN GaN piezoresistive sensors are thought to exhibit high sensitivity due to the added polarization affect. However, their technology is still difficult, and sensor structures have been investigated in various configurations, mostly with their sapphire substrate still attached. N-type (Bykhovski *et al.* 1996) and P-type GaN (Gaska *et al.* 2000) have been used as well as Schottky diodes (Strittmatter *et al.* 2003a) and metal–insulator semiconductor (MIS) diodes (Strittmatter *et al.* 2003b). Careful analysis of the materials stack as well as the various piezocontributions lead to the conclusion that Schottky diodes on P-doped GaN films should exhibit the highest GF. Indeed a GF of 250 had been identified (Gaska *et al.* 2000). In an attempt to employ the high spontaneous polarization of AlN an n-GaN/i-AlN/n-GaN sandwich structure was investigated, where the 50-nm-thick AlN interlayer acts as piezoelectric insulator (Gaska *et al.* 1997). The response was measured laterally between ohmic contacts, thus, for the current flow all three layers have been in parallel and the response also included the formation of interfacial polarization induces 2DEG and 2D hole gas (2DHG) channels. The GF(=50) measured was

in between that measured for n-type Schottky diodes and p-type Schottky diodes.

With respect to membranes and cantilever structures realized on Si substrates, only proof-of-concept experiments have been reported up to now (Brueckner *et al.* 2005, Davies *et al.* 2004). An example for the undercut etching of AlN nanobeams is given in Kang *et al.* (2005). Since this reflects an early stage of technology, a cantilever experiment shall be described here shortly. The GaN active layer had been deposited by MOCVD onto a (111)-oriented of Si substrate. This process includes also the deposition of a thin low-temperature AlN nucleation layer. The rear side was masked by a metal mask and a window etched into the Si by DRIE, the GaN serving as selective etch stop. The cantilevers are often bent downward due to the residual built-in stress in the GaN buffer layer, although containing the stress relief layers as described in Section 1.06.3.2. **Figure 18** shows the top side of a cantilever with a mesa-etched piezoresistor at the anchor point. In this case, the cantilever was realized by dry undercut etching (Zimmermann *et al.* 2006). The results are discussed in Section 1.06.4.1.6.

1.06.4.1.1.(iii) Diamond In diamond, a resistor realized with boron doping ($E_A = 0.37$ eV) will show a strong temperature activation (see **Figure 15**). Diamond piezoresistors used in the temperature regime will therefore show a mixed response (Adamschik *et al.* 2001). At RT the GF is mostly below 20, due to the effect of the incomplete acceptor activation, the grain boundaries, and the built-in stress profile within the grains. At high temperature, the doping concentration becomes fully activated but the mobility is reduced. The GF starts to drop to close of that of metals. An application in an acceleration sensor is discussed in Section 1.06.6.1.

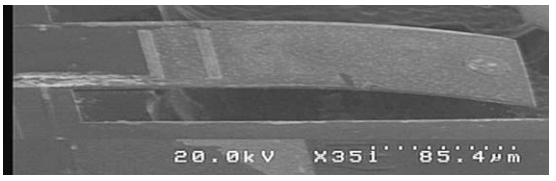


Figure 18 GaN cantilever with piezoresistor on Si substrate realized by undercut etching. (Source: Neuberger M, Zimmermann T, Berkhart P, Kunz M, Daumiller I, Dadgar A, Krost A, Kohn E 2004 GaN based piezo sensor. 62nd Device Research Conference (DRC), South Bend, IN, USA, Abstracts 45–6, June, with permission.)

Single-crystal diamond stones are used in high-pressure, high-temperature experiments performed in high-pressure cells (Boehler 2005). Diamond anvils withstand high pressures and high temperatures simultaneously without plastic deformation or cracking. However, it is difficult to control the environment at the working surface of the anvil. Typically, the temperature is determined by Raman spectroscopy with limited resolution. It has, therefore, been attempted to deposit a resistive pressure sensor structure on the working surface of the diamond stone by means of an epitaxial boron-doped resistor structure or boron-doped P-I-P diode (Zaitsev *et al.* 2001). The sensor had been tested up to 70 kbar and 850°C and was concluded, that it may work even at 1 Mbar and 1200°C. However, the piezoresistive effect is counterbalanced by the thermal activation of the deep boron acceptor and a temperature-stable high sensitivity can only be expected for temperatures, where the deep boron acceptor is fully activated and the activation of other deep centers does not interfere.

1.06.4.1.2 Thermistors

In thermistor structures, a large change of resistivity with temperature is preferable. In single-crystal material, this can be obtained through activation of a deep impurity, in a polycrystalline film carrier transport across grain boundary barriers may also lead to a high-temperature dependence.

In SiC, thermistors have been realized with RF-sputtered films with ceramic-like characteristics and applied up to 300°C (Nagai *et al.* 1982).

In diamond, the activation of the deep boron acceptor ($E_A = 0.37$ eV) has been employed. As can be seen from **Figure 15**, the temperature dependence of the carrier activation can be exponential across more than six orders of magnitude. This is also seen for the resistance, as the change in mobility is comparably small. The temperature dependence may therefore be expressed by a characteristic temperature β:

$$\beta = \frac{\ln(RT/R_N)}{(1/T) - (1/T_N)}$$

where R_N and T_N are the resistance and temperature values for the normalized conditions.

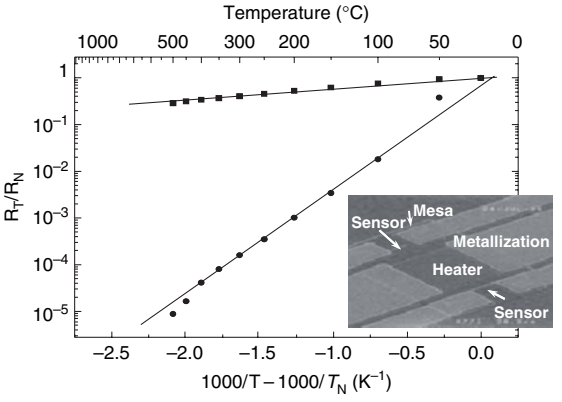


Figure 19 Diamond heater element and temperature sensor using boron-doped layers. High doping case: $\beta = 700$ K, $E_A = 61$ meV, $R_N = 230$ kΩ. Low doping case: $\beta = 4200$ K, $E_A = 360$ meV, $R_N = 6.6 \times 10^{10}$ Ω, $T_N = 300$ K. (Source: Gluche P, Leuner R, Vescan A, Ebert W, Kohn E, Rembe C, aus der Wiesche S, Hofer E P 1998 Actuator–sensor technology on electronic grade diamond films. *Microsyst. Technol.* **5**, 38–43 with permission.)

Figure 19 (inset) shows an application of such a diamond thermistor in an ink-jet spot heater for temperature control (Gluche *et al.* 1998). Two doping levels were used for the heater and the thermistor with two activation energies and two characteristic temperatures. The thermistor with low doping displays a characteristic temperature of $\beta = 4200$ K, which means that between RT and 500°C, its impedance changes by more than five orders of magnitude, well in agreement with Figure 15 (in case of a single-crystal substrate). Such thermistors have been found extremely temperature stable during operation at 500°C (Windheim *et al.* 1995).

1.06.4.1.3 Radiation detectors

All three materials: SiC, GaN, and diamond are investigated for application in DUV sensing, X-ray radiology, nuclear reactor monitoring, and high-energy physics experiments (Bruzzi *et al.* 2002). The GaN domain is however in essence the UV and DUV regime. Both SiC and diamond have been extensively investigated with respect to particle detection reaching from neutrons and α, β, γ particles to vacuum ultraviolet (VUV) and DUV. The penetration depth of these particles is in the range of 5 µm to full penetration in the case of neutrons. Detectors are therefore commonly vertical bulk devices with an active layer thickness up to several hundreds of micrometers and with contacts on both sides.

1.06.4.1.3.(i) Silicon carbide

To achieve detector grade properties, SiC active layers are generally realized epitaxially on single-crystal hexagonal (H4 or H6) polytypes. Substrates are still plagued by defects like micropipes and insulating characteristics need mostly vanadium doping to compensate for residual deep levels. Nevertheless, due to a continuously reduced defect density in epitaxial films, SiC detectors can be an alternative to Si devices in particle tracking in very high radiation environment (Sellin *et al.* 2006). Schottky diodes on 4H N^+-doped substrates and high resistive epitaxial material show very low leakage even at elevated temperature (<1.0 nA cm^{-2} at 100°C), allowing X-ray spectroscopy with high-energy resolution at elevated temperature (Bertuccio *et al.* 2004). The high materials quality is also indicated by the fact that no priming was needed to fill deep trap levels (Bruzzi *et al.* 2002). In another experiment, the resolution at 100°C was limited by the noise of the Si-JFET (junction field effect transistor) preamplifier also attached to the board (Bertuccio *et al.* 2004, Lee *et al.* 2006). Similarly, Schottky diodes have been tested as radiation detectors for 22 MeV electrons and 6 MeV photons (Bruzzi *et al.* 2003a). Beta particles (from a 0.1 mCi ^{90}Sr source) could be detected with 100% charge collection efficiency at 400 V (Bruzzi *et al.* 2003b). Neutrons have been detected with a LiF conversion layer (Dulloo *et al.* 1999).

Due to the large substrate size available the SiC, Schottky diode detector technology has already been extended to pixel detector arrays (Stevenson 2006). However, the high number of polytypes indicates already that under high fluences, carrier recombination can induce migration of defects, especially related to stacking faults.

1.06.4.1.3.(ii) Diamond

The advantage of diamond for particle detection over other materials is its radiation hardness, its high electrical breakdown strength, its high atomic density, its tissue equivalence to organic tissue, and its robustness in hostile environment. Using commercial simulation packages, the radiation damage of high doses of neutrons, electrons, protons, and α-particles have been studied and the results compared to those of Si (Mainwood 1998). It was concluded that, vacancies in Si migrate to form complex defects with dopants, whereas in diamond, the vacancies are immobile and the detector material remains usually undoped. Thus, despite the fact that diamond single-crystal material is not yet available in wafer size, a number of highly specific applications

have emerged, which may be referenced shortly in the following:

Diamond detectors in space missions:

Diamond CVD single-crystal metal–semiconductor–metal (MSM) detectors and PIN diodes (using boron acceptor and phosphorus donor doping) have been investigated for the large yield radiometer (LYRA) instrument, a solar VUV radiometer on board of the Solar Orbiter (Hochedez *et al.* 2002), and the ESA PROBA-2 mission (BenMoussa *et al.* 2006). LYRA will monitor the solar irradiance in four selected UV bandpasses from 220 to 1 nm.

C13-enriched diamond detectors have been developed to identify the isotopic composition of the solar wind (especially oxygen) in NASA's Genesis Discovery Mission (Baker *et al.* 2000).

X-ray and synchrotron beam metrology:

Quadrant structures have been fabricated on CVD diamond films as beam positioners to monitor drift in X-ray beams (Bergonzo *et al.* 1999, Marczewska *et al.* 2003, Shu *et al.* 2001). They allow permanent *in situ* monitoring of beam instabilities during experiments.

Diamond in radiology:

Single-crystal diamond has been used as solid state ionization chamber with high linearity, no memory effects, and no priming has been needed (Balducci *et al.* 2006a).

CVD diamond has been investigated for heavy ion beam tracking with energies ranging from 80 to 400 MeV u^{-1}, counting the number of impinging particles (Rebisz *et al.* 2006). The 100 MeV u^{-1} particles from a ^{76}Ge source have been counted with rates of up to 10^6 particles s^{-1} (Stolz *et al.* 2006). In this experiment, the single-crystal film had been grown epitaxially on Ir.

A 2″ diamond-on-Si wafer has been used to measure special beam distributions (Marczewska *et al.* 2003).

Diamond in nuclear physics:

CVD diamond has been investigated by several laboratories to replace existing detector materials and substitute for selected natural stones for high-dose particle detection, especially high-energy neutrons. Tests have been performed at high-energy physics and nuclear reactor facilities like at CEA, TRINITI (Bergonzo *et al.* 2001), in the Superconducting Super Collider (SSC), and the Large Hadron Collider (LHC) facilities (Allers *et al.* 1997). Using several hundreds of micrometers thick polycrystalline free-standing large area CVD substrates microstrip sensors and pixel detector arrays have been developed for the Large Hadron Collider Project (Adam *et al.* 1999, 2000).

CVD diamond has been investigated for neutron detection in several configurations:

Neutrons passing through ^{235}U will produce a number of fission products (charged particles with several tens of mega electron volt energy), which have been monitored by high-grade polycrystalline CVD diamond films (Lardon *et al.* 2006).

A ^{239}Pu radiation source diluted in nitric acid has been monitored by a diamond detector immersed into the solution (Bergonzo *et al.* 2000). Here, the high corrosion resistance of diamond has been important.

Boron deposited onto diamond has been used as conversion material, with transmutation to ^{7}Li and emission of α-particles (Mer *et al.* 2004). Alpha particles are then detected in diamond, on the other hand, the interaction with background γ-rays is less likely. This has resulted in a high neutron to γ selectivity.

The sensitivity to α-particle detection, has for example, been tested with irradiation from an Am and Cm source as well as triple ^{239}Pu, ^{241}Am, ^{244}Cm source, emitting 5.16, 5.48, and 5.8 MeV α-particles (Balducci *et al.* 2006b).

As may have become apparent, particle detectors are with few exceptions discrete devices wire connected to a low noise amplifier tailored to the bandpass characteristics of the radiation. This amplifier is generally located at some distance outside the harsh environment and the sensitivity and signal-to-noise ratio of the detector may be crucial.

1.06.4.1.4 Electrochemical electrodes

As already discussed, electrodes in electrochemistry are the electronic interfaces to liquids. Quasi-metallically doped inert electrodes are discussed here, which can be operated with high overpotentials without corrosion or fouling. SiC will form CO, CO_2, and SiO_2 at its surface under anodic polarization (Lauermann *et al.* 1997) and the electrochemical stability will in essence be similar to that of Si. GaN and AlN will form Ga and Al oxides and hydroxides on the surface (Huygens *et al.* 2000). The most inert semiconductor material is diamond.

1.06.4.1.4.(i) Gallium nitride In a similar experiment, as described below, for the case of diamond, a quasi-metallically n-type-doped GaN electrode has been bias cycled six times (**Figure 20**). The first scan indicated a potential window of nearly 4 V. However, the following scans indicate that the activity of the electrode is increasingly decreased. The electrode becomes unstable at high anodic potentials. There are several possibilities for this behavior. Most likely the surface becomes oxidized and shielded by an adsorbate layer. The detailed mechanisms are however not clear at present. Similar behavior is seen for n-type AlN (Kohn *et al.* 2005).

1.06.4.1.4.(ii) Diamond Diamond can be highly doped with boron and is therefore an ideal electrode material for applications in harsh and hazardous liquid environments. The inertness and durability of diamond is highlighted with the experiment shown in **Figure 21**. Here a single-crystal diamond electrode has been cycled many times in a wide potential window-detectible corrosion. The graph shows the surface after a 28 500 current voltage scans at 1 V s^{-1} scan speed from −3 V to +7 V DC. A total charge density of 17.85 A h cm^{-2} had been transferred. During this time the electrochemical window had been reduced from −2/+3.2 to −1/+2.7 V DC. The surface roughness has changed from 2.6 to 2.8 nm, and has thus been essentially maintained.

This high stability in the anodic regime is the basis for the application of diamond electrodes in electrooxidation and electroanalysis (Haenni *et al.* 2004, Swain 2004). This is illustrated for the three cases shown in **Figure 22**.

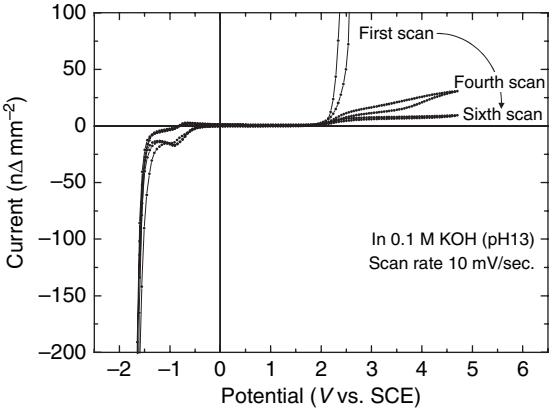


Figure 20 Voltammogram of metallically doped (n-type) GaN electrode during various number of scans as indicated, solution used 1 M KOH (pH 13); scan speed: 10 mV s^{-1} (Courtesy: Denisenko A, University of Ulm with permission.) SCE: standard calomel electrode

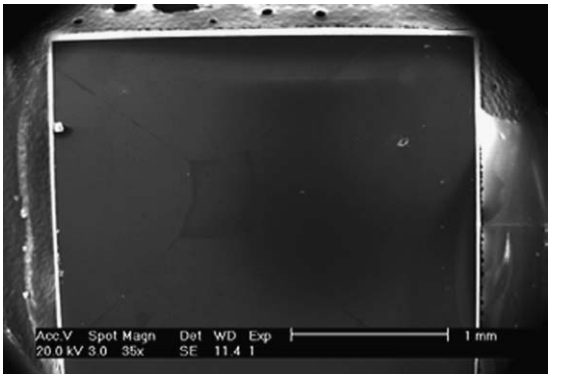


Figure 21 Single-crystal diamond electrode after 28 500 scans as described in the text; SHE, standard hydrogen electrode; SCE, saturated calomel electrode. (Source: Kohn E, Denisenko A, Ive T, Brand O, Dadgar A, Krost A, Haenni W 2005 Liquid junctions to wide bandgap semiconductors. *13th Int. Workshop Physics of Semiconductor Devices (IWPSD-2005)*, New Delhi, India, The Physics of Semiconductor Devices 2005, December 2005, pp. 750–9.)

Here, the oxidation of histamine (a biomolecule), phenol, and sulfate are shown among a large number of substances (Haenni *et al.* 2004, Rao 2000). It may be added that the oxidation of organic molecules goes through several steps and the ones shown here do not represent full oxidation. These oxidation potentials are compared with the potential window for water dissociation, the energetic location of which is dependent on the pH of the solution. A typical potential window of a diamond electrode is also shown. The correlation of the electrochemical potential scales with the physics-based energy levels is added (compare also with **Figure 5**). It can be seen that all the three reactions would not be possible with a Pt electrode.

Further improvement in sensitivity and selectivity can be obtained with microelectrode arrays. Here, the micron size of the active areas leads to a field enhancement and local modification of the Helmholz layer, again leading to an enhanced reaction rate, but also an increased influence of the grain boundary network (Colley *et al.* 2006, Soh *et al.* 2004). Such an electrode array is shown in **Figure 23**, which has been developed by CSEM (Haenni *et al.* 2004).

1.06.4.1.5 Diodes for high-temperature gas sensing

The control of gases in a high-temperature environment is a fundamental requirement for high-temperature reactions like in combustion and jet

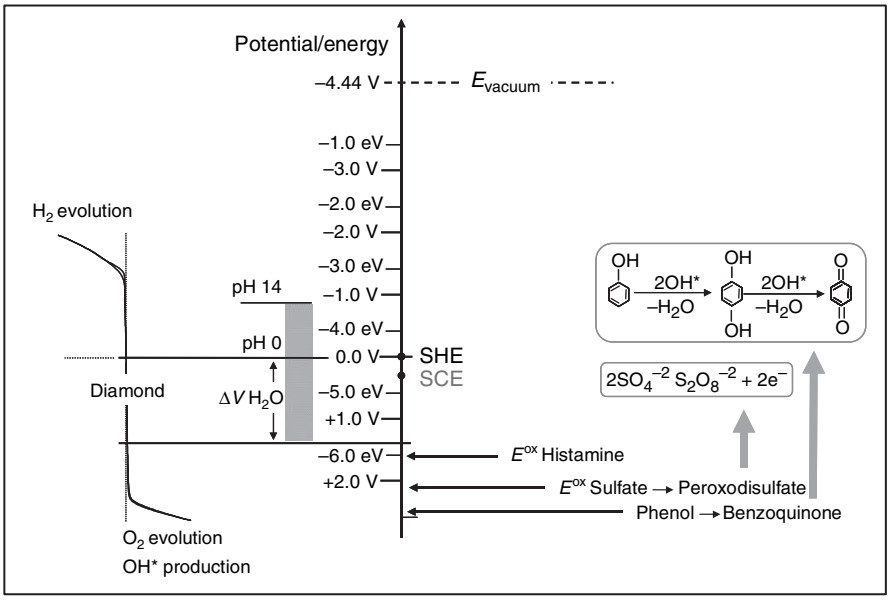


Figure 22 Redox potentials for various reactions in the semiconductor energy scale and potential window of water dissociation depending on pH. Also shown is the voltammamogram of a typical quasi-metallic-doped diamond electrode in 1 M H_2SO_4.

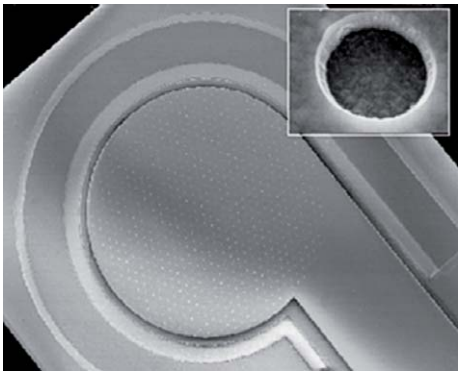


Figure 23 Diamond microelectrode array; pixel diameter 23 μm. (Source: Haenni W, Rychen P, Fryda M, Comninellis C 2004 Industrial applications of diamond electrodes. *Thin Film Diamond II (Semiconductor and Semimetals)*. Elsevier, Chap. 5.)

engines. The identification of polluting exhaust fumes is highly desirable. Gases may be NO_x, NH_3, HCl, CO, CO_2, and volatile hydrocarbons. Typical operation conditions are high temperature, chemically reactive environments, and varying oxygen and hydrogen concentrations.

Semiconductor detectors are typically based on high-temperature stable nanocrystalline metal oxides with wide bandgap like IrO_2, TiO_2, or SnO_2, which can be operated at temperatures in excess of 1000°C. Gases are detected by adsorption at the surface of the nanocrystals, causing changes in the residual electronic conduction across the grain boundaries and at very high-temperature ionic transport through the dielectric. However, single-crystal semiconductors like SiC, GaN, and diamond can also be used in combination with catalytic metals, which dissociate gas molecules into radicals at their surface like Pt, Pd, or Ir. In many cases the diffusion of protons through a thin metal film toward the semiconductor surface at high temperature is also an important contribution. The adsorbed radicals form a dipole layer on the semiconductor surface and modify its surface potential. In the case of a Schottky diode, the change in surface potential is detected by changes in the IV characteristics. In the case of a MOS diode, the change in surface potential is detected as shift in threshold voltage. The MIS diode has two drawbacks: first, the thermal activation of oxide leakage currents should not degrade the quality factor of the MOS diode in the temperature range of application and second, for the capacitive readout an RF circuit is needed. To be able to operate at elevated temperature, the semiconductor bulk and surface have to be highly stable and corrosion resistant. Both SiC and AlGaN/GaN diodes have been used successfully in this configuration. The exposure to oxygen is critical. Both materials form an oxide at the surface at high

temperature and diffusion of oxygen through this surface layer will eventually cause drift. Unlike diamond, they are stable against etching by oxygen exposure. Diamond is therefore limited to a temperature range below 500°C.

1.06.4.1.5.(i) Silicon carbide Dielectric-assisted SiC Schottky diodes have been used at temperature as high as 1000°C with Pt as catalytic Schottky metal (Spetz *et al.* 1997). Reducing gases like H_2, C_xH_y, and NH_3 are dissociated into radicals and the adsorption of protons used for detection. The sensitivity to non-hydrogen containing oxidizing molecules like NO_x or CO_2 has been linked to the porosity of the Pt Schottky metal layer and reactions of dissociated molecules with residual H radicals. High sensitivities have been obtained, when the catalytic metal is in contact with a thin interfacial layer of SiO_2 (Spetz *et al.* 2001), WO_3 (Kandaswamy *et al.* 2005), Ga_2O_3 (Trinchi and Wlodarski 2004), or mixed oxide $TaSi_xO_y$ (Tobias *et al.* 2000), indicating that the adsorption/desorption site at the interface is surface oxygen related. SiC Schottky diodes contain therefore a thin (approximately 1 nm) SiO_2 interfacial layer and may be considered a dielectric-assisted Schottky contact (MIS SiC-Schottky diode (Uneus *et al.* 1999). **Figure 24** shows a basic design and the shift in barrier potential (with Pt as a catalytic metal) by exposure to O_2 and H_2. The response time is in the millisecond range. The devices have been realized on bulk SiC crystals mainly in the 6 H polytype.

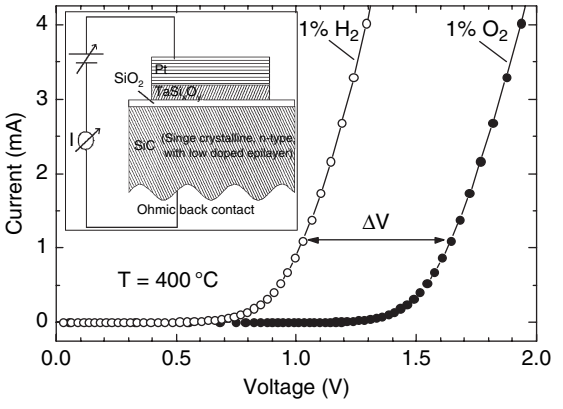


Figure 24 SiC gas sensor diode with catalytic layers as indicated in the figure with shift of threshold voltage as function of hydrogen and oxygen content. (Source: Tobias P, Rask P, Göras I, Lundström A, Solomonsson P, Spetz A L 2000 Platinum-insulator-silicon carbide structures as small and fast sensors for exhaust gases. *Proc. Sensoren und Messysteme 2000*, VDI Berichte 1530, pp. 179–87 with permission.)

Again, in the MOS diode sensor, the gases are dissociated by a catalytic metal, protons diffuse through the metal and form a dipole layer at the SiO_2 interface. This changes the field distribution across the dielectric and leads to a flatband voltage shift. In the case of gases, which do not contain hydrogen like CO and NO_x, the formation of the dipole layer is difficult, as diffusion of other radicals than protons through the catalytic metal is suppressed. Either detection is indirect by reaction with residual hydrogen, or the metal has to be porous. At triple points, where the gas phase, the insulator, and the catalytic metal meet within the diffusion length of the split molecules, radicals can be adsorbed on the insulator surface. Such points are used in the dissociation of NH_3 followed by proton adsorption. To retard diffusion of oxygen through the dielectric, with related continued SiC oxidation, the oxide layer had been modified to a $SiO_2/Si_3N_4/SiO_2$ layer stack (Spetz *et al.* 2001). The capacitance is read out under weak depletion conditions (near flatband) to avoid complications with the deep depletion/inversion behavior, depending on thermal minority carrier generation at high temperature. The device has been operated as MISiCFET (SiC metal oxide semiconductor field effect transistors (MOSFETs)) (**Figure 27**) at 800°C (Spetz *et al.* 2001).

1.06.4.1.5.(ii) Gallium nitride The same detection principles as discussed above have also been applied to GaN Schottky diodes covered with Pt and Pd. Here also a shift in barrier potential has been observed depending on the exposure to H_2, HC1, CO, and NO_x. A maximum response to H_2 has been observed between 450°C and 500°C with a shift in barrier potential from 1.15 eV by approximately 0.5 eV (Schalwig *et al.* 2001). Overall, the results indicate that gas detection in these diodes follows essentially the same fundamental principles as in the case of the corresponding SiC devices. The response is mostly nonlinear and saturates at high concentrations. This can be explained by the balance between reactivity and diffusion-limited processes and a kinetic phase transition of the adsorbate layer similar to the Pt/SiC case (Kim *et al.* 2003).

In a recent experiment, an epitaxial GaN resistor has been used for H sensing (Yun *et al.* 2005). Here the change in surface potential of the free surface and thus the change in channel resistance due to changes in surface depletion have been employed. However, the active surface states have not yet been identified.

It should be noted that the stable surface condition of Ga face GaN will be an oxygen-terminated surface. However, Ga_2O_3 will become instable in reducing the atmosphere above 700°C (Fleischner and Meixner 1992). In the atmosphere, GaN will decompose at approximately 650°C (Pisch *et al.* 1998).

A higher stability is therefore expected for an AlGaN Schottky contact, still grown on GaN. Such Pt–AlGaN/GaN diodes have indeed been operated up to 800°C in H_2/N_2 atmospheres (Song *et al.* 2005). A cross section is shown in **Figure 25**. The barrier potential increased monotonically (reaching 120 meV at 800°C). The temperature limit is thought to be reached by alloy formation between the Pt Schottky metal and AlGaN semiconductor, rendering the junction ohmic.

1.06.4.1.5.(iii) Diamond On polycrystalline diamond structures for CO gas sensing have been realized, where SnO_2 is the catalytic layer deposited onto a (nominally) undoped diamond layer on a highly P-doped substrate (Gurbuz *et al.* 1998). Thus, the IV characteristic is that of a glossy heterojunction diode (**Figure 26**). The oxygen detection mechanism has been found to be due to the modification of the oxygen vacancy density in the SnO_2 film generated by the dissociation of the oxygen-containing gas. These oxygen vacancies act as donors, donating electrons to the conduction band. This leads to a potential redistribution across the SiO_2/i-diamond junction and modification of its IV characteristics. An activation energy of vacancy generation of 1.1 kcal mol^{-1} has been found and a stable response increasing with temperature. The applied temperature range was up to 300°C, which means well below the oxygen temperature stability limit of diamond.

Using Pd instead of SnO_2, the structure has recently also been tested for the detection of benzene (C_2H_4) and toluene (C_6H_8) proposed for use in the soil as an indicator for oil contamination (Gurbuz *et al.* 2004).

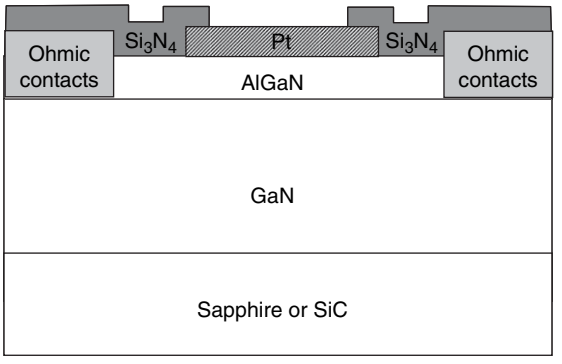


Figure 25 AlGaN/GaN heterostructure gas sensor with Pt catalytic Schottky layer.

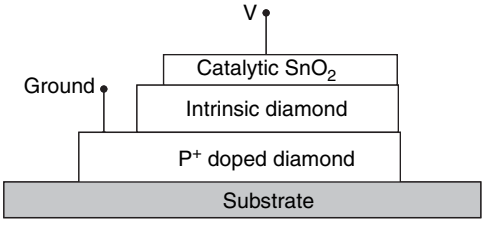


Figure 26 Schematic cross section of catalyst-adsorptive oxide-insulator-semiconductor (CAIS) structure on diamond.

1.06.4.1.6 Field effect transistors

The technology and harsh operating conditions of the ceramic-like wide bandgap semiconductors is rarely compatible with Si-CMOS technologies despite the fact that Si is the substrate in most cases. In sensing, transistor structures may directly amplify or convert signals without interconnect and matching losses, thus they are realized in the sensor material itself. Most commonly considered are FETs, as here unipolar charge transport is modulated and injection phenomena, which are in competition with deep level activation, are not dominating. Such FET structures have emerged in three areas:

gas sensing at high temperature
chemical sensing in liquids
in electromechanical sensing as extension of the piezo-resistor

1.06.4.1.6.(i) SiC MOSFETs SiC MOSFETs (MISiCFETs) have been developed primarily for high-temperature gas sensing. They are based on the MOS sensors as discussed in the previous section and the catalytic diode structures are implemented into the MOS gate. Usually Si-based MOSFETs are operated with inversion channels. In SiC-based devices, the channel is in general created by doping and has thus been buried to avoid complications with surface states and minority carrier generation at high temperature (up to 1000°C) (Spetz *et al.* 2001). Thus, the device can operate in the depletion/accumulation mode, depending on the channel doping level. The channel may be prebiased by a backgate to adjust for a high sensitivity bias point near pinch-off or may be used for drift compensation (Gurbuz *et al.* 1998) (**Figure 27**).

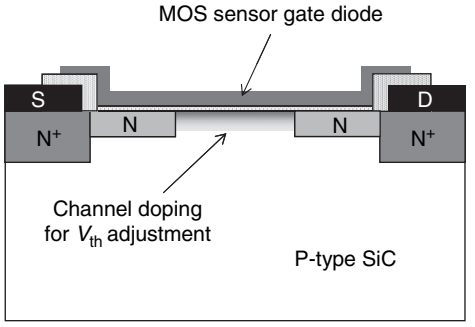


Figure 27 Schematic view of SiC metal oxide semiconductor field effect transistor (MOSFET) for high-temperature gas sensing.

1.06.4.1.6.(ii) AlGaN/GaN HEMTs Because the hexagonal GaN-based materials system is highly polar, the advantage of GaN-based HEMTs is seen in the highly efficient modulation of the channel polarization charges and the FETs may be labeled piezo-HEMTs. The piezoresistive effect has several contributions as discussed in Section 1.06.4.1.1. The example mostly employed is the AlGaN/GaN heterostructure, where an interfacial 2DEG channel of high density (with a sheet charge density in the order of 10^{13} cm^{-2}) is formed entirely by the polarization discontinuity between the AlGaN barrier layer and GaN substrate. This interfacial 2DEG is used in the FET (which is a HEMT structure) as channel. Here, the AlGaN barrier layer is strained on GaN and the polarization discontinuity is determined by the piezopolarization induced by the stress and the difference in spontaneous polarization between the two materials. The various contributions to the polarization charges are given in **Figure 11**. A general cross section of the AlGaN/GaN piezo-HEMT is shown in **Figure 28**, and a band structure diagram in **Figure 29**. The 2DEG channel is generated at the AlGaN/GaN interface by the polarization fields with

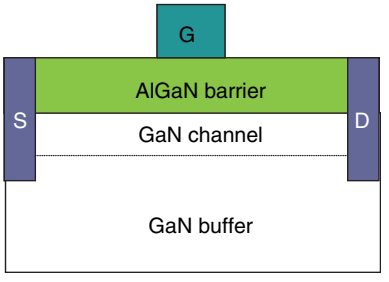


Figure 28 Cross section of AlGaN/GaN piezo-HEMT (high electron mobility transistor).

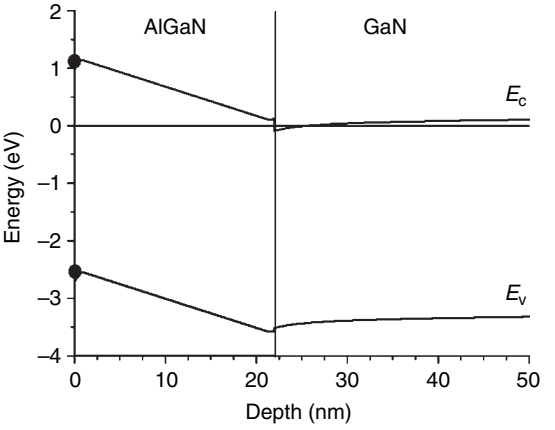


Figure 29 Associated bandstructure of the heterojunction of **Figure 28**; AlGaN with 30% Al content.

a countercharge at the surface (mostly in surface states often called surface donors) (Ibbetson *et al.* 2000).

Therefore, this structure is sensitive to stress as well as to changes in the charge state of the surface due to the absorption of polar molecules. Thus, they have been used in all three applications: as high-temperature gas sensors employing the heterostructure diode as discussed in Section 1.06.4.1.4; as ChemFET in biochemical applications based on the electrode characteristics as discussed in Section 1.06.4.1.3; and as a membrane pressure sensor and as a force sensor on a cantilever as a bias-controlled piezoresistor.

As with HEMT gas sensors, they are based on the gate electrode with catalytic metal coverage. The signal is then capacitively coupled to the 2DEG channel of the FET. The operation with Pt gate has been demonstrated up to 300°C for H_2, C_2H_4, and CO detection with high sensitivity, but also a low level of saturation (Schalwig *et al.* 2001).

For biochemical applications, the ChemFET structures have usually been realized on sapphire as the thermal properties of SiC are not important. In this case, the surface of the AlGaN barrier layer is the electrochemical electrode, capacitively coupled to the channel. The surface in contact with the solution is in general an Al, Ga oxide or Al, Ga oxynitride, naturally formed by the contact with the water-based solution with an associated redox potential most likely generated by hydroxyl active sites (Steinhoff *et al.* 2003). The stability of this oxide surface will determine the stability of the ChemFET, especially under high anodic overpotentials. The stability of a strongly

oxygen terminated GaN electrode surface is shown in **Figure 20** and reflects in essence the characteristics of an AlGaN surface with 30% Al content. Thus, the area of safe operation is limited by the onset of anodic current transfer at approximately +2.5 V (against SHE). This potential appears in the FET operation between gate and drain and limits the region of safe operation in the output characteristics. This is illustrated in **Figure 30** for operation of the ChemFET in dilute HC1 with pH 3.8. However, it should be noted that the gate bias range covered by a solution of pH 1 and 13 is only approximately 1 V. Therefore, for high sensitivity, enhancement mode devices with short gatelength operating in a small gate and drain bias window are needed, which are however difficult to realize. Nevertheless, such AIGaN/GaN HEMT structures have been used in biochemical investigations (Steinhoff *et al.* 2005). Studies on reliability and degradation are not available yet.

With respect to AlGaN/GaN piezo-HEMTs they have initially been realized on sapphire and tested with the substrate still attached (Kang *et al.* 2003). Therefore, the response had been moderate. Fabricated on Si substrates as discussed in Section 1.06.4.1.1, FETs could be realized on membranes and cantilevers (Kang *et al.* 2004). On GaN, cantilevers have been investigated as bias-controlled piezoresistors (Zimmermann *et al.* 2006). Approximately half of the channel current could be modulated, when bending the 120-μm-long cantilever 30 μm upwards. The beam had been approximately 2.3 μm thick and contained one stress relief layer and a 20-nm-thick strained AlGaN HEMT barrier layer. The YM of cantilever extracted from mechanical bending measurements was 190 GPa. This may be compared with a value of 250, obtained by nanoindentation (Jian *et al.* 2003). The highest relative change in output current of 50% has been obtained for operation with a gate voltage ($V_{gs} = -3.0$ V) near pinch-off (**Figure 31**). However, the experiment did not allow to separate the individual contributions of the overall piezoresponse and the thermal stability of the device. The operation of AlGaN/GaN polarization-based HEMTs up to 750°C (Daumiller *et al.* 1999) may however indicate that operation at high temperature is possible without loosing the piezopolarization.

1.06.4.1.6.(iii) Diamond FETS Diamond FET technologies are still rather immature and just emerging. Two fields of application drive this development. First, these are high-power, high-temperature electronics and second, electrochemistry and biochemistry. FETs have been realized based on two channel concepts. As a first concept, electrolyte–solution–gate FETs have been realized using the hydrogen-terminated surface, which induces a near-surface 2DHG (Kawarada *et al.* 2001). This structure does not contain extrinsic channel doping. The surface chemistry

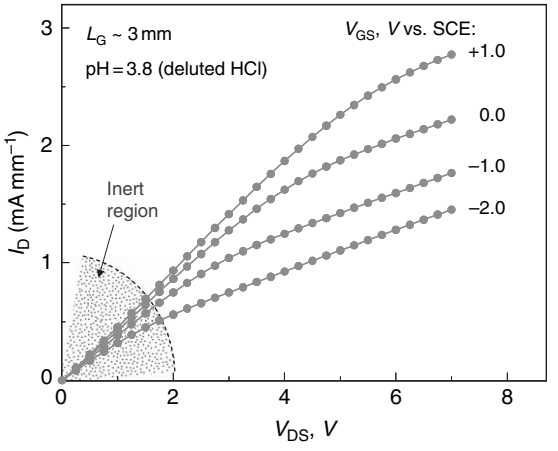


Figure 30 GaN-based high electron mobility transistor (HEMT) with AlGaN electrode and area of safe operation as described in the text. (Source: Denisenko A, Pietzka C, Kohn E, Kunze M, Daumiller I, Dadgar A, Krost A 2006 Effect of anodic polarization on electrochemical properties of GaN surface. *E-MRS 2006 Conference*, Symposium Q, Presentation Q17, May/June 2006; private communication, with permission). SCE: saturated calomel electrode

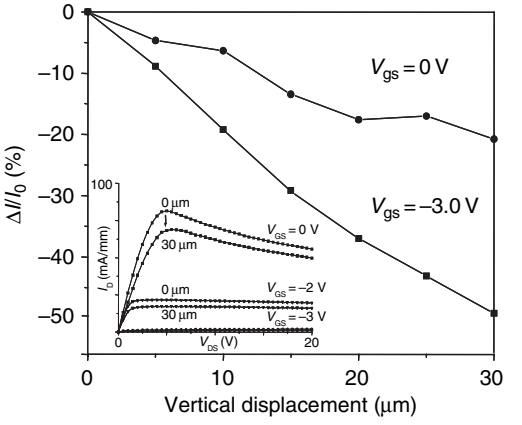


Figure 31 Response of AlGaN/GaN piezo-HEMT (high electron mobility transistor) for two different gate potentials. The structure is similar to the one shown in **Figure 18**. (Source: Zimmermann T, Neuburger M, Benkart P, Hernandez-Guillen F J, Pietzka C, Kunze M, Daumiller I, Dadgar A, Krost A, Kohn E 2006 Piezoelectric GaN sensor structures. *IEEE Electron Devices Lett.* **27**, 309–12.)

leading to this configuration is still in debate, but commonly transfer doping from surface adsorbates to form the 2DHG is made responsible (Strobel *et al.* 2004). Thus, this surface should be extremely sensitive to liquid environments. However, the hydrogen-terminated surface is chemically not active, active sites have to be added, for example, by oxidized spots or electrochemically active defects. Second, the hydrogen termination is not stable at anodic potentials (Kondo *et al.* 2005). Therefore, the response of such FETs in the liquid environment is in general only conditional stable or may even be subject to degradation (Müller *et al.* 2003).

In a second approach, the channel is created by a thin boron doping spike with a peak concentration high enough for full activation. Because the sheet charge that can fully be modulated by the gate without breakdown, is limited, the thickness of the spike needs to be in the nanometer range, which is difficult to realize. This configuration has been called delta doping (Kunze *et al.* 1999). However, in this case, the surface can be oxygen terminated in the same way as used for electrochemical electrodes. It is, therefore, an extremely reliable system, which can operate in the entire range of pH and under high overpotentials. A first result is shown with **Figure 32**, also indicating the measurement arrangement, where the shift in transfer characteristic is shown between pH 1 (1 M H_2SO_4) and pH 13 (1 M KOH). The experiment has been repeated several times and slight changes are observed, which had been associated with degradation of the passivation surrounding the active gate electrode area (Denisenko *et al.* 2006).

1.06.4.2 Actuator Elements

Many microsystems contain moving parts, which have to be actuated. Concerning a basic cantilever, the actuation dynamics is basically reflected in the mechanical resonance frequency of the beam as seen from eqn [1], where E is the YM, ρ is the density of the material, t is the thickness of the cantilever beam, and l is the cantilever length.

$$f_{\text{res}} \cong 0.16\sqrt{\frac{E}{\rho}}\frac{t}{l^2} \qquad [1]$$

Two groups of parameters influence the dynamics: the materials properties with the ratio of E/ρ and the geometry with the ratio t/l^2. For fast response thick short cantilevers are desired, pointing toward an activation mechanism, which can generate high forces. Because the materials discussed here possess a high YM, again a high actuation force is required. Common actuation principles are the electrostatic, piezoelectric, and the electrothermal actuation.

On the other hand, devices may be operated in aggressive liquids and a different type of actuator is needed. The best-known device in this respect is

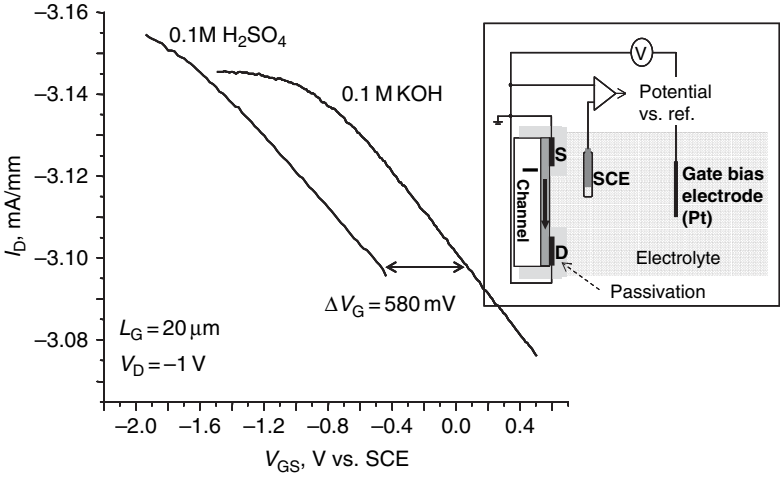


Figure 32 Diamond ion-sensitive field effect transistor (ISFET) transfer characteristics for two different solutions with pH 1 and 13 and measurement arrangement. The structure was realized with boron-doped delta channel and oxygen-terminated surface. (Source: Denisenko A, Jamornmarn G, El-Hajj H, Kohn E 2006. pH Sensor on O-terminated diamond using boron doped channel. *Diamond 2006*, Estoril, Portugal, Abstract 16.5, September 3–8, 2006; private communication, with permission.) SCE: saturated calomel electrode.

probably the ink-jet and the actuator either a piezo-element or a microheater. Microheaters are discussed first.

1.06.4.2.1 Microheaters

Microheaters have been commonly realized with high melting materials like graphite or BN and HfB_2 films (Solzbacher *et al.* 2000) and used in a temperature range of up to and above 1000°C. At this high temperature, oxidation and corrosion are critical and only doped oxides like SnO_2:Sb can withstand these conditions reliably (Spannhake *et al.* 2006). Substrates are often alumina or SiC ceramics. N-type-doped single-crystal epitaxial SiC itself has been used up to 1000°C in air, although it is not totally oxidation resistant, and above 1200°C in an inert and reducing atmosphere (Mlcak *et al.* 2004). Thus, SiC allows use of doped layers as heaters and to realize the entire stack, which is thermally stressed, in monolithic form. SiC is therefore the mostly used material for heater platforms.

Diamond can also substitute all elements needed in a microheater structure (Yang and Aslam 1996). Although attacked by oxygen above approximately 750°C, heaters are still extremely durable during heat cycling even without external passivation. This is shown in **Figure 33**. Here a heater element of 1 mm^2 surface area is covered with a thin water film and operated in the open mode of an ink-jet without cavity (Kaiser *et al.* 2002). The graph illustrates water ejection from the heater surface with a pulse repetition rate of 2 kHz and a pulse power density of approximately 1 MW cm^{-2}. When the power is applied, the surface is heated up to the spinodalian temperature (of 324°C for water), igniting spontaneously the phase transition. A thin vapor layer develops across the heater, ejecting a bubble from the water film, thermally isolating the heater surface immediately. Thus, the heater temperature will be essentially higher than the spinodalian temperature. When the vapor layer collapses, high mechanical cavitation forces act on the heater surface. Obviously, it is advantageous, to employ a monolithic system with intrinsically balanced thermal stresses. Additionally, in the case of diamond, the grain size of the film can be used to tailor the thermal properties. When grown with low grain size like in the case of UNCD, the film will act as a thermal barrier toward the substrate. When outgrown with larger grain size (or in a 2D-NCD configuration), the heater itself and its connection to the metallic contacts will be an efficient heat spreader.

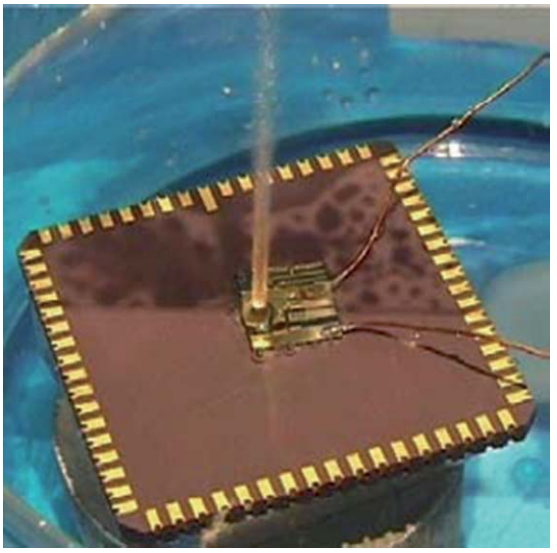


Figure 33 Diamond heater with 1-mm^2 surface area in water driven by a 10 μs pulse train of 200 V at 50 Ω and with 2 kHz repetition rate, resulting in microfountain action. Height of the water column is approximately 20 cm.

1.06.4.2.2 Electrostatic actuator

In the electrostatic case, the two opposite plates of an air gap capacitor are attracted by an applied field, the limit being the breakdown field either across the air gap or the anchor isolation. For a constant driving voltage, the force is a hyperbolical function with displacement, whereby the displacement changes along the beam. If the cantilever is curved, as it is often the case for a vertically gradually stressed films, additional to the deflecting force a bending moment M_z needs to be taken into account (Schmid 2006). Because the beam is increasingly accelerated toward the counterplate, this results in small forces in the beginning, and in snapping, when the separation becomes small. Therefore, it is difficult to hold the beam in an intermediate position. Thus, this driving principle is predestined for switching across small gaps but not the adjustment of a displacement. An overview for the material selection of electrostatic actuators can be found in (Srikar and Spearing 2003).

Figure 34 shows the switching speed for an identical geometry of a cantilever fabricated of diamond, AlN, SiC, and Si, respectively. It can be seen that the stiffer material needs a higher voltage for the displacement to close the switch. However, applying high driving voltages (approximately five times the threshold voltage) all cantilevers can be driven with comparable speed. Therefore, the high YM is only a minor disadvantage, when high driving forces can be

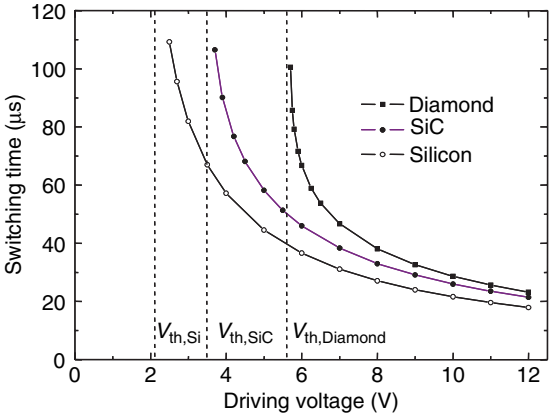


Figure 34 Switching time versus switching voltage of single-ended cantilevers of identical geometry for Si, SiC, and diamond.

generated. However, this means also high driving voltages. It holds especially for curved beams, with high tip displacement. High driving voltages may interfere with the signal voltage. This is especially true when operated at high temperature, where the isolation of the anchor is reduced due to temperature-activated leakage. Otherwise, this actuation principle is not temperature sensitive. The switch stays closed as long as the voltage is applied, however, no DC loss occurs.

1.06.4.2.3 Piezoactuator

In the case of piezoactivation, the driving force is the induced strain in a piezoelectric crystal due to an external electrical field. In a certain range, the interaction between the vector of the electrical field $\mathbf{E}_{el}$ and the vector of induced strain $\mathbf{S}$ can be described by a linear relationship and a matrix of piezoelectric coefficients.

Perovskites like $BaTiO_3$, $KNbO_3$, $KTaO_3$, $SrTiO_3$, and $PbTiO_3$ are typical piezoelectrics with large piezoelectric coefficients. They are deposited by techniques like MOCVD (Zhao *et al.* 2004), sol–gel spin-on (Zeng *et al.* 1999), pulsed laser deposition (Wan *et al.* 2002), or ECR-PECVD (Chung *et al.* 2000). After deposition, they generally need to be poled by an annealing step under bias, where the individual dipoles are oriented. They are then operated in a regime of temperature and field, where this orientation is maintained. A commonly used compound of $SrTiO_3$, and $PbTiO_3$ is PZT (lead zirconate titanate) (which can however have various composition ratios). It can show high piezoelectricity up to a Curie temperature of approximately 450°C and a linear elastic regime up to approximately $10\,kV\,cm^{-1}$, where depolarization effects become apparent (Klaasse *et al.* 2004). Thus, the range of application is rather limited. Nevertheless, high driving forces can be developed.

A material with high polarization and high-temperature stability is wurzite AlN. It is deposited with a preferred orientation of the polarization, which is usually perpendicular of the surface (*c*-axis). To avoid random orientation in sputtering processes, orientation is obtained by the self-bias developed during deposition or an external DC bias and bombardment of Ar ions, providing the energy for the surface ion reorientation (Elmazria *et al.* 2004, Olivares *et al.* 2005).

For use in MEMS structures, these materials are deposited as thin films and for narrow thin-film cantilevers two common modes are exploited, namely the d_{31} mode and d_{33} mode (see, for example, APC 2002) (**Figure 35**).

In a thin-film arrangement, the d_{31} mode is used with vertical polarization (*z*-direction in **Figure 35**), and the dipoles of the ferroelectric have to be poled

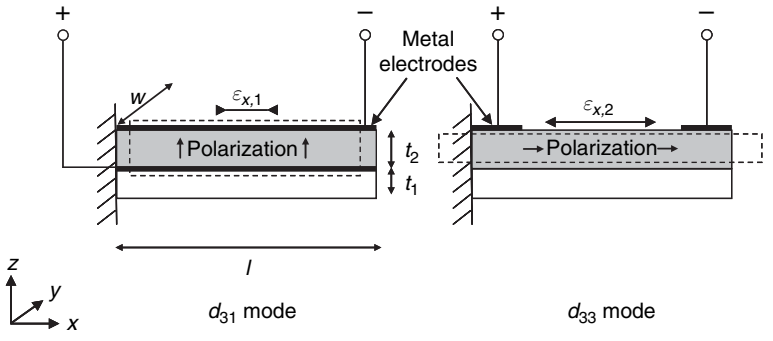


Figure 35 Schematics of common modes used in piezoelectric unimorph cantilever actuators.

Table 5 Materials parameters of piezoelectrics

	d_{31} (pm V^{-1})	d_{33} (pm V^{-1})	Curie temperature (°C)	Depolarization field (V cm^{-1})
PZT	−180	360	~350	2×10^4
AlN	−2	4	>1000	5×10^6
$BaTiO_3$	−78	190	118	3×10^4
$LiNbO_3$	−0.85	6	1150	–

in this direction. In this mode, the metal electrodes are arranged on top and bottom of the film, and the mechanical response perpendicular to the electrical field acts with $\varepsilon_{x,1}$ in direction of the cantilever beam length (**Table 5**).

The approach has recently also been applied to PZT/diamond unimorph. The cantilever used had been a 40-μm-long UNCD beam of 1.0-μm thickness with a top PZT layer of 0.1-μm thickness deposited by MOCVD (Srinivasan *et al.* 2006). A tip deflection of approximately 2 μm for a driving voltage of 5 V was demonstrated. With its high piezocoefficients, the PZT/diamond pair will generate high contact forces and high resonance frequencies, it will however only operate in a limited temperature range. The other alternative would be AlN/diamond with smaller piezocoefficients of AlN, but with the ability to operate at high temperature. **Figure 36** shows a calculation of the contact force, which can be generated with an AlN/diamond cantilever of different length operating in the d_{31} mode. In this calculation, a deflection of 2 μm has been assumed. The maximum field applied to AlN was 1 MV cm^{-1}, which is approximately two orders of magnitude higher than possible for the case of PZT. To achieve the full potential, high voltages need to be applied or a multiplayer stack with intermediate contact plains would be needed. The graph illustrates that high contact forces can be obtained, but also that the structural design is critical.

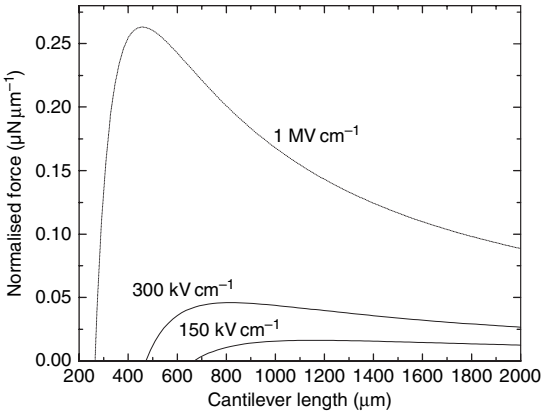


Figure 36 Simulated contact force (normalized on the cantilever width) of an AlN/diamond unimorph actuator in the d_{31} mode as a function of cantilever length (layer thicknesses: $t_{AlN} = 3.2$ μm, $t_{dia} = 2.0$ μm). The cantilever had a separation of 2 μm with the substrate. There is no contact force before the cantilever touches the substrate.

The second possibility is the exploitation of the d_{33} mode. Here, the piezoelectric film needs to be poled horizontally and thus parallel to the surface. For single-crystal wurzite AlN films this means that they need to be grown laterally in M face direction, which is difficult to obtain. In the d_{33} mode, case contacts are deposited at the ends of the structure and an electrical field applied in x-direction (**Figure 36**). Usually the arrangement is an interdigitated finger structure and large forces can also be obtained for long beams with low driving voltages. With d_{33} being normally larger than d_{31}, larger deflections can be obtained. Taking the case of diamond, which possess the highest YM, a PZT film on a diamond cantilever poled for d_{33} mode can expand by more than 500 ppm and can generate compressive stresses of more than 350 MPa (for a YM of PZT of 70 GPa). This value is larger than the stress, which can be generated by the electrostatic actuation and is in the same order of stresses, which can be generated by thermoelectric actuation (bimetal effect, see below). The static power consumption is determined by the electrical loss of the piezoelectric and can be extremely low for highly insulating films.

1.06.4.2.4 Bimetal actuator

The thermoelectric actuation is based on the thermally induced strain in a material in its elastic regime. As bimetal actuator, consisting of a two-layer stack it makes use of the thermally induced strain differences between the two layers. The vertical stress $\sigma_x(z)$ developed across bimetal pair is

$$\sigma_x(z) = E_i(\varepsilon_x - \alpha_i \Delta T) \qquad [2]$$

where α_i is the difference in the coefficient of thermal expansion of the two layers and ε_x is the strain induced by a nonthermal intrinsic stress and external loads.

Both SiC and diamond possess low coefficients of thermal expension, show no plastic range before fracture, and are elastic over a wide range in temperature. They are therefore ideal in combination with metals of high coefficient of thermal expansion (**Table 4**). With high temperatures, high bending moments can be generated, the limit imposed by the interfacial adhesion of the films and high-temperature softening of the metal. Thus, such metals are usually hard metals like Ni and Cu.

High bending moments can be obtained using a materials combination with a high YM for the beam structure and large coefficient of thermal expansion for the bimetal. A proper figure-of-merit is the product $E_{\text{beam}} \times \sigma_{\text{bimetal}}$ (**Table 2**). Diamond possesses the highest figure of merit and **Figure 37** shows the tip deflection achieved by a bimetal pair of nickel/diamond. The actuation is largely linear and no snap-in effect is expected. Such cantilevers can therefore also be used for a defined tip placement.

This quasi-static relationship is obtained after the heat flow is balanced. Thus, fast switching requires a fast heat flow, which on the other hand means a high heating power level for standby. To circumvent this problem, bistable or monostable bridge configurations are needed. They require tight geometrical control and stress engineering of the films. A diamond-based bistable bridge actuator is discussed in Section 1.06.6.2.

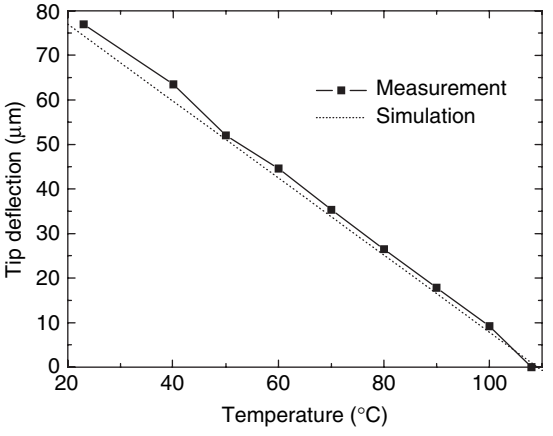


Figure 37 Tip deflection of electrothermally actuated diamond cantilever with Ni bimetal. (Source: Schimid P, Hernandes-Gullien F J, Kohn E 2003 Diamond switch using new thermal actuation principle. *Diamond Relat. Mater.* **12**, 418–21, with permission.)

1.06.5 Technologies

Microsystems technologies are very heterogeneous, depending on the application. Nevertheless, most technologies are aimed to be compatible with CMOS technologies. If the sensor technology is not CMOS compatible, discrete device structures are hybrid integrated with a systems carrier. The materials discussed here, are aimed at applications in extreme conditions, beyond the capability of Si. Therefore, the development is still largely on the basis of discrete devices. Elements are bulk machining for membrane structures, surface machining for 3D structures including a sacrificial layer technology. These will be discussed in the following sections.

1.06.5.1 Bulk Machining

Bulk machining describes processing technologies based on substrate etching. It had originally been developed for the etching of Si for membrane fabrication and the preparation of molds.

Hexagonal SiC single-crystal wafers (4H and 6H) are available and can be used as substrate for all-SiC MEMS. The SiC substrate can be dry etched using a high-density fluorine-based plasma by DRIE or inductively coupled plasma-reactive ion etching (ICP-RIE) (Beheim and Evans 2006), mostly containing digital etch/passivation cycling (Bosch process, Lämmer and Schlip). Such a process has been used to obtain freestanding SiC cantilevers (Hossain *et al.* 2006).

Single-crystal diamond is not available in wafer size and polycrystalline diamond wafers, which are several hundreds of micrometers thick and polished, are mainly used as quasi-substrates for optical windows or heat sinks (Wild 1998, Wörner 1998). They are also etched in oxygen containing plasma of high etch rate by ICP-RIE.

The majority of MEMS processes in SiC and diamond therefore make use of an Si substrate. Thus, to a large extent the Si bulk machining processes have been copied onto the case considered here, where the ceramic-like semiconductor is the membrane material or the active pattern realized by the mold. The Si dry etching follows the normal routine used in Si-MEMS and is either by wet etching in hot KOH or dry in a high-density plasma.

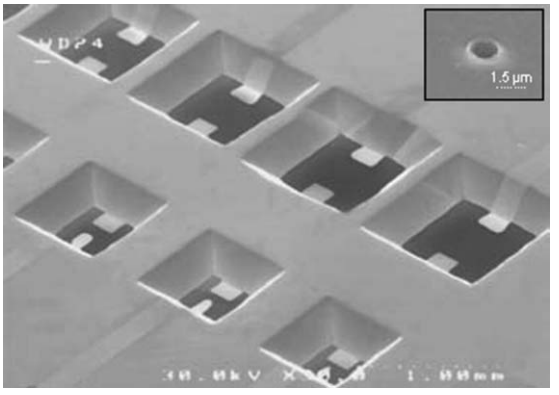


Figure 38 Planar cell patch clamp array with nanocrystalline diamond (NCD) membrane and Pt electrodes on front and rear side (the diamond film is transparent to the e-beam exposure). In the inset, the clamp hole located in the center of each compartment is shown. (Source: Kusterer J, Alekov A, Pasquarelli A, Müller R, Ebert W, Lehmann-Horn F, Kohn E 2005 A diamond-on-Si patch-clamp-system. *Diamond Relat. Mater.* **14**, 2139–42, with permission.)

Vertical sidewalls of high aspect ratio are again obtained by DRIE. Wet etching by KOH results in (111)-oriented slopes.

Figure 38 illustrates an example of a Si bulk machining process for diamond membrane fabrication, showing an array for a planar patch clamp system (Kusterer *et al.* 2005). In the SEM micrograph, the diamond compartments with the membrane are transparent to the e-beam exposure. In the center of the membrane, a clamp hole of approximately 1-µm diameter has been etched. In this application, the compartments are filled with liquid and sloped walls as obtained by Si wet etching in hot KOH are advantageous. The top and bottom Pt electrode metallizations are clearly visible.

1.06.5.2 Surface Machining

Surface machining is the extension of the microelectronics technology into the third dimension to obtain freestanding beams for cantilevers, bridges, hinges, and gears and for enclosed structures like capillaries and cavities.

The desired relief structure may be obtained by metal plating of a prestructured pattern, undercut etching, or conformal deposition onto a prefabricated pattern with subsequent etching of the sacrificial support material (molding). The sacrificial layer technology is especially a key aspect in the fabrication of closed structures like capillaries and cavities. Sacrificial layer materials used are SiO_2, metals and polyimides deposited by evaporation, plating, plasma CVD, or spin-on. Top layers are usually poly-Si, poly-SiC, and metals like Ni. The sacrificial layer is etched selectively, generally by a liquid solution, with respect to the deposited film. Such a technology has been developed for SiC and diamond structures, however, not as yet for GaN- and AlN-based devices.

In the case of SiC, such patterns are usually not grown by selective epitaxy or back etching, but by overgrowth of the relief structure of a sacrificial SiO_2, SiN_4, or poly-Si layer (Cheng *et al.* 2002). This step is then followed by polishing the surface until the isolated SiC pattern is left. Finally, the mold is etched completely to obtain freestanding structures, or the SiC structure processed further (Mehregany and Zorman 1999). After conformal overgrowth of the mold by the SiC film, its surface is in general not planar and strongly textured and needs to be polished (using diamond powders). In full MEMS processes like the MUSiC foundry process originating from Case Western Reserve University (CWRU, Ohio) (Trade mark, used by FLX Micro, Ohio) the realization of a SiC mold can be realized three times for the stacking of three layers in total.

With respect to diamond, molding has been used to realize field emitter arrays (Davidson *et al.* 2003), microchannels (Ralchenko and Givargizo 2003), and AFM tips (Beuret *et al.* 2000, Kulisch *et al.* 1997). In all cases, the molding material was Si. A flow chart for the realization of a micro emitter array and individual field emitter elements are shown in **Figure 39(a)** and **39(b)**. As can be seen, tip ends can be very sharp depending on the Si mold preparation.

In diamond, back polishing is difficult because of its hardness. Instead, selective growth and dry etching (in an oxygen plasma) have been used extensively (Auciello *et al.* 2004, Kohn *et al.* 1999, Rangsten *et al.* 1999, Yang *et al.* 1999). This is illustrated in **Figure 40**, where the process flow for an electrostatically driven cantilever switch is shown (Adamschik *et al.* 2000).

The main sacrificial layer material is SiO_2. However, it is difficult to deposit with a thickness of several micrometers without stress and cracking. Here, plated metal structures are an alternative. Such a process with plated Cu as sacrificial layer material has been developed for diamond (Müller *et al.* 2004). The process is sketched in **Figure 41** for the case of the fabrication of an all-diamond capillary, where the liquid will only be in contact with a diamond surface.

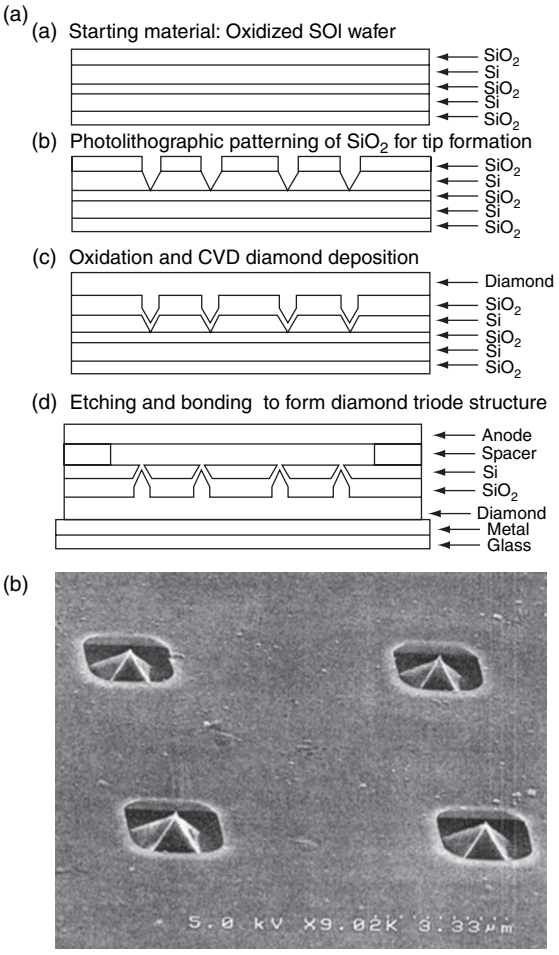


Figure 39 (a) Flow chart of the fabrication routine of a diamond field emitter array. (Source: Davidson D, Kang W P, Wisitsora At 2003 Diamond field emission devices. *Diamond Relat. Mater.* **12**, 429–33 with permission.) (b) Individual field emitter elements realized by the process as sketched in (a).

It was designed for applications in aggressive media like the all-diamond ink-jet, which is discussed below.

1.06.5.3 Surface Processing

During processing, surfaces have been in contact with many materials like metals, hard dielectrics, and polymers. They have to withstand high temperatures, aggressive liquids, and thermal stresses. Although being inert, they need to provide high chemical activity in the liquid and good adhesion and controlled interdiffusion in contact with solid films.

1.06.5.3.1 Ohmic contacts

Contact metallization schemes have been developed mainly in connection with electronic and optoelectronic device structures, and there exists extensive literature on this field. They are in general based on silicide-, nitride-, or carbide-forming conducting alloys with refractory metals like Ti, Ni, or W. They are therefore alloyed at high temperature, even when tunneling is provided by high doping. Nevertheless, although annealed at high temperature, in most cases, the devices are not operated at high temperature. Thus, the reliability of each contact system is still critical. In general, for devices operating in aggressive media contacts are passivated or hermetically encapsulated.

SiC, Ti, Ni, Cr, Ti:Ni, and Ti:W have been extensively used and annealed at temperatures up to 1000°C. Si diffusion barriers have been TiN and $TaSi_2$. The complete stacks used have been Ni/Cr/W, Ti/TiN/Pt, and Ti/$TaSi_2$/Pt. For high-temperature gas sensing, ohmic contacts used were $TaSi_x$ with Pt overlay (Spetz *et al.* 2001).

For GaN/AlGaN heterostructures Ti/Al/Ni/Au, annealing above 850°C has become a standard contact system, which has allowed short time operation up to 1000°C (in the case of AlInN/GaN FETs) (Ruvimov *et al.* 1996). Both Ti and Al have a small work function difference to Ga, which dominates the Ga face surface properties (Schmitz *et al.* 1998).

With respect to diamond, contacts have been developed to the hydrogen- and oxygen-terminated surfaces. The hydrogen-terminated surface is chemically unstable and contacts are not reliable under harsh conditions. On oxygen-terminated surfaces, the surface potential is pinned at approximately 1.7 eV and contacts are best provided on degenerately boron-doped layers by tunneling. Contacts have been developed around Cr/Au and Ni alloyed at a moderate temperature. For high-temperature operation, a W:Si/Pt/Au contact has been used on highly boron-doped contact layers (Vescan *et al.* 1996).

1.06.5.3.2 Schottky contacts

Schottky contacts should form a reliable (temperature stable) interface with the semiconductor. SiC, Ni, and $TaSi_2$ have been found stable up to around 600°C (Saxena *et al.* 1999). However, most contacts for gas sensing are based on Pd or Pt.

In the case of GaN, the surface potential should not be pinned as seen from the dependence of the barrier height on the work function difference

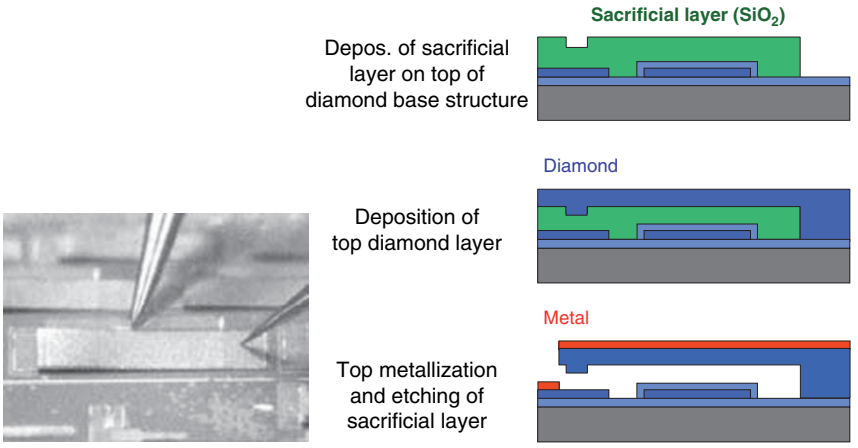


Figure 40 Sketch of SiO_2 sacrificial layer technology in diamond for realization of cantilever switch. (Source: Adamschik M, Schmid P, Ertl S, Ebert W, Käb N E Kohn 2000a CVD-Diamond for high temperature MEMS applications: First example: Diamond microswitch. *Proc. 5th International High Temperature Electronics Conference (HiTEC)*, Albuquerque, NM, USA, June 2000 with permission.)

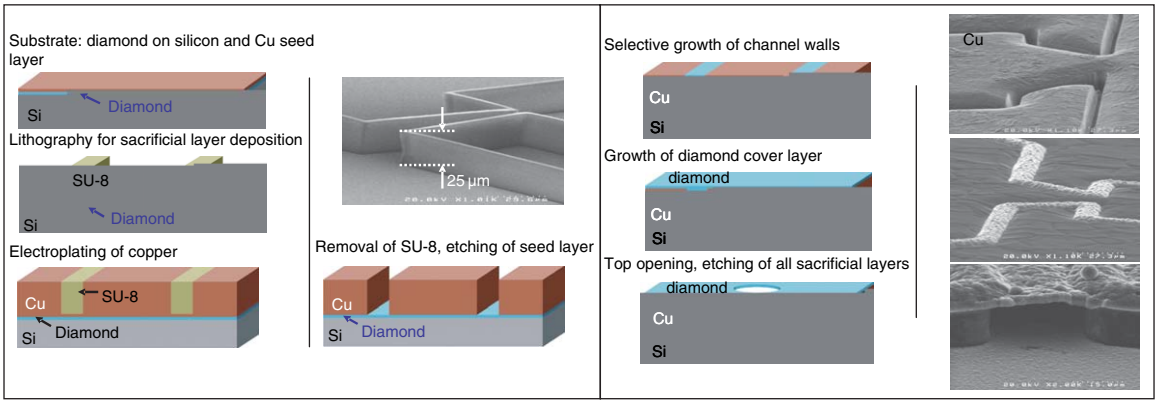


Figure 41 Sacrificial layer technology for diamond capillaries, based on Cu-plating, SU-8 patterning, and nanocrystalline diamond (NCD) overgrowth. (Source: Müller R, Schmid P, Munding A, Gronmaier R, Kohn E 2004 Elements of surface microfluidics in diamond. *Diamond Relat. Mater.* **13**, 780–4, with permission.)

(Schmitz *et al.* 1998). Therefore, Schottky barrier materials are mostly Pt and Ni. However, on the technical surface, a native oxide coverage may determine the interface and pin the surface potential. This is especially observed for AlGaN, where for an Al content of 30%, a barrier height of approximately 1.5 eV is observed (Ibbetson *et al.* 2000).

With respect to diamond, reliable Schottky barrier contacts are only possible on oxygen-terminated surfaces. Here, the barrier height is pinned at approximately 1.7 eV, with an experimental spread between 1.5 and 2.0 eV. High-temperature stable contacts have been realized with W:Si/Au, where the W:Si had been kept amorphous by addition of nitrogen during deposition (Vescan *et al.* 1998).

1.06.5.3.3 Passivation layers

Passivation layers should normally protect the active semiconductor surface from the surrounding environment. In sensors the active surface is exposed to the environment and passivation is needed to protect the peripheral areas. Depending on the environment, passivation layers need to have good adhesion, also during stressing, they are usually chemically inert, corrosion-resistant dielectrics (acting as diffusion barriers to water diffusion) and with wide bandgap

(and thus high electrical breakdown strength). They are therefore in general Al_2O_3, SiO_2, Si_3N_4, and nanocrystalline AlN in the case of diamond (McGoech *et al.* 1999). In electrochemistry, epoxies are used extensively.

1.06.5.3.4 Heteroepitaxial growth and interfaces

Most wide bandgap semiconductors are grown on foreign substrates with high mismatch, preventing direct overgrowth with low dislocation density. Therefore, in nearly all cases, a thin buffer layer is acting as nucleation layer for oriented outgrowth. Epitaxially ordered layers are formed already a few nanometers above this film. On Si, such buffers are for SiC, a carborization layer to nucleate SiC (Morales *et al.* 2003), for GaN and AlN a low-temperature AlN buffer, on Al_2O_3, SiC, or Si as discussed in Section 1.06.3.2, and for diamond a carborization step together with bias enhancement of nucleation (Yun and Dandy 2005). In the case of cubic SiC and diamond, the overlayers are textured. In the case of diamond for growth of nanocrystalline layers, seeding is often performed with nanodiamond particles deposited from a suspension or with nanoparticles attached to a plasma-generated a-C layer (Rotter 1999).

1.06.5.3.5 Etching

All ceramic-like semiconductors discussed here are highly etch resistant. Etching in wet solutions is difficult, but generally possible by photochemical reaction when illuminated with above bandgap UV illumination. In the case of SiC, Si, diamond etching of deep trenches is often performed with high density, high energetic plasmas (ICP-RIE), often with chemically assisted erosion by addition of Ar. Deep trenches with vertical walls and high aspect ratio are usually etched in by cycling between etching (e.g., using dilute SF_6) and sidewall passivation by polymerization (e.g., using dilute CHF_3) in a large number of sequences (Bosch process) (Lämmer and Schlip).

SiC can be etched in hot KOH at approximately 600°C (which will also remove Si at a higher rate). Wurzite SiC has also been etched electrochemically in HF (Chang *et al.* 2006, Mikami *et al.* 2005) or in KOH-solutions, activated by UV light (Kato *et al.* 2003). Thus, it is usually dry etched (by ICP-RIE as mentioned above) in fluorine-containing gases like SF_6 or NF_3 mixed with Ar and O_2 (Beheim and Evans 2006). Such etching processes have been used to obtain suspended microstructures in bulk 6H–SiC substrates (Hossain *et al.* 2006) and to obtain nanometer-sized double-clamped SiC beams by underetching of a SiC layer on a Si substrate (Yang *et al.* 2001).

GaN with N face polarity is etched in KOH (Li *et al.* 2001). Material with Ga face polarity is etched in KOH only when assisted by DUV light (Youtsey *et al.* 1997). However, GaN, AlGaN, and InGaN can be dry etched in plasmas mostly containing Cl for chemical activation or Ar for physical activation by chemically assisted ion beam etching (CAIBE) (Ping *et al.* 1996), ECR, RIE, or ICP (Basak *et al.* 2000, Pearton *et al.*, 1994, Shul *et al.* 1996). Controlled etching needs to be performed in a mild plasma, to prevent loss of nitrogen. High precision is obtained by digital etching, where in sequence a nanometer-thin oxide layer is formed, which is subsequently removed (Buttari et al. 2002). InGaN has been used as thin sacrificial layer grown between two GaN layers and removed by photoelectromechanical etching (Stones *et al.* 2001). Photochemical etching has also been used in conjunction with selective etching on PN structures (Stones *et al.* 2001, Strittmatter *et al.* 2001). Because the bandgap of InGaN is smaller than that of GaN, at the chosen wavelength, the GaN film is not absorbed and thus not etched. AlN has been dry etched in an Ar/C_2 and H_4/H_2 ECR plasma (Pearton *et al.* 1994).

AlN is etched by KOH solutions, depending on quality. While the N face is attacked by the etch solution, the Al face is not etched, presumably due to oxide passivation (Vartuli *et al.* 1996, Zhuang *et al.* 2002). Diamond cannot be etched in any wet solution, but is dry etched by oxygen. The process is critical, because the texture of the nanocrystalline film may result in a roughened surface or even needles. Masks that are used need to be oxygen resistant and are therefore mostly metals like Ti/Al. Perpendicular sidewalls can be obtained depending on the plasma conditions as well as by sidewall passivation by metal oxides, originating from back-sputtered mask material (Ding *et al.* 2005). Etching of nanocrystalline material in an O_2 atmosphere becomes noticeable above 750°C, starting at grain boundary sites. Microbody structures have been etched and patterned by a focused ion beam (Murakawa *et al.* 2003).

1.06.5.3.6 Chemical functionalization

Most functionalized surfaces in biochips are metal-oxide surfaces using the hydroxyl group as linker. Their reactivity is generally high, and are used

oxides with high stability. This excludes in many cases SiO_2 and therefore also the native oxide on SiC. GaN and AlN have been functionalized with organosilane (Baur *et al.* 2005).

In the case of diamond, on the other hand, the high intrinsic stability of its surface may make its functionalization difficult. But, indeed the diamond surface can provide stable bridges to biomolecules through specific reactive groups. These functional groups can then be used as a starting point for the attachment of more complex molecules such as DNA or proteins to the surface (Härtle *et al.* 2004, Wenmakers *et al.* 2005). Clean surfaces, oxygen-, and hydrogen-terminated surfaces have been modified and functionalized by electrochemical and photochemical processes to form covalent C═C links, carboxylic acid groups, N–C bonds, and C–F groups (Freedman and Stinespring 1990, Ohtani *et al.* 1998, Tsubota *et al.* 2004). To date, the photochemical method is mostly used as developed by Hamers (Nichols *et al.* 2005) with UV illumination (Hg line at 245 nm) of hydrogen-terminated diamond to produce a carboxyl (C═C) link. This process involves electron/hole pair generation and electron emission into the liquid. The main advantage seen is that the C═C linker is extremely stable, for example, in the hybridization of oligonucleotides (Yang *et al.* 2002).

Most biomolecules are charged. Therefore, detection of the covalent linkage can be via field effect. This is indeed used in ChemFET structures based on (H induced) surface channel devices (which are however only conditionally stable as already mentioned) (Garrido *et al.* 2005, Strobel *et al.* 2004). After functionalization, direct electron transfer could be observed between the redox center of an enzyme and the diamond electrode (Härtle *et al.* 2004). This means diamond can combine biofunctionalization not only with ISFET but also with electrode properties.

1.06.6 Diamond-Based Applications

Some applications have already been discussed, where the complexity and technological requirements have been limited. These have been electrochemical electrodes, radiation detectors, and GaN-based piezosensors. In the following, three examples based on CVD diamond with performances under extreme and harsh conditions are discussed to further illuminate the potential of this young material. While SiC can already be explored commercially in foundry processes, diamond microsystems are just emerging.

1.06.6.1 Acceleration Sensor

Both SiC and diamond possess high stiffness and high fracture strength. They are, therefore, especially suited for applications, where high forces are developed. The following is an example, where a diamond membrane has been structured into a bridge structure with Si seismic mass for sensing shock and fast acceleration (**Figure 42**). The thickness of the diamond film had been 15 µm, the YM 850 GPa (approximately 80% of the ideal value), and the fracture strength (bending the cantilever downward) was 4.7 GPa (approximately half of the ideal value), the data obtained by resonance frequency measurements of

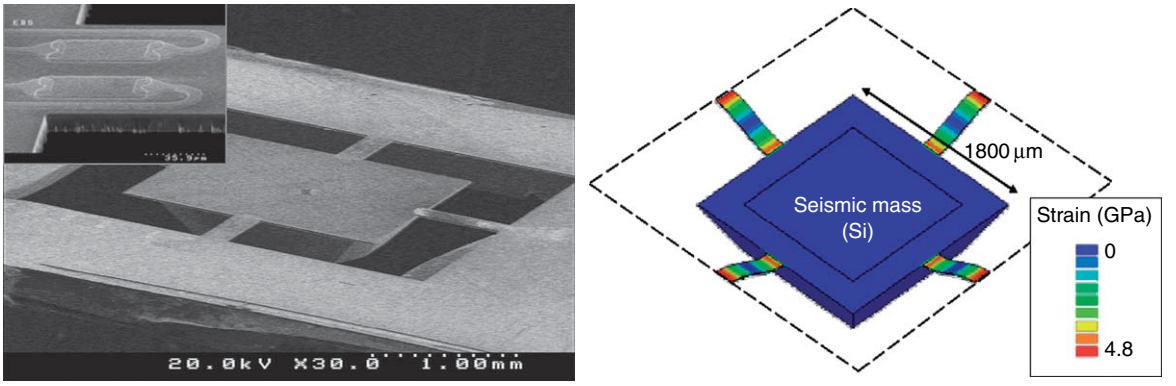


Figure 42 Bridge acceleration sensor with Si center seismic mass. Piezoresitors shown in the inset. *On the left side* simulation shows displacement at 7000 ***g*** acceleration, creating a critical stress of 4.8 GPa, leading to fracture. (Source: Kohn E, Gluche P, Adamschik M 1999 Diamond MEMS – a new emerging technology. *Diamond Relat. Mater.* **8**, 934–40 with permission.)

single anchored cantilevers and nanoindention measurements. The active elements are boron-doped piezoresistors integrated at the anchor position (see inset in **Figure 42**) realized as described in Section 1.06.4.1.1. In this case, at moderate activation of the boron doping (indicating a doping level above $10^{18}\,cm^{-3}$), the GF was 8.0. The piezoresistors were arranged in a fourfold bridge. Two resistors were located at the end of the beams, two outside the suspended structure for reference. The signal was read out by a Wheatstone bridge circuit on a circuit board. The accelerator was tested in shock experiments up to 40 GPa. However, it was realized that this had been earlier than the fracture of bridges could be expected. The fracture limit of the sensor was simulated using the measured materials data assuming the experimental fracture strength. This is shown in **Figure 42** (right). As can be seen, the critical stress appears at the outside anchor position of the bridges. At this condition, the acceleration was 7000 $\boldsymbol{g}$ (Kohn *et al.* 1999).

1.06.6.2 High-Temperature/High-Power Electrostatic Switch

Microswitches are a basic component of RF MEMS because of their low insertion losses and high current switching capability. They are either discrete devices or part of integrated circuits (ICs). In most cases, they are cantilevers or bridges and either DC or RF coupled and electrostatically actuated. Materials used are poly-Si or metals like Ni. Interstructural insulation is often provided by oxides or nitrides with their mechanical and thermal properties essentially deviating from those of the conductor materials. Both SiC and diamond are attractive as cantilever materials because of their high stiffness translating into high switching speeds, but also high switching voltages as discussed in Section 1.06.4.2.2. Both materials also open up the possibility of high-temperature and high-power operation.

The case discussed is an all-diamond microswitch, the advantage being that diamond is used as multifunctional material to replace all the materials normally needed in the stack. The basic technology has been described in Section 1.06.5.2. Some more details are given in Ertl *et al.* (2000) and Adamschik *et al.* (2002) and may be apparent from **Figure 43**. The switch has been designed as a DC coupled switch for RF applications to switch bias level and signal simultaneously.

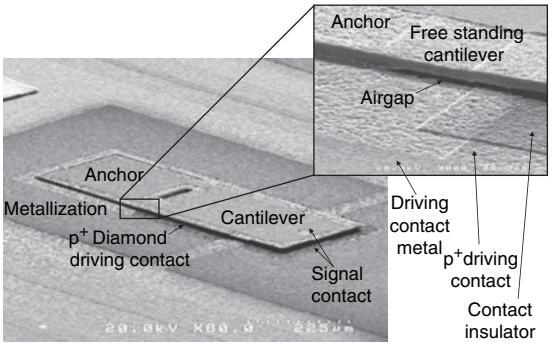


Figure 43 Basic structure of electrostatic cantilever switch. The various parts have been realized by selective growth and SiO_2 as sacrificial layer material. (Source: Adamschik M, Kusterer J, Schmid P, Schad K B, Flöter A, Kohn E 2002 Diamond microwave micro relay. *Diamond Relat. Mater.* **11**, 672–6.)

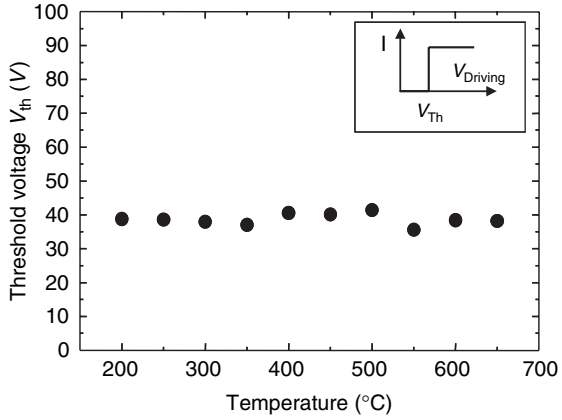


Figure 44 Threshold voltage for switching versus operating temperature. (Source: Adamschik M, Schmid P, Ertl S, Ebert W, Käb N E Kohn 2000 CVD-Diamond for high temperature MEMS applications: First example: Diamond microswitch. *Proc. 5th International High Temperature Electronics Conference (HiTEC)*, Albuquerque, NM, USA, June 2000.)

The switching properties had been evaluated up to 650°C. In essence, the threshold voltage remained constant, indicating no change in YM due to plastic deformation (**Figure 44**) (Adamschik *et al.* 2000). Mechanically, the device was switched in a vacuum of up to 800°C. In this experiment, the cantilever was deflected by a micromanipulator. No sticking or fatigue was encountered.

The DC on/off ratio of the switching resistance was $>10^9$ at RT. At high temperature, leakage across the diamond anchor, the diamond base plate, and Si

substrate became noticeable in this specific case. However, even at 800°C, the ratio was still better than 10^4 (**Figure 45**). This residual conductivity was thought to be generated by activated grain boundary conduction, activation of nitrogen, which has been used to support nanocrystalline growth and the activation of residual boron.

For an allowed temperature of the contact tip during operation at 600°C, which may be realistic for operation under vacuum, a DC current density of more than 250 kA cm^{-2} should be permissible. The simulation was based on thermal conductivity values of 25% of the ideal value for the nanodiamond cantilever and 50% of the ideal value for the base plate of the highly oriented diamond (Adamschik *et al.* 2000a). This result had stimulated overdrive experiments to determine the breakdown limit. The result of such an experiment is shown in **Figure 46** (see also Adamschik *et al.* 2000b). The thermal overload resulted in burning a hole into the cantilever due to reaction with atmospheric oxygen. Molten Au is piled up around the hole. Thus, the local temperature reached was essentially higher than 1000°C. Based on the contact resistance and contact area measured before destruction, a power density of 0.74 MW cm^{-2} was applied. Nevertheless, the cantilever did not stick and switching was possible even after the destructive event (as shown in the inset).

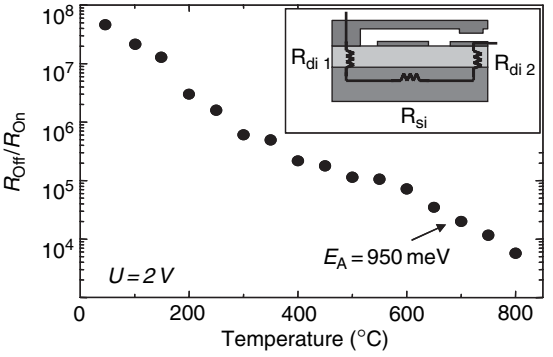


Figure 45 Ratio of the switch resistance in the off-state and on-state versus operating temperature. The reduction of R_{off} with temperature is mainly caused by activation of deep levels in the undoped nanocrystalline diamond (NCD) film, the bypass current path also being indicated. (Source: Adamschik M, Schmid P, Ertl S, Ebert W, Käb N E Kohn 2000 CVD-Diamond for high temperature MEMS applications: First example: Diamond microswitch. *Proc. 5th International High Temperature Electronics Conference (HiTEC)*, Albuquerque, NM, USA, June 2000.)

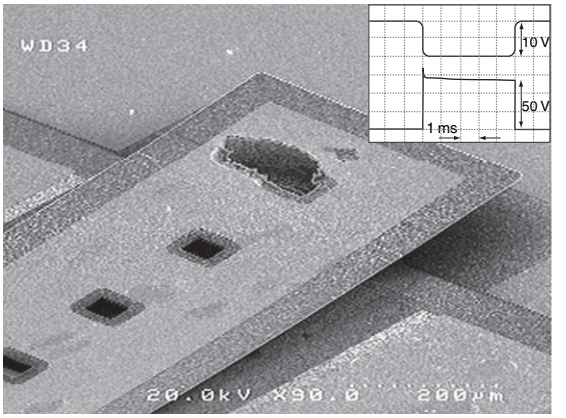


Figure 46 Cantilever switch after overdrive experiment (see text) also shown is the switching characteristics afterward. (Source: Adamschik M, Shmid P, Ertl S, Gluche P, Flöter A, Kohn E 2000b Performance of high speed diamond micro switch. *Proc. MicroMat 2000 (3rd Int. Conf. Poster Exhibition Micro Materials)*, Berlin, Germany, April 2000, pp. 863–866 with permission.)

1.06.6.3 Ink-Jet for Aggressive Media

Ink-jet devices are one of the most popular microfluidic devices, mainly used for ink printing. Many different, mostly heterogeneous, configurations with various actuation principles have been developed. Mostly they are Si-based MEMS devices, containing either a thermoelectric or a piezoelectric actuator element, a polymer capillary system, and a metal nozzle plate (Aden *et al.* 1994).

While they are very well optimized for their application, the hybrid nature of the materials stack imposes strong limits on their operating environment. The materials employed may not be compatible with aggressive chemicals and/or organic solvents because of corrosion of the metal nozzle plate or dissolution of the polymer capillary system and/or the glue used for bonding. In the case of the bubble jet with electrothermal actuation, thermal stress due to different thermal expansion coefficients of the individual layers within the stack may result in cracking or even delamination. This limits the lifetime of such devices as well as their range of operating temperature and their compatibility with aggressive fluids and organic solvents. Other actuator elements other than the spot heater are difficult to integrate and therefore rarely used in on-chip systems.

With respect to the bubble jet concept, various approaches can be found in the literature to facilitate the technology for on-chip integration. For example, to avoid a polymer capillary system, Si bulk

micromachining is used, but requires the heater integration on the nozzle plate (backshooter) (Krause *et al.* 1996, Lee *et al.* 2002). In another approach, capillaries have been realized in an SOI approach (Baek *et al.* 2004). Usually planar capillaries are realized by a high aspect ratio sacrificial layer technology, employing SiO_2, Ge, and SiN (Wang *et al.* 2004). Heater materials are refractory metals like Pt or refractory compounds like HfB_2 and need passivation.

Therefore, the replacement of the entire stack by only one material seems desirable. Both SiC and diamond are heater materials and highly corrosion resistant and seem thus to be appropriate. In the approach described here, diamond (NCD) has been used as a multifunctional material. The entire chamber including the heater and capillary structure has been realized in a diamond-on-Si technology and has thus been labeled an all-diamond ink-jet. Therefore, the entire structure is chemically inert, corrosion resistant and thermally stable and can be used to dispense any liquid like concentrated acids or hot KOH.

A key requirement for microfluidic elements is a technology for encapsulated microchannels. They have been fabricated by conformal deposition of NCD and sacrificial layer etching. High aspect ratios are needed, which are difficult to obtain with SiO_2. Therefore, a sacrificial layer technology based on electroplated Cu has been developed as described above. In this process, the Cu sacrificial layer has been deposited onto a sputter-deposited seed layer and plated in between an SU-8 pattern with steep sidewalls. The Cu pattern was overgrown with another NCD film after deposition of a diffusion barrier and nucleation layer. Despite the different thermal coefficients of expansion of diamond, the Si substrate and the Cu, no cracking was observed during overgrowth at 700°C. The heater and wire connections within the chamber have been realized by quasi-metallic p-type-doped diamond resistors, need no passivation, and are highly corrosion resistant. They were fabricated before the capillary walls were grown as already discussed in Section 1.06.4.2.1. To open the capillary, the Cu pattern was etched in 30% HNO_3. The resulting ink-jet pattern is schematically shown by the cross section in **Figure 47** and the micrograph of the realized structure is shown in **Figure 48**.

In the bubble jet, actuation is generated by superheating of the liquid. By applying short heating pulse, a thin liquid film is heated at the interface to its spinodalian temperature (324°C for water), resulting in a liquid to vapor phase transition within microseconds. The vapor bubble expands instantaneously, separating the liquid from the heater, ejecting a droplet through the nozzle hole. The droplet reaches a velocity between 5 and 17 $m\,s^{-1}$ depending on the liquid (mainly its density) and the heater pulse power and duration.

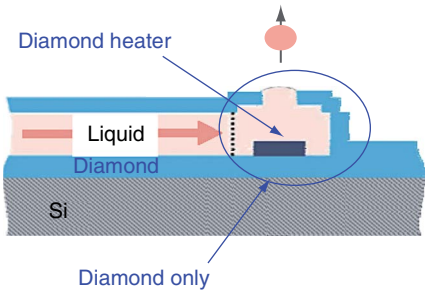


Figure 47 Schematic cross section of all diamond ink-jet. (Source: Mueller R, Gronmaier R, Janischowsky K, Khon E 2005 An 'all-diamond-inkjet' using sacrificial layer technology. *Diamond Relat. Mater.* **14**, 504–8 with permission.)

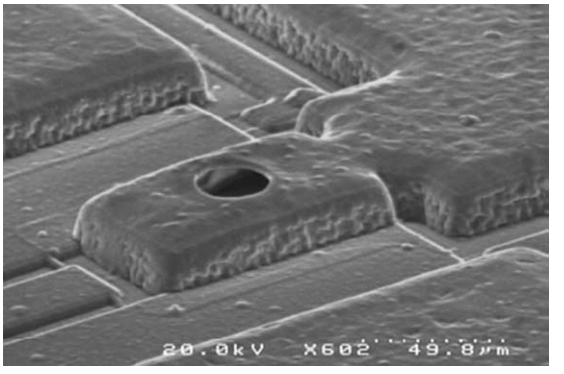


Figure 48 Scanning electron microscope (SEM) micrograph of the all-diamond ink-jet cavity and nozzle opening. (Source: Mueller R, Gronmaier R, Janischowsky K, Khon E 2005 An 'all-diamond-inkjet' using sacrificial layer technology. *Diamond Relat. Mater.* **14**, 504–8 with permission.)

The inertness and chemical stability of the system has been tested dispensing 40% concentrated sulfuric acid to etch Cu patterns into a Cu thin-film sputter deposited onto a glass slide. The glass slide was mounted 10 mm over the ink-jet nozzle plate on a motorized stage. The result can be seen in **Figure 49**, where holes were etched into the Cu film. The droplet velocity was measured to 13 $m\,s^{-1}$ using pseudocinematography, allowing operation at a frequency of 10 kHz.

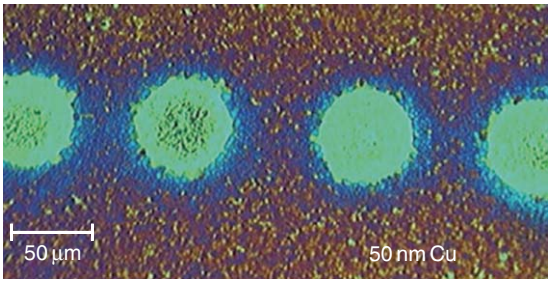


Figure 49 Micrograph of HNO_3 ink-jet drops on 50-nm Cu on glass. (Source: Müller R, Berger W, Janischowsky K, Kusterer J, Kohn E 2005 'All-diamond-inkjet' for dispension of aggressive liquids. *63rd Device Research Conference (DRC)*, Santa Barbara, CA, USA, Abstracts 229–30, June 2005.)

1.06.7 Integration Aspects

Technologically, integration means the interconnection of prefabricated parts into a complex structure or system. Thus, it could be part of the technology of the sensor element itself or part of the sensor or actuator element integration with their readout or driver circuitry. Many aspects are then also relevant for packaging.

Sensors need signal readout by low noise amplifiers, which should be closely integrated. Equally, actuators need drivers directly connected to minimize wiring losses. Candidates are MOSFETs and JFETs on α-SiC (mostly 6H, Friedrichs and Stephani 2006, Slater *et al.*) and β-SiC (Nagasawa *et al.* 2006) and also AlGaN/GaN HEMTs (Daumiller *et al.* 1999). Diamond is not an option yet. While MISiCFETs have been realized in CMOS configuration (although with low hole mobility), GaN-based HEMTs have only shown attractive characteristics for n-type channels. However, up to now, all these configurations have not been used in monolithic integration of sensor elements, despite improvements in the sensor component technology resulting in 1D and 2D arrays like in the case of SiC and diamond particle detector arrays discussed above. Thus, in general, the readout and driver electronics are hybrid integrated. Such technologies are discussed in the following section.

1.06.7.1 Wafer Bonding

Active layer transfer technologies have been developed to connect active layer structures of various materials with a variety of other substrates. It is usually performed in two steps: wafer splitting and wafer bonding. In the first step, the wafer containing the active layer structure is prepared for splitting. This may be obtained by high-dose proton implantation – the wafer is bonded top side down onto the new base substrate, and finally the active surface layer split off from the original wafer.

Implantation with a high dose of protons leads to the formation of planar defects or platelets, which develop into microcracks and microcavities along the peak implant concentration, eventually leading to blister formation at elevated temperatures. The technique has been applied to Si, SiC, GaN, and diamond (Tong *et al.* 1995, 1997). The different bond strength of the materials is reflected in different activation energies for blister formation, which is approximately 1.2 eV for Si and 4.2 eV for diamond. Therefore critical temperatures range from 200°C for Si to 850°C for diamond. In the case of GaN on sapphire, the interface has been laser irradiated with a wavelength absorbed by the GaN but not the sapphire. At high irradiation dose the GaN is dissociated forming N_2 at the interface and the GaN layer is split off (Miskys *et al.* 2003).

Wafer bonding is obtained on hydroxy-terminated surfaces, where the hydroxyl bonds act as bridges to the other surface. The best-known process is the smart-cut process (Aspar *et al.* 1997), where SiC and GaN layer structures have been bonded to silicon base substrates (Cong *et al.* 2004). A similar process has also been developed for the bonding of Si onto diamond (Aleksov *et al.* 2005). Here, the diamond had been grown highly oriented on Si. However, it had been difficult to obtain the high degree of flatness needed across a large surface area.

1.06.7.2 Device Transfer Technology

Hybrid integration will also allow the heterogeneous integration of various types of substrates. One such hybrid technology is the transfer of device structures without substrate onto a base substrate. It involves a lift-off process for the device to be removed from its substrate. Such a process for the transfer of GaN disks, involving a sacrificial layer technology is described in Stones *et al.* (2001).

1.06.7.3 Chip Soldering

In many cases, the sensor chip has to be attached to the base plate by a high-temperature stable soldering or a brazing process. Concerning high-temperature stability, alloys with refractory metals are used, mostly centering around the alloys also used for contact

formation. However, they may be rather brittle, introducing high local stresses. In addition, if small patterns of micrometer size need to be joined together, the kinetics of the formation of the liquid phases is difficult to control, resulting in the formation of rough surfaces and squeeze out. An alternative is solid liquid interdiffusion processes based on isothermal solidification also called SOLID (Hübner *et al.* 2002). In solid–liquid interdiffusion, a low melting metal (i.e., Sn) is placed in between two contacts of a high melting metal (i.e., Cu, Au, and Ni) and heated until the contacts are wetted by the liquid phase of the low melting component. During annealing at a constant temperature, alloying reactions transform the liquid phase into solid intermetallics from both sides. In the Cu/Sn system the bond can be formed slightly above the melting point of tin (232°C), with a solid fast growing η-Cu_6Sn_5 intermetallic phase being transformed into ε-Cu_3Sn, which is stable up to 640°C. If all tin is consumed in the reaction and transformed into the high-temperature stable alloy from both interfaces, the final joint is also stable up to 640°C. The concept is illustrated in **Figure 50**, also showing the phase transformations.

The high-temperature stability of the final joint allows stacking of several chips onto each other without softening the previous joints. In this way, up to six 10 µm thin Si chips have been stacked (Benkart *et al.* 2005). The technology has also been applied to solder a diamond DUV array onto a Si substrate, which could also be an array in SiC or GaN (**Figure 51**). Despite the formation of the joint at approximately 300°C, a high thermal stability of the final joint of approximately 500°C is obtained.

The same technology has also been applied to a bistable diamond switch based on a bistable bridge configuration (**Figure 52**). The actuator is based on the bimetallic concept using a diamond/nickel pair (see Section 1.06.4.2.4) and switched by cycling the temperature. Here, the device is assembled from two parts, the base plate and the actuator. They are also soldered together by the Sn/Cu SOLID process.

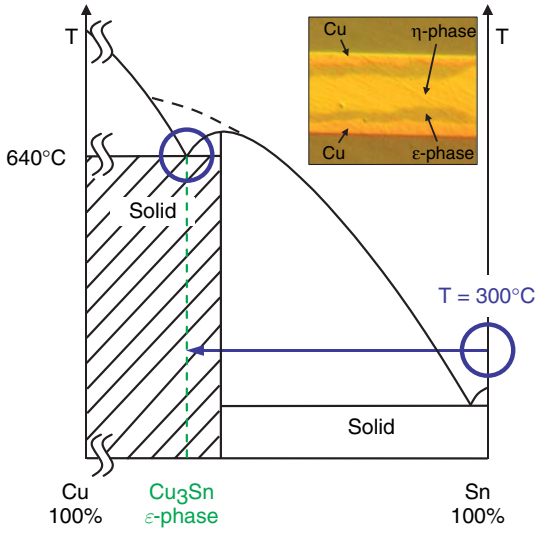


Figure 50 Schematic phase diagram of Cu:Su alloy and indication of liquid solid soldering process (SOLID), see also text. The inset shows schematically the formation of the various phases during alloying. The final ε phase is stable up to 640°C. (Source: Hübner H, Ehrmann O, Eigner M, Gruber W, Klumpp A, Merkel R, Ramm P, Roth M, Weber J, Wieland R 2002 Face-to-face chip integration with full metal interface. *Proc. Advanced Metallization Conference*, *AMC 2002*, Vol. 69, pp. 53–8.)

1.06.7.4 Hermetic Sealing

Devices sensing in a harsh environment have their surface exposed to this environment, however their peripheral structures may need to be hermetically sealed in many cases. Such sealed cavities have been traditionally realized by capping with ceramics (SiC, Al_2O_3, and AlN) or glass–ceramic sandwiches. In a smaller scale this technology has borrowed from the traditional flip-chip bump packaging, where the chip is directly attached to another level of metallization, eliminating wire bond connections. In such a direct chip attach technology, an SiC membrane pressure sensor for an operation at 600°C has been packaged and sealed using dielectrics like AlN, high-temperature melting glass, and Kovar (Okojie *et al.* 2001). In this case, it has been especially important to balance the thermal stress components to eliminate fatigue. Increasingly, micron-sized structures are however sealed with approaches based on 3D thin-film structures with materials like SiO_2, Si_3N_4, and poly-Si (Gogoi and Mastrangelo 2001, Lin *et al.* 1998). The connection between the lid and the substrate may also be a soft metal like thick Au, which can absorb thermal stress (Savrun 2002). Recently, approaches using diamond have also emerged (Zhu and Aslam 2006, Zhu *et al.* 2004) using a molding technique.

1.06.8 Conclusions

In this chapter MEMS materials and devices have been discussed as they have emerged within the last decade. In many cases, the application had been very

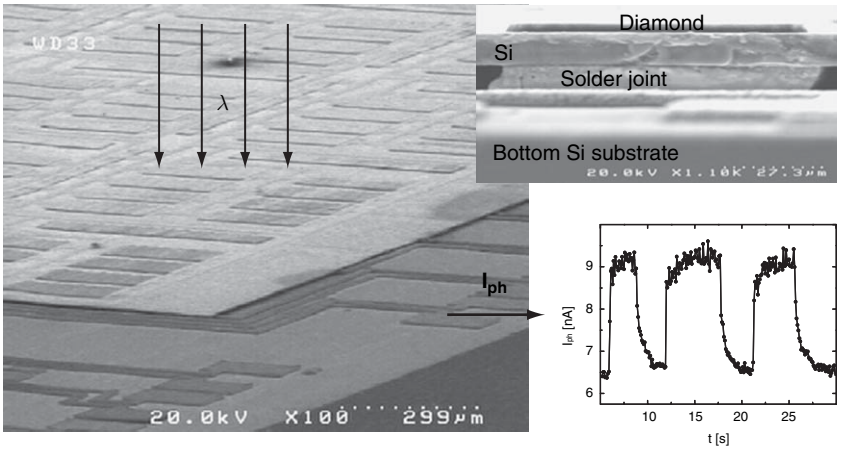


Figure 51 Micrograph of vertically integrated diamond deep ultraviolet (DUV) detector array with Cu:Sn solder joint (soldered as described in **Figure 50**) shown in the inset. Also shown is the photocurrent response. (Source: Kaiser A, Kueck D, Benkart P, Munding A, Prinz G M, Haittmann A, Huebner H, Sauer R, Kohn E 2006 Concept for 3-D intergrated UV sensor. ICNDST-ADC 2006, Research Triangle Park, NC, USA, May 15–18. *Diamond Relat. Mater.*, to be published, with permission.)

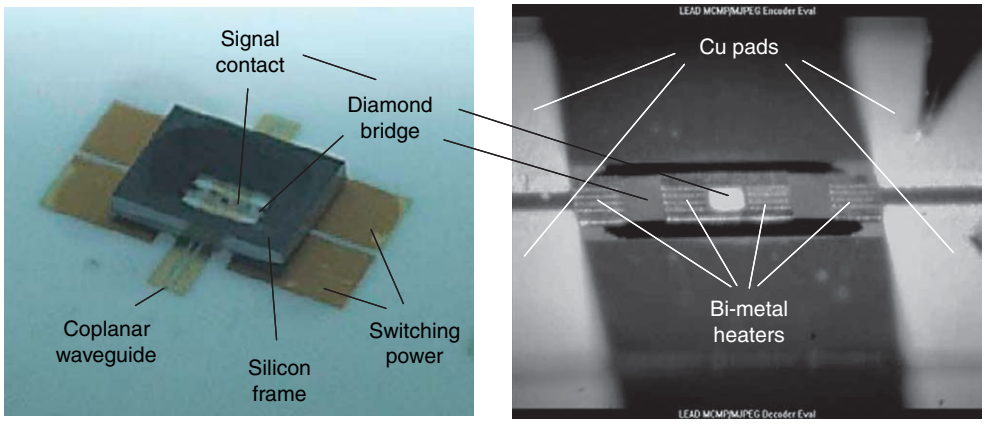


Figure 52 *Left*: Bistable bridge relay with bistable diamond bridge. *Right*: Diamond bridge upside down with signal contact (center, bright) and diamond heaters and Ni bimetal structures. (Source: Kusterer J, Hernandez F J, Guillen K, Janischowsky K, Kusterer J, Ebert W, Kohn E 2006 Bi-stable micro actuator based on stress engineered nano-diamond. *Diamond Relat. Mater.* **15**, 773–6 with permission.)

specific and the driving force had been conditions outside the materials capability of Si. Three materials have emerged: SiC, GaN, and diamond, where GaN represents a materials matrix including InAIN, AlGaN, and InGaN and thus heterostructures. Each of the materials has its own specific advantage and disadvantage and it seems that they can complement each other.

SiC is certainly the most mature material within this group. Most of its ideal properties are second best to ideal diamond. It can be cubic, when grown on Si, and also hexagonal, where it develops a large number of polytypes. Here it is available with a large substrate area, but the stability may be limited by polytype transformation at defects. It sublimes at around 1800°C, and SiC heaters have been operated at 1200°C, the highest temperature of operation of all three materials candidates. It can be doped n-type and p-type and CMOS devices are in essence available for a smart sensor integration in 4H or 6H polytype. The main fields at present are high-temperature piezo-MEMS and high-temperature gas sensors, both oriented toward combustion engine and jet engine control and space applications. It is

the most widely used material for many fields and first MEMS components are already available commercially.

GaN is still a young material and basically confined to optoelectronics and high-power electronics. However, it is the first highly polar semiconductor material. AlN is a traditional electroceramic with a high polarization and high thermal stability, surpassing the depolarization temperature of other piezoelectric materials like PZT essentially. However, single-crystal semiconducting AlN is just emerging and not yet available in large size and high crystalline quality. The high-temperature stability of single-crystal AlN and the AlGaN/GaN and InAlN heterojunctions is very promising despite of their high defect density. If GaN is capped, like in the case of a heterostructure, its thermal stability has been found essentially higher than the GaN decomposition temperature in the atmosphere. Thus, it is one of the few piezoelectric materials with very high depolarizing and decomposition temperatures, but used in MEMS only in proof-of-concept experiments up to now.

Diamond is a material with extreme properties, but burns in oxygen at a high temperature and starts to graphitize at defect sites above approximately 1500°C. In an inert H atmosphere the diamond phase itself has been found stable at 2200°C. However, to obtain large surface areas, deposition onto a foreign substrate is still polycrystalline or nanocrystalline. Certain properties are then determined by the grain boundary characteristics rather than those of diamond. Diamond electronics is not yet available for integration and the film growth conditions are not yet compatible with CMOS processes on Si and SiC. It is therefore still confined to discrete device applications. However, high quality (mostly thick freestanding) films are very attractive for use under extreme conditions like for the detection of high-energy particles in the nuclear environment, in high energy physics, or space. It is also the only inert, highly corrosion-resistant nonmetallic electrode in electrochemistry and therefore increasingly used in poisonous and hazardous environments for analysis as well as for wastewater treatment. In both of the above-mentioned fields, detectors (particles and UV) and electrochemical microprobes, the first diamond arrays have emerged in a first step toward integrated systems.

With respect to electromechanical structures, a similar performance can be obtained in diamond, SiC and Si by scaling the geometry. At present Si-MEMS device structures are scaled down into the nanometer range using single-crystal high-quality material. However, with improved quality of SiC and nanodiamond substrates, the geometries will shrink. A rather important step will be taken, when single crystal large area diamond wafer becomes available. Several groups are investigating single-crystal growth on Ir, upscaling the present substrate size of $1\,cm^2$ to 4″-diameter wafers. This would enable nanometer-sized MEMS in single-crystal diamond, which are presently only investigated by proof-of-concept experiments (Prawer *et al.* 2006).

Acknowledgments

This work contains many contributions of our diamond and GaN-based technology groups over several generations of Ph.D. students. Here I would like to thank in particular Joachim Kusterer for his contribution to the MEMS technology and intensive help with the preparation of the manuscript. A special mention is the contribution of Andrej Denisenko with respect to many fundamental aspects of diamond electrochemistry. With respect to the advanced materials configurations investigated and implemented into device structures, the cooperation with Zlatko Sitar's group at NCSU (AlN) and of Alois Krost at the University of Magdeburg (GaN-on-Si) and Dieter Gruen at Argonne Laboratories (ultra-nanocrystalline-diamond) has to be acknowledged readily. Many thanks also to Werner Haenni (CSEM, Switzerland) for joint experiments and stimulating discussions concerning diamond electrochemistry. Last but not the least, many stimulating discussion within the community should not be left unmentioned. The work cited has been supported financially in part by the European Union, the German Ministry of Education and Research and the German Research Society (DFG).

References

Adam W, Bauer C, Berdermann E, Bergonzo P, Bogani F, Borchi E, Brambilla A, Bruzzi M, Colledani C, Conway J, Dabrowski W, Delpierre P, Deneuville A, Dulinski W, van Eijk B, Fallou A, Fizzotti F, Foulon F, Friedl M, Gan K K, Gheeraert E, Grigoriev E, Hallewell G, Hall-Wilton R, Han S, Hartjes F, Hrubec J, Husson D, Kagan H, Kania D, Kaplon J, Karl C, Kass R, Krammer M, Logiudice A, Lu R, Manfredi P F, Manfredotti C, Marshall R D, Meier D, Mishina M, Oh A, Palmieri V G, Pan L S, Peitz A, Pernicka M, Pirollo S, Polesello P, Pretzl K, Re V, Riester J L, Roe S, Roff D, Rudge A,

Schnetzer S, Sciortino S, Speziali V, Stelzer H, Steuerer J, Stone R, Tapper R J, Tesarek R, Trawick M, Trischuk W, Turchetta R, Vittone E, Wagner A, Walsh A M, Wedenig R, Weilhammer P, Zeuner W, Ziock H, Zoeller M, Charles E, Ciocio A, Dao K, Einsweiler K, Fasching D, Gilchriese M, Joshi A, Kleinfelder S, Milgrome O, Palaio N, Richardson J, Sinervo P, Zizka G 1999 The first bump-bonded pixel detectors on CVD diamond. *Nucl. Instrum. Meth. Phys. Res. A* **436**, 326–35

Adam W, Berdermann E, Bergonzo P, Bertuccio G, Bogani F, Borchi E, Brambilla A, Bruzzi M, Colledani C, Conway J, D'Angelo P, Dabrowski W, Delpierre P, Deneuville A, Dulinski W, van Eijk B, Fallou A, Fizzotti F, Foulon F, Friedl M, Gan K K, Gheeraert E, Hallewell G, Han S, Hartjes F, Hrubec J, Husson D, Kagan H, Kania D, Kaplon J, Kass R, Koeth T, Krammer M, Logiudice A, Lu R, mac Lynne L, Manfredotti C, Meier D, Mishina M, Moroni L, Oh A, Pan L S, Pernicka M, Peitz A, Perera L, Pirollo S, Procario M, Riester J L, Roe S, Rousseau L, Rudge A, Russ J, Sala S, Sampietro M, Schnetzer S, Sciortino S, Stelzer H, Stone R, Suter B, Tapper R J, Tesarek R, Trischuk W, Tromson D, Vittone E, Walsh A M, Wedenig R, Weilhammer P, Wetstein M, White C, Zeuner W, Zoeller M, and RD42 Collaboration 2000 Micro-strip sensors based on CVD diamond. *Nucl. Instrum. Meth. Phys. Res. A* **453**, 141–8

Adamschik M, Schmid P, Ertl S, Ebert W, Käb N, Kohn E 2000a CVD-Diamond for high temperature MEMS applications: First example: Diamond microswitch. *Proc. 5th International High Temperature Electronics Conference (HiTEC),* Albuquerque, NM, USA, June 2000

Adamschik M, Schmid P, Ertl S, Gluche P, Flöter A, Kohn E 2000b Performance of high speed diamond micro switch. *Proc. MicroMat 2000 (3rd Int. Conf. Poster Exhibition Micro Materials),* Berlin, Germany, pp. 863–6

Adamschik M, Müller R, Gluche P, Flöter A, Limmer W, Sauer R, Kohn E 2001 Analysis of piezoresistive properties of CVD-diamond films on silicon. *Diamond Relat. Mater.* **10**, 1670–75

Adamschik M, Kusterer J, Schmid P, Schad K B, Flöter A, Kohn E 2002 Diamond microwave micro relay. *Diamond Relat. Mater.* **11**, 672–6

Aden J S, Bohorquez J H, Collins D M, Crook M D, Garcia A, Hess U E 1994 The third generation HP thermal inkjet printhead. *Hewlett-Packard J.* **45**, 41–5

Aleksov A, Kubovic M, Kasu M, Schmid P, Grobe D, Ertl S, Schreck M, Stritzker B, Kohn E 2004 Diamond-based electronics for RF-applications. *Diamond Relat. Mater.* **13**, 233–40

Aleksov A, Li X, Govindaraju N, Gobien J M, Walter S D, Pater J T, Sitar Z 2005 Silicon-on-diamond: An advanced silicon-on-diamond technology. *Diamond Relat. Mater.* **14**, 308–13

Allers L, Howard A S, Hassard J F, Mainwood A 1997 Neutron damage to CVD diamond. *Diamond Relat. Mater.* **6**, 353–5

Angus J C, Martin H B, Landau U, Evstefeeva Y E, Miller B, Vinokur N 1999 Conducting diamond electrodes: Applications in electrochemistry. *New Diamond Frontier Carbon Technol.* **9**, 175–88

APC 2002 *Piezoelectric Ceramics: Principles and Application.* APC International Ltd

Aspar B, Bruel M, Moriceau H, Maleville C, Poumeyrol T, Papon A M, Claverie A, Benassayag G, Auberton-Herve A J, Barge T 1997 Basic mechanisms involved in the smart-cut process. *Microelectron. Eng.* **36**, 233–40

Auciello O, Birrell J, Carlisle J A, Gerbi J E, Xiao X, Peng B, Espinosa H D 2004 Materials and fabrication processes for a new MEMS technology based on ultrananocrystalline diamond thin films. *J. Phys.: Condens. Matter* **16**, R539–52

Avigal Y, Hoffman A 1999 A new method for nucleation enhancement of diamond. *Diamond Relat. Mater.* **8**, 127–31

Baek S S, Lim H-T, Song H, Kim Y-S, Bae K-D, Cho C-H, Lee C-S, Shin J-W, Shin S-J, Kuk K, Oh Y-S 2004 A novel backshooting inkjet using trench filling and SOI wafer. *Sens Actuators A* **114**, 392–7

Baker C, Laplante J, McNamara K M 2000 Conductive low impurity diamond for solar wind collection. *Diamond Relat. Mater.* **9**, 1951–6

Balducci A, Garino Y, Lo Giudice A, Manfredotti C, Marinelli M, Pulcella G, Verona-Rinati G 2006a Radiological X-ray dosimetry with single crystal CVD diamond detectors. *Diamond Relat. Mater.* **15**, 797–801

Balducci A, Marinelli M, Milani E, Morgada M E, Pucella G, Scoccia M, Tucciarone A, Verona-Rinati G, Angelone M, Pillon M, Potenza R, Tuve C 2006b Growth and characterization of single crystal CVD diamond film based nuclear detectors. *Diamond Relat. Mater.* **15**, 292–5

Basak D, Nakanishi T, Sakai S 2000 Reactive ion etching of GaN using BCl_3, BCl_3/Ar and BCl_3/N_2 gas plasmas. *Solid State Electron.* **44**, 725–8

Bauer T, Gsell S, Schreck M, Goldfuß J, Lettieri J, Schlom G D, Stritzker B 2005 Growth of epitaxial diamond on silicon via iridium/$SrTiO_3$ buffer layers. *Diamond Relat. Mater.* **14**, 314–7

Baur B, Steinhoff G, Heranado J, Purrucker O, Tanaka M, Nickel B, Stutzmann M, Eickhoff M 2005 Chemical functionalization of GaN and AlN surfaces. *Appl. Phys. Lett.* **87**, 263901-1–3

Becker D, Jüttner K 2003 Influence of surface inhomogeneities of boron doped CVD-diamond electrodes on reversible charge transfer reactions. *J. Appl. Electrochem.* **33**, 959–67

Beheim G M, Evans L J 2006 Deep reactive ion etching for bulk micromachining of silicon carbide. *MEMS Handbook: MEMS-Design and Fabrication.* Boca Raton, FL, USA CRC Press, Chap. 8, 8-1 to 8-15

Benkart P, Kaiser A, Munding A, Bschorr M, Pfleiderer H-J, Kohn E 2005 3D chip stack technology using through-chip interconnects. *IEEE Design Test Comput.* **22**, 512–8

BenMoussa A, Hochedez J F, Schühle U, Schmutz W, Haenen K, Stockman Y, Soltani A, Scholze F, Kroth U, Mortet V, Theissen A, Laubis C, Richter M, Koller S, Defise J-M, Koizumi S 2006 Diamond detectors for LYRA, the solar VUV radiometer on board PROBA2. *Diamond Relat. Mater.* **15**, 802–6

Bergonzo P, Jackman R 2004 *Diamond based radiation and photon detectors. Thin Film Diamond II (Semiconductor and Semimetals)*, **77**, Elsevier, Chap. 6, 197–389

Bergonzo P, Brambilla A, Tromson D, Marshall R D, Jany C, Foulon F, Gauthier C, Sole V A, Rogalev A, Goulon J 1999 Semitransparent CVD diamond detectors for *in situ* synchrotron radiation beam monitoring. *Diamond Relat. Mater.* **8**, 920–6

Bergonzo P, Foulon F, Brambilla A, Tromson D, Jany C, Haan S 2000 Corrosion hard CVD diamond alpha particle detectors for nuclear liquid source monitoring. *Diamond Relat. Mater.* **9**, 1003–7

Bergonzo P, Brambilla A, Tromson D, Mer C, Guizard B, Foulon V, Amosov V 2001 CVD diamond for radiation detection devices. *Diamond Relat. Mater.* **10**, 631–8

Bergvogel P 2003 Thirty years of ISFETOLOGY – What happened in the past 30 years and what may happen in the next 30 years. *Sens. Actuators B,* **88**, 1–20

Bertuccio G, Casirahgi R, Cetronio A, Lanszieri C, Nava F 2004 Low-noise silicon carbide X-ray sensor with wide operating temperature range. *Electron. Lett.* **40(3)**, 173–4

Beuret C, Akiyama T, Staufer U, de Rooji N F, Niedermann P, Hänni W 2000 Conical diamond tips realized by a

double-molding process for high-resolution profilometry and atomic force microscopy applications. *Appl. Phys. Lett.* **76**, 1621–23

Bhattacharyya S 2004 Mechanism of high n-type conduction in nitrogen-doped nanocrystalline diamond. *Phys. Rev. B* **70**, 125412-1–10

Bhattacharyya S, Auciello O, Birrell J, Carlisle J A, Curtiss L A, Goyette A N, Gruen D M, Krauss A R, Schlueter J, Sumant A, Zapol P 2001 Synthesis and characterization of highly-conducting nitrogen-doped ultrananocrystalline diamond films. *Appl. Phys. Lett.* **79**, 1441–3

Boehler R 2005 Diamond cells and new materials. *Mater. Today* November, **8**, 34–42

Borst T H, Weis O 1996 Boron-doped homoepitaxial diamond layers: Fabrication, characterization, and electronic applications. *Phys. Stat. Sol. (a)* **154**, 423–44

Brueckner K, Forster Ch, Tonisch K, Cimalla V, Ambacher O, Stephan R, Blau K, Hein M A 2005 Electromechanical resonances of SiC and AlN beams under ambient conditions. *Proc. 2005 European Microwave Conference,* Paris, France, October 2005

Bruzzi M, Bucciolini M, Nava F, Pini S, Russo S 2002 Advanced materials in radiation dosimetry. *Nucl. Instrum. Meth. Phys. Res. A* **485**, 172–7

Bruzzi M, Lagomarsino S, Nava F, Sciotino S 2003a Characterisation of epitaxial SiC Schottky barriers as particle detectors. *Diamond Relat. Mater.* **12**, 1205–8

Bruzzi M, Nava F, Russo S, Sciortino S, Vanni P 2003b Characterisation of silicon carbide detectors response to electron and photon irradiation. *Diamond Relat. Mater.* **10**, 657–61

Buttari D, Heikman S, Keller S, Mishra U K 2002 Digital etching for highly reproducible low damage gate recessing on AlGaN/GaN HEMTS. *Proc. IEEE Lester Eastman Conf. High Performance Devices,* Newark, Delaware, USA, pp. 461–9

Bykhovski A D, Kaminski V V, Shur M S, Chen Q C, Khan M A 1996 Piezoresistive effect in n-type GaN. *Appl. Phys. Lett.* **68**, 818–19

Caliendo C, Imperatori P 2003 High-frequency, high-sensitivity acoustic sensor implemented on AlN/Si substrate. *Appl. Phys. Lett.* **83**, 1641–3

Callaway J 1961 Low-temperature lattice thermal conductivity. *Phys. Rev.* **122**, 787–90

Chang W-H, Schellin B, Obermeier E, Huang Y-C 2006 Electrochemical etching of n-type 6H-SiC without UV illumination. *J. Microelctromech. Syst.* **15**, 548–52

Chen J, Deng S Z, Chen J, Yu Z X, Xu N S 1999 Graphitization of nanodiamond powder annealed in argon ambient. *Appl. Phys. Lett.* **74**, 3651–3

Cheng L, Pan M, Scofield J, Steckl A J 2002 Growth and doping of SiC-thin films on low-stress, amorphous Si_3N_4/Si substrates for robust microelectromechanical systems applications. *J. Electron. Mater.* **31**, 361–5

Chen S W, Lin H F, Sung T T, Wu J D, Kao H L, Chen J S 2003 Synthesis and SAW characteristics of AlN thin films fabricated on Si and GaN using helicon sputtering system. *Electron. Lett.* **39**, 1691–3

Chung S-O, Lee H-C, Lee W-J 2000 Effects of Electrodes on the electric properties of $Pb(Zr,Ti)O_3$ film deposited by electron cyclotron resonance plasma enhanced chemical vapor deposition. *Jpn. J. Appl. Phys.* **39**, 1203–5

Colder H, Rizk R, Morales M, Marie P O, Vicens J 2005 Influence of substrate temperature on growth of nanocrystalline silicon carbide by reactive magnetron sputtering. *J. Appl. Phys.* **98**, 024313-1–10

Colinge J-P 1998 Fully-depleted SOI CMOS for analog applications. *IEEE Trans. Electron Devices* **45**, 1010–6

Colley A L, Williams C G, Johnsson U D H, Newton M E, Unwin P R, Wilson N R J, Macpherson V 2006 Examination of the special heterogeneous electroactivity of boron-doped diamond microarray electrodes. *Anal. Chem.* **78**, 2539–48

Collins A T 1989 Diamond electronic devices – A critical appraisal. *Semicond. Sci. Technol.* **4**, 605–11

Cong P, Ko W H, Young D J 2004 Single crystal SiC MEMS fabrication technology using smart-cut process for harsh environment application. *Proc. IEEE Sensors 2004, Vienna, Austria*, pp. 1161–4

Dadgar A, Poschenrieder M, Daumiller I, Kunze M, Strittmatter A, Riemann T, Bertram F, Bläsing J, Schulze F, Reiher A, Krtschil A, Contreras O, Kaluza A, Modlich A, Kamp M, Reißmann L, Diez A, Christen J, Ponce F A, Bimberg D, Kohn E, Krost A 2002 Gallium-nitride-based devices on silicon. *Phys. Stat. Sol.(c)* **0**, 1940–9

Daumiller I, Kirchner C, Kamp M, Ebeling K J, Kohn E 1999 Evaluation of the temperature stability of AlGaN/GaN heterostructure FETs. *IEEE Electron Devices Lett.* **20**, 448–50

Davidson D, Kang W P, Wisitsora At 2003 Diamond field emission devices. *Diamond Relat. Mater.* **12**, 429–33

Davies S, Huang T S, Gass M H, Papworth A J, Joyce T B, Chalker P R 2004 Fabrication of GaN cantilevers on silicon substrates for microelectromechanical device. *Appl. Phys. Lett.* **84**, 2566–8

Denisenko A, Jamornmarn G, El-Hajj H, Kohn E 2007 pH Sensor on O-terminated diamond using boron doped channel. *Diamond and Related Materials*, **16**, 905–10

Denisenko A, Pietzka C, Kohn E, Kunze M, Daumiller I, Dadgar A, Krost A 2006 Effect of anodic polarization on electrochemical properties of GaN surface. *E-MRS 2006 Conference,* Nice, France Symposium Q, Presentation Q17, May/June 2006

Ding G F, Mao H P, Zhang Y H, Yao X, Zhao X L 2005 Micromachining of CVD diamond by RIE for MEMS applications. *Diamond Relat. Mater.* **14**, 1543–8

Dulloo A R, Ruddy F H, Seidel J G, Adams J M, Nico J S, Gilliam D M 1999 The neutron response of miniature silicon carbide semiconductor detectors. *Nucl. Instrum. Meth. Phys. Res. A* **422**, 47–8

Eickhoff M, Möller H, Stoemenos J, Zappe S, Kroetz G, Stutzmann M 2004 Influence of crystal quality on the electronic properties of n-type 3C-SiC grown by low temperature low pressure chemical vapor deposition. *J. Phys.* **95**, 7908–17

Elmazria O, Assouar M B, Renard P, Alnot P 2003 Electrical properties of piezoelectric aluminium nitride films deposited by reactive dc magnetron sputtering. *Phys. Stat. Sol. (a)* **196**, 416–21

Ertl S, Adamschik M, Schmid P, Gluche P, Flöter A, Kohn E 2000 Surface micromachined diamond microswitch. *J. Diamond Relat. Mater.* **9**, 970–4

Espinosa H D, Peng B, Prorok B C, Moldovan N, Auciello O, Carlisle J A, Gruen D M, Mancini D C 2003 Fracture strength of ultrananocrystalline diamond thin films – Identification of Weibull parameters. *J. Appl. Phys.* **94**, 6076–84

Feltin E, Beaumont B, Laügt M, De Mierri P, Vennegues P, Leroux M, Gibart P 2001 Crack-free thick GaN layers on silicon (111) by metalorganic vapor phase epitaxy. *Phys. Stat. Sol.(a)* **188**, 531–5

Fleischer M, Meixner H 1992 Sensing reducing gases at high temperatures using long term stable Ga_2O_3 thin films. *Sens. Actuators B* **6**, 257–61

Freedman A, Stinespring C D 1990 Fluorination of diamond (100) by atomic and molecular beams. *Appl. Phys. Lett.* **57**, 1194–6

Friedrichs P, Stephani D 2006 Unipolar SiC power devices and elevated temperature. *Microelectron. Eng.* **83**, 181–4

Fu X, Dunning J L, Zorman C A, Mehregany M 2005 Measurement of residual stress and elastic modulus of polycrystalline 3C-SiC films deposited by low-pressure chemical vapour deposition. *Thin Solid Films* **492**, 195–202

Garrido J A, Härtl A, Kuch S, Stutzmann M, Williams O A, Jackmann R B 2005 pH sensors based on hydrogenated diamond surfaces. *Appl. Phys. Lett.* **86**, 073504–1

Gaska R, Yang J W, Bykhovski A D, Shur M S, Kaminski V V, Soloviov S 1997 piozoresistive effect in GaN–AlN–GaN structures. *Appl. Phys. Lett.* **71**, 3817–19

Gaska R, Shur M S, Bykhovski A D, Yang J W, Khan M A, Kaminski V V, Soloviov S 2000 Piezoresistive effect in metal–semiconductor–metal structures on p-type GaN. *Appl. Phys. Lett.* **76**, 3956–8

Gaska R, Chen C, Yang J, Kuokstis E, Khan A, Tamulaitis G, Yilmaz I, Shur M S, Rojo J C, Schowalter L J 2002 Deep-ultraviolet emission of AlGaN/GaN quantum wells on bulk AlN. *Appl. Phys. Lett* **81**, 4658–60

Georgakilas A, Mikroulis S, Cimalla V, Zervos M, Kostopoulos A, Komninou Ph, Kehagias Th, Karakostas Th 2001 Effects of the sapphire nitridation on the polarity and structural properties of GaN layers grown by plasma-assisted MBE. *Phys. Stat. Sol. (a)* **188**, 567–70

Glassbrenner C J, Slack G A 1964 Thermal conductivity of silicon and germanium from 3 K to the melting point. *Phys. Rev.* **134**, A1058–69

Gluche P, Leuner R, Vescan A, Ebert W, Kohn E, Rembe C, aus der Wiesche S, Hofer E P 1998 Actuator–sensor technology on electronic grade diamond films. *Microsyst. Technol.* **5**, 38–43

Gogoi B P, Mastrangelo C H 2001 Force balanced micromachined pressure sensors. *IEEE Trans. Electron Devices* **48**, 1575–84

Gu Z, Edgar J H, Speakman S A, Blom D, Perrin J, Chaudhuri J 2005 Thermal oxidation of polycrystalline and single crystalline aluminium nitride wafers. *J. Electron. Mater.* **34**, 1271–9

Gurbuz Y, Kang W P, Davidson J L, Kern D V 1998 High temperature tolerant diamond-based microelectronic oxygen gas sensor. *Sens Actuators B* **49**, 115–20

Gurbuz Y, Kang W P, Davidson J L, Kerns D V 2004 Diamond miroelectronic gas sensor for detection of benzene and toluene. *Sens Actuators B* **99**, 207–15

Haenni W, Rychen P, Fryda M, Comninellis C 2004 Industrial applications of diamond electrodes. *Thin Film Diamond II (Semiconductor and Semimetals).* Elsevier, Chap. 5, 149–96

Härtl A, Schmich E, Garrido J A, Hernando J, Catharino S C R, Walter S, Feulner P, Kromka A, Steinmüller D, Stutzmann M 2004 Protein-modified nanocrystalline diamond thin films for biosensor applications. *Nat. Mater.* **3**, 736–42

Hernandez-Guillen F J, Janischowsky K, Kusterer J, Ebert W, Kohn E 2005 Mechanical characterization and stress engineering of nanocrystalline diamond films for MEMS applications. *Diamond Relat. Mater.* **14**, 411–15

Hochedez J-F, Alvarez J, Auret F D, Bergonzo P, Castex M-C, Deneuville A, Defise J M, Fleck B, Gibart P, Goodman S A, Hainaut O, Kleider J-P, Lemaire P, Manca J, Monroy E, Muñoz E, Muret P, Nesladek M, Omnes F, Pace E, Pau J L, Ralchenko V, Roggen J, Schühle U, Van Hoof C 2002 Recent progress of the BOLD investigation towards UV detectors for the ESA solar orbiter. *Diamond Relat. Mater.* **11**, 427–32

Hossain T K, McLaren S, Engel J M, Lui C, Adesida I, Okojie R S 2006 The fabrication of suspended micromechanical structures from bulk 6H–SiC using an ICP-RIE system. *J. Micromech. Microeng.* **16**, 751–6

Hu X, Deng J, Pala N, Gaska R, Shur M S, Chen C Q, Yang J, Simin G, Khan M A, Rojo J C, Schowalter L J 2003 AlGaN/GaN heterostructure field-effect transistors on single-crystal bulk AlN. *Appl. Phys. Lett.* **82**, 1299–301

Hübner H, Ehrmann O, Eigner M, Gruber W, Klumpp A, Merkel R, Ramm P, Roth M, Weber J, Wieland R 2002 Face-to-face chip integration with full metal interface. *Proc. Advanced Metallization Conference, AMC 2002*, San Diego, CA, USA, MRS Proceedings, V18, pp. 53–8

Huygens I M, Strubbe K, Gomes W P 2000 Electrochemistry and photoetching of GaN. *J. Electrochem. Soc.* **147**, 1797–1802

Ibbetson J P, Fini P T, Ness K D, DenBaars S P, Speck J S, Mishra U K 2000 Polarization effects, surface states, and the source of electrons in AlGaN/GaN heterostructure field effect transistors. *Appl. Phys. Lett.* **77**, 250–2

Internet libraries: www.webelements.com, www.memsnet.org/material/environmentalchemistry.com/Yogi/periodic/, www.ioffe.rssi.ru/SVA/NSM/Semicond

Isberg J, Hammersberg J, Johansson E, Wikstroem T, Twitschen D J, Whithead A J, Scarsbrook S E G A 2002 High carrier mobility in single-crystal plasma-deposited diamond. *Science* **297**, 1670–2

Janischowsky K, Ebert W, Kohn E 2003 Bias enhanced nucleation of diamond on Si (100) in a HFCVD system. *Diamond Relat. Mater.* **12**, 336–9

Jena D, Heikman S, Green D, Buttari D, Coffie R, Xing H, Keller S, Denbaars S, Speck J S, Smorchkova I, Mishra U K 2002 Realization of wide electron slabs by polarization doping in graded III–V nitride semiconductor alloys. *Appl. Phys. Lett.* **81**, 4395–7

Jian S-R, Fang T-H, Chuu D-S 2003 Analysis of physical properties of III-nitrides thin films by nanoindentation. *J. Electron. Mater.* **32**, 496–500

Jiang X, Paul M, Klages C P 1996 Deposition of heteroepitaxial diamond films on 2 in silicon substrates. *Diamond Relat. Mater.* **5**, 251–5

Jiang L, Cheung R, Hedley J, Hassan M, Harris A J, Burdess J S, Mehregany M, Zorman C A 2006 SiC cantilever resonators with electrothermal actuation. *Sens. Actuators A* **128**, 376–86

von Kaenel Y, Stiegler J, Mischler J E, Blank E 1997 Stress distribution in heteroepitaxial chemical vapor deposited diamond films. *J. Appl. Phys.* **81**, 1726–36

Kaiser A, Hintz M, Adamschik M, Schmid P, Müller R, Maier C, Brugger H, Hofer E P, Seliger H, Kohn E 2002 Diamond based injection system for spotting and synthesis in biochemistry. *AICHE Annu. Meeting, Symp. Sensor Technology,* Indianapolis, IN, USA, Proceedings 175–85, November 2002

Kaiser A, Kueck D, Benkart P, Munding A, Prinz G M, Haittmann A, Huebner H, Sauer R, Kohn E 2006 Concept for 3-D integrated UV sensor. *Diamond Relat. Mater.,* **15**, 1967–71

Kandasamy S, Trinchi A, Wlodarski W, Comini E, Sberveglieri G 2005 Hydrogen and hydrocarbon gas sensing performance of Pt/WO_3/SiC MROSiC devices. *Sens. Actuators B* **111–12**, 111–16

Kang B S, Kim S, Ren F, Baik K, Pearton S J, Gila B P, Abernathy C R, Pan C-C, Chen G-T, Chyi J-I, Chandrasekaran V, Sheplak A, Nishida T, Chu S N G 2003 Effect of external strain on the conductivity of AlGaN/GaN high-electron-mobility transistors. *Appl. Phys. Lett.* **83**, 4845–7

Kang B S, Kim S, Ren F, Johnson J W, Terrien R J, Rajagopal R J, Roberts J C, Piner E L, Linthicum K J, Chu S N G, Baik K, Gila B P, Abernathy C R, Pearton S J 2004 Pressure-induced changes in the conductivity of AlGaN/GaN high-electron mobility-transistor membranes. *Appl. Phys. Lett.* **85**, 2962–4

Kang B S, Kim J, Jang S, Ren F, Johnson J W, Therrien R J, Rajagopal P, Roberts J C, Piner E L, Linthicum K J, Chu S N G, Baik K, Gila B P, Abernathy C R, Pearton S J 2005 Capacitance pressure sensor based on GaN high-electron-mobility transistor-on-Si membrane. *Appl. Phys. Lett.* **86**, 253502-1–3

Kato M, Ishimura M, Arai E, Ramasamy P 2003 Electrochemical etching of 6H-SiC using aqueous KOH solutions with surface roughness. *Jpn. J. Appl. Phys.* **42**, 4233–6

Kawarada H 1998 Heteroepitaxy and highly oriented diamond deposition. *Low-PressureSynthetic Diamond – Manufacturing and Applications.* Springer, Chap. 8, 139–62

Kawarada H, Araki Y, Sakai T, Ogawa T, Umezawa H 2001 Electrolyte-solution-gate FETs using diamond surface for biocompatible ion sensors. *Phys. Stat. Sol. (a)* **185**, 79–83

Kim M-H, Do Y-G, Kang H C, Noh D Y, Park S-J 2001 Effect of step-graded $Al_xGa_{1-x}N$ interlayer on properties of GaN grown on Si(111) using ultrahigh vacuum chemical vapour deposition. *Appl. Phys. Lett.* **79**, 2713–5

Kim J, Ren F, Gila B P, Abernathy C R, Pearton S J 2003 Reversibe barrier height changes in hydrogen-sensitive Pd/GaN and Pt/GaN diodes. *Appl. Phys. Lett.* **82**, 739–41

Klaasse G, Puers B, Tilmans H A C 2004 Piezoelectric actuation for application in RF-MEMS switches; MEMS, MOEMS and Micromachining, *Proc. SPIE* **5455**, 174–80

Klages C-P, Schäfer L 1998 Hot-filament deposition of diamond. *Low-Pressure Synthetic Diamond – Manufacturing and Applications.* Springer, Chap. 5, 85–101

Kohn E, Adamschik M, Schmid P, Ertl S, Flöter A 2001 Diamond electro-mechanical micro devices-technology and performance. *Diamond Relat. Mater.* **10**, 1684–91

Kohn E, Gluche P, Adamschik M 1999 Diamond MEMS – A new emerging technology. *Diamond Relat. Mater.* **8**, 934–40

Kohn E, Denisenko A, Ive T, Brand O, Dadgar A, Krost A, Haenni W 2005 Liquid junctions to wide bandgap semiconductors. *13th Int. Workshop Physics of Semiconductor Devices (IWPSD-2005)*, New Delhi, India, The Physics of Semiconductor Devices 2005, December 2005, pp. 750–9

Konda T, Honda K, Tyrk D A, Fujishima A 2005 Covalent modification of single-crystal diamond. *J. Electrochem. Soc.* **152**, E18–23

Krause P, Obermeier E, Wehl W 1996 A micromachined single-chip injet printhead. *Sensors Actuators A* **53**, 405–9

Krost A, Dadgar A 2002 GaN based devices on Si. *Phys. Stat. Sol. (a)* **194**, 361–75

Kuizumi S 2003 *Thin-Film Diamond I (Semiconductors and Semimetals)* Elsevier, Chap. 5, 239–59

Kulisch W, Malave A, Lippold G, Scholz W, Mihalcea C, Oesterschulze E 1997 Fabrication of integrated diamond cantilevers with tips for SPM applications. *Diamond Relat. Mater.* **6**, 906–11

Kumagai Y, Murakami H, Seki H, Koukitu A 2002 Thick and high-quality GaN growth on GaAs (111) substrates for preparation of free standing GaN. *J. Crystal Growth* **246**, 215–222

Kumar S, Ravindranathan P, Dewan H S, Roy R 1996 Survival of diamond at 2200°C in hydrogen. *Diamond Relat. Mater.* **5**, 1246–8

Kunze M, Vescan A, Gluche P, Ebert W, Kohn E 1999 Delta-doping in diamond. *Carbon* **37**, 787–791

Kusterer J, Alekov A, Pasquarelli A, Müller R, Ebert W, Lehmann-Horn F, Kohn E 2005 A diamond-on-Si patch-clamp-system. *Diamond Relat. Mater.* **14**, 2139–42

Kusterer J, Hernandez F J, Haroon S, Schmid P, Mürding A, Müller R, Kohn E 2006 Bi-stable micro actuator based on stress engineered nano-diamond. *Diamond Relat. Mater.* **15**, 773–6

Lämmer F A, Schlip A Method of anisotropically etching silicon. *US Pat.* 5501893 (1996), *German Pat.* DE4241045 (1994)

Lardon P, Mer C, Delacour P, Bazin B, Tromson D, Normand S, Nesladek M, Foulon F, Bergonzo P 2006 Investigations of high quality diamond detectors for neutron fluency monitoring in a nuclear reactor. *Diamond Relat. Mater.* **15**, 815–21

Lauermann I, Memming R, Meissner D 1997 Electrochemical properties of silicon carbide. *J. Electrochem. Soc.* **144**, 73–80

Lee K-W, Yu K-S, Boo J-H, Kim Y, Hatayama T, Kimoto T, Matsunami H 1997 Epitaxial growth of cubic SiC films on Si substrates of by high vacuum chemical vapour deposition using 1,3-disilabutance. *J. Electrochem. Soc.* **144**, 1474–6

Lee S-W, Kim H-C, Kuk K, Oh Y-S 2002 A monolithic inkjet print head: Domejet. *Sens Actuators A* **95**, 114–19

Lee S M, Lee H S, Lee H H 2006 Development of hybrid real time wide-band electronic radiation dosimeter for a mobile robot. *Proc. SAS 2006 – IEEE Sensors Applications Symposium*, Houston, TX, USA, February

Li D, Sumiya M, Fuke S, Yang D, Que D, Suzuki Y, Fukuda Y 2001 Selective etching of GaN polar surface in potassium hydroxide solution studied by x-ray photoelectron spectroscopy. *J. Appl. Phys.* **90**, 4219–23

Liaw H M F S 1994 SAW characteristics of sputtered aluminium nitride on Si and gallium arsenide. *Proc. IEEE Ultrasonics Symposium 1994,* Frroelectric and Frequency Control Society, Cannes, France , November 1994

Lim D-C, Lee H-G, Kim J W, Moon J-S, Lee S-B, Choi S S, Boo J-H 2004 Deposition of epitaxial silicon carbide films using high vacuum MOCVD method for MEMS application. *Thin Solid Films* **459**, 7–12

Lin L, Howe R T, Pisano A P 1998 Microelectromechanical filters for signal processing. *J. Microelectromech. Syst.* **7**, 286–93

Mainwood A 1998 CVD diamond particle detectors. *Diamond Relat. Mater.* **7**, 504–9

Marczewska B, Bilski P, Olko P, Nesladeck M, Bergonzo P, Rebisz M, Waligorski M P R 2003 CVD diamond wafers as large-area thermoluminescence detectors for measuring the special distribution of dose. *Phys. Stat. Sol. (a)* **199**, 119–24

McGoech S P, Placido F, Gou Z, Wort C J H, Savage J A 1999 Coatings for the protection of diamond in high-temperature environments. *Diamond Relat. Mater.* **8**, 916–9

Medjdoub F, Carlin J-F, Gonschorek M, Feltin E, Py M A, Ducatteau D, Gaquiere C, Grandjean N E 2006 Can InA1N/GaN be an alternative to high power/high temperature AlGaN/GaN devices? *2006 IEEE International Electron Devices Meeting,* San Francisco, CA, USA, Technical Digest, December 11–13, 2006

Mehregany M, Zorman C A 1999 SiC MEMS: Opportunities and challenges for applications in harsh environments. *Thin Solid Films* **355–6**, 518–24

Mer C, Pomorski M, Bergonzo P, Tromson D, Rebisz M, Domenech T, Vuillemin J-C, Foulon F, Nesladek M, Williams O A, Jackman R B 2004 An insight into neutron detection from polycrystalline CVD diamond films. *Diamond Relat. Mater.* **13**, 791–5

Mikami H, Tatayama T, Yo H, Uraoka Y, Fuyuki T 2005 Analysis of photochemical processes in α-SiC substrates with automatically flat surfaces. *Jpn. J. Appl. Phys.* **44**, 8329–32

Miskys C R, Kelly M Y, Ambacher A, Stutzmann M 2003 Freestanding GaN-substrates and devices. *Phys Stat. Sol. (c)* **0**, No 6, 1627–50

Mlcak R, Doppalapudi D, Whitfield G, Tuller H 2004 Single crystal SiC microhotplate conductometric chemical sensor array. *206th Meeting of the Electrochemical Society*, Honolulu, HI, USA, Abstract 2493

Morales F M, Molina S I, Araujo D, Garcia R, Cimalla V, Pezoldt J 2003 SiC voids, mosaic microstructure and dislocation distribution in Si carbonized layers. *Diamond Relat. Mater.* **12**, 1227–30

Mueller R, Gronmaier R, Janischowsky K, Kohn E 2005 An 'all-diamond-inkjet' using sacrificial layer technology. *Diamond Relat. Mater.* **14**, 504–8

Müller R, Denisenko A, Kohn E 2003 Effect of surface quality on the ion sensitivity of H-terminated diamond. *Diamond Relat. Mater.* **12**, 554–559

Müller R, Schmid P, Munding A, Gronmaier R, Kohn E 2004 Elements of surface microfluidics in diamond. *Diamond Relat. Mater.* **13**, 780–4

Müller R, Berger W, Janischowsky K, Kusterer J, Kohn E 2005 All-diamond-inkjet' for dispension of aggressive liquids. *63rd Device Research Conference (DRC),* Santa Barbara, CA, USA, Abstracts 229–30, June 2005

Murakawa M, Noguchi H, Takeuchi S 2003 Fabrication of 3-D shaped micro body structures of diamond by use of focused ion beam. *7th Applied Diamond/3rd Frontier AFC 2003)/AFC 2003), Proceedings (NASA/CP-2003-212319),* August 2003, Tsukuba, Japan, pp. 185–90

Nagai T, Yamamoto K, Kobayashi I 1982 SiC thin-film thermistor. *J. Phys. E: Sci. Instrum.* **15**, 520–4

Nagasawa H, Yagi K, Kawahara T, Hatta N, Abe M 2006 Hetero- and homo-epitaxial growth of 3C-siC of MOS-FETs. *Microelectron. Eng.* **83**, 185–8

Neuburger M, Zimmermann T, Benkart P, Kunze M, Daumiller I, Dadgar A, Krost A, Kohn E 2004 GaN based piezo sensor. *62nd Device Research Conference (DRC),* South Bend, IN, USA, Abstracts 45–6, June 2004

Neuburger M, Aleksov A, Schlesser R, Kohn E, Sitar Z 2007 *Electr. Lett.* **43**, 592–4

Nichols B M, Butler J E, Russell J N, Hamers R J 2005 Photochemical functionalization of hydrogen-terminated diamond surfaces: A structural and mechanistic study. *J. Phys. Chem. B* **109**, 20938–47

Ohtani B, Kim Y, Yano T, Hashimoto K, Fujishima A, Uosaki K 1998 Surface functionalization of doped CVD diamond via covalent bond. An XPS study on the formation of surface-bond quaternary pyridinium salt. *Chem. Lett.* 953–4

Okada Y, Tokumaru Y 1984 Precise determination of lattice parameters and thermal expansion coefficient of silicon between 300 and 1500 K. *J. Appl. Phys.* **56**, 314–20

Okojie R S 2006 Single-crystal silicon carbide MEMS: Fabrication, characterization and reliability. *MEMS Handbook: MEMS-Design and Fabrication,* 2nd edn. CRC Press, Chap. 7, Taylor and Francis, CRC Press, Boca Raton, FL, USA

Okojie R S, Ned A A, Kurtz A D 1998a Operation of α(6H)–SiC pressure sensor at 500°C. *Sens Actuators A* **66**, 200–4

Okojie R S, Ned A A, Kurtz A D, Carr W N 1998b Characterization of highly doped n- and p-type 6H–SiC piezoresistors. *IEEE Trans. Electron Devices* **45**, 785–90

Okojie R S, Benheim G M, Saad G J, Savrun E 2001 Characteristics of a hermetic 6H–SiC pressure sensor at 600°C. *AIAA Space 2001 Conference and Exposition,* Albuquerque, NM, USA, Abstract AIAA 2001-4652 August 2001, Abstract AIAA 2001-4652 August 2001, pp. 1–8

Olivares J, Iborra E, Clement M, Vergara L, Sangrador J, Sanz-Hervas A 2005 Piezoelectric actuation of microbridges using AlN. *Sens. Actuators A* **123–4**, 590–5

Palmour J W, Kong H S, Waltz D G, Edmond J A, Carter C H Jr. 19916H–silicon carbide transistors for high temperature operation. Proc 1st Int. High Temperature Electronic Conference, p. 511

Pearton S J, Abernathy C R, Ren F 1994 Low bias electron cyclotron resonance plasma etching of GaN, AlN and InN. *Appl. Phys. Lett.* **64**, 2294–5

Perez I, Torvik J B 2003 4H–SiC Rf bipolar junction transistors. *61st, Device Research Conference,* Salt Lake City, UT, USA, Abstracts 27

Ping A P, Schmitz A C, Khan M A, Adesida I 1996 Dry etching of GaN using chemically assisted ion beam etching with HCl and H_2/Cl_2. *J. Electron. Mater.* **25**, 825–9

Pisch A, Schmid-Fetzer R 1998 *In situ* decomposition study of GaN thin films. *J. Crystal Growth* **187**, 329–32

Powell A R, Rowland L B 2002 SiC materials – Processes, status, and potential road blocks. *Proc. IEEE* **90**, 942–55

Prawer S, Namieson D N, Huntington S T, Greentree A D, Rabeau J, Oliviero P, Gibson B, Reichart P, Salzman J, Jelezko F 2006 Diamond based quantum information processing. *ICNDST and ADC 2006*. Research Triangle Park, NC, USA , Abstract, May 2006

Ralchenko V G, Givargizo E I 2003 Fabrication of very-high aspect-ratio micro channels in CVD diamond by a molding technique. *7th Applied Diamond/3rd Frontier Carbon Technology Joint Conference (ADC/AFC 2003), Tsubuka, Japan Proceedings (NASA/CP-2003-212319),* August 2003, pp. 540–3

Rangsten P, Björkman H, Hjort K 1999 Microfluidic components in diamond. *Proc.Tranducers 99,* June 1999 Sendai, Japan, June 1999, pp. 190–3

Rao T N, Fujishima A 2000 Recent advances in electrochemistry of diamond. *Diamond Relat. Mater.* **9**, 384–9

Rebisz M, Martemiyanov A, Berdermann E, Pomorski M, Marczewska B, Voss B 2006 Synthetic diamonds for heavy-ion therapy dosimetry. *Diamond Relat. Mater.* **15**, 822–6

Rotter S 1999 Growth of conformal thin diamond films by CVD for coating application. *Proc.Appl. Diamond Conf./Frontier Carbon Technologies–ADC/FCT 99*, Tokyo, Japan, pp. 25–7; publ. by i-cube, Tsukuba, Japan

Ruvimov S, Lilienthal-Weber Z, Washburn J, Duxstad K J, Haller E E, Fan Z-F, Mohammad S N, Kim W, Botchkarev A E, Morkoc H 1996 Micostructure of Ti/Al and Ti/Al/Ni/ Au ohmic contacts for n-GaN. *Appl. Phys. Lett.* **69**, 1556–8

Savrun E 2002 Packaging considerations for very high temperature microsystems. *IEEE Sensors 2002*, Orlando, FL, USA, pp. 1139–43

Saxena V, Su J N, Steckl A J 1999 High-voltage Ni– and Pt–SiC Schottky diodes utilizing metal field plate termination. *IEEE Trans. Electron Devices* **46**, 456–63

Schalwig J, Müller G, Ambacher O, Stutzmann M 2001 Group-III-nitirde based gas sensing devices. *Phys. Stat. Sol. (a)* **185**, 39–45

Schmid P 2006 Ph.D. thesis, University of Ulm

Schmid P, Lipka K M, Ibbetson J, Nguyen N, Mishra U, Pond L, Weitzel C, Kohn E 1998 High-temperature performance of GaAs-based HFET structure containing LT-AlGaAs and LT-AlGaAs. *IEEE Electron Devices Lett.* **19**, 225–7

Schimid P, Hernandes-Gullien F J, Kohn E 2003 Diamond switch using new thermal actuation principle. *Diamond Relat. Mater.* **12**, 418–21

Schmitz A C, Ping A T, Khan M A, Chen Q, Yang J W, Adesida I 1998 Metal contacts to n-type GaN. *J. Electron. Mater.* **27**, 255–260

Sekaric L, Parpia J M, Craighead H G, Feygelson T, Houston B H, Butler J E 2002 Nanomechanical resonant structures in nanocrystalline diamond. *Appl. Phys. Lett.* **81**, 4455–4457

Sellin P J, Vaitkus J 2000 New materials for radiation hard semiconductor detectors. *Nucl. Instrum. Meth. Phys. Res. A* **557**, 479–89

Serre C, Perez-Rodriguez A, Romano-Rodrigurez A, Calvo-Barrio L, Morante J R, Esteve J, Acero M C, Skorupo W, Kögler R 1997 Synthesis of SiC microstructures in Si technology by high dose carbon implantation: Etch stop properties. *J. Electrochem. Soc.* **44**, 2211–2215

Sevillano E 1998 Microwave plasma deposition of diamond. *Low-Pressure Synthetic Diamond – Manufacturing and Applications.* Springer, Chap. 2, pp. 11–39

Sharda T, Bhattacharya S Diamond nanocrystals. Encyclopedia of Nanoscience and Nanotechnology (ENN). American Scientific Publishers (ASP)/Vol.2, pp. 337–70

Sheldon B W, Csencsits R, Rankin J, Boekenhauer R E, Shigesato Y 1994 Bias enhanced nucleation of diamond during microwave-assisted chemical vapour deposition. *J. Appl. Phys.* **75**, 5001–8

Shikata S H 2004 Diamond surface acoustic wave devices. *Thin Film Diamond II (Semiconductor and Semimetals)*, ElsevierVol. 77,, Chap. 8, pp. 339–358

Shu D, Job P K, Kuzay T M, Korenov S 2001 CVD-diamond-based position-sensitive detector test with electron beam from a Rhodotron accelerator. *Proc. 2001 Particle Accelerator Conference*, Chicago, IL, USA, pp. 2435–7

Shul R J, McClellan G B, Pearton S J, Abernathy C R, Constantine C, Barratt C 1996 Comparison of dry etch techniques for GaN. *Electron. Lett.* **32**, 1408–9

Siechel E K, Pankove J I 1975 Thermal conductivity of GaN. *J. Chem. Solids* **38**, 330

Sitar Z, Private communication

Slack G A 1964 Thermal conductivity of pure and impure silicon, silicon carbide, and diamond. *J. Appl. Phys.* **38**, 3460–6

Slack G A, Bartram S F 1975 Thermal expansion of some diamond like cystals. *J. Appl. Phys.* **46**, 89–98

Slack G A, McNelly T F 1976 Growth of high purity AlN crystals. *J. Crystal Growth* **34**, 263–79

Slack G A, Tanzilli R A, Pohl R O, Vandersande J W 1987 The intrinsic thermal conductivity of A1N. *J. Chem. Solids* **48**, 641–47

Slater D B, Johnson G M, Lipkin L A, Suvorov A V, Palmour J P Demonstration of 6H-SiC CMOS technology. 3rd International High Temperature Electronics Conference (HiTEC), Albuquerque (NM, USA), 1996, Transactions 2, XVI-27-32

Soh K L, Kang W P, Davidson J L, Basu S, Wong Y M, Cliffel D E, Bonds A B, Swain G M 2004 Diamond-derived microelectrode array for electrochemical analysis. *Diamond Relat. Mater.* **13**, 2009–15

Solzbacher F, Imawan C, Steffes H, Obermeier E, Möller H 2000 A modular system of SiC-based microhotplates for the application in metal oxide gas sensors. *Sens Actuators B* **64**, 95–101

Song J, Lu W, Flynn J S, Brandes G R 2005 Pt–AlGaN/GaN Schottky diodes operating at 800°C for hydrogen sensing. *Appl. Phys. Lett. 133501-1–3*

Spannhake S, Schulz O, Helwig A, Krenkow A, Müller G, Doll T 2006 *High-temperature MEMS heater platforms: Long-term performance of metal and semiconductor heater materials. Sensors 2006,* Daegu, Korea, pp. 405–19

Spetz A L, Baranzahi A, Tobias P, Lundström I 1997 High temperature sensors based on metal-insulator-silicon carbide devices. *Phys. Stat. Sol. (a)* **162**, 493–511

Spetz A L, Uneus L, Svenningstorp H, Tobias P, Ekedahl L-G, Larsson O, Göras A, Savage S, Harris C, Martensson P, Wigren R, Solomonsson P, Häagendahl B, Ljung L P, Mattson Lundström M I 2001 SiC based field effect gas sensors for industrial applications. *Phys. Stat. Sol. (a)* **185**, 15–25

Srikar V T, Spearing S M 2003 Materials selection for microfabricated electrostatic actuators. *Sens Actuators A* **102**, 279–85

Srinivasan S, Divan R, Zhong L, Carlisle J, Auciello O 2006 Ferroelectric/piezoelectric-diamond hybrid heterostructures for high performance MEMS/NEMS devices. *ICNDST& ADC 2006, Research Triangle Park, NC, USA,* Abstracts Paper B5.5, May 2006, pp. 15–18

Steinhoff G, Baur B, Wrobel G, Ingebrandt S, Offenhäuser A, Dadgar A, Krost A, Stutzmann M, Eickhoff M 2005 Recording of cell action potentials with AlGaN/GaN field-effect-transistors. *Appl. Phys. Lett.* **86**, 033901

Steinhoff G, Hermann M, Schaff W J, Eastman L F, Stutzmann M, Eickhoff M 2003 pH response of GaN surfaces and its application for pH-sensitive field-effect-transistors. *Appl. Phys. Lett.* **83**, 177–9

Stevenson R 2006 Silicon carbide opens the door to radiation-detection market. *Compd. Semicond.* Vol, 12, Jan/Feb 29–30

Stolz A, Behravan M, Regmi M, Golding B 2006 Heteroepitaxial diamond detectors for heavy ion beam tracking. *Diamond Relat. Mater.* **15**, 807–10

Stonas A R, MacDonald N C, Turner K L, DenBaars S P, Hu E L 2001 Photoelectrochemical undercut etching for fabrication of GaN microelectromechanical systems. *J. Vac. Sci. Technol. B* **19**, 2838–41

Stones A R, Margalith T, DenBaars S B, Coldren L A, Hu E L 2001 Development of selective lateral photochemical etching of InGaN/GaN for lift-off applications. *Appl. Phys. Lett.* **78**, 1945–7

Strittmatter R P, Beach R A, McGill T C 2001 Fabrication of GaN suspended microstructures. *Appl. Phys. Lett.* **78**, 3226–8

Strimatter R P, Beach R A, Brooke J, Preisler E J, Picus G S, McGill T C 2003a GaN Schottky diodes for piezoelectric strain sensing. *J. Appl. Phys.* **93**, 5675–81

Strittmatter R P, Beach R A, Picus G S, McGill T C 2003b Piezoelectrically enhanced capacitive strain sensors using metal-insulator-semiconductor diodes. *J. Appl. Phys.* **94**, 5958–63

Strobel P, Riedel H, Ristein J, Ley L 2004 Surface doping of diamond. *Nature* **430**, 439–41

Swain G M 2004 Electroanalytical applications of diamond electrodes. *Thin Film Diamond II (Semiconductor and Semimetals).* vol. 77, Elsevier, Chap. 4, 121–48

Takagaki Y, Hesjedal T, Brand O, Ploog K H 2004a Surface-acoustic-wave transducers for the extremely-high-frequency range using AlN/SiC(0001). *Semicond. Sci. Technol* **19**, 256–9

Takagaki Y, Sun Y J, Brand O, Ploog K H 2004b strain relaxation in AlN/GaN bilayer films grown on γ-$LiAlO_2$ (100) for nanoelectromechanical systems. *Appl. Phys. Lett.* **84**, 4756–8

Taniyasu Y, Kasu M, Kobayashi N 2002 Intentional control of n-type conduction for Si-doped AlN and $Al_xGa_{1-x}N$ ($0.42 < x < 1$). *Appl. Phys. Lett.* **81**, 1255–7

Taniyasu Y, Kasu M, Makimoto T 2006 An aluminium nitride light-emitting diode with wavelength of 210 nanometre. *Nature* **441**, 325–8

Tobias P, Rask P, Göras I, Lundström A, Solomonsson P, Spetz A L 2000 Platinum-insulator-silicon carbide structures as small and fast sensors for exhaust gases. *Proc. Sensoren und Messysteme 2000,* Ludwigsburg, Germany, VDI Berichte 1530, pp. 179–87

Tong Q-Y, Gösele U, Yuan C, Steckl A J, Reiche M 1995 Silicon carbide wafer bonding. *J. Electrochem. Soc.* **142**, 232–6

Tong Q-Y, Gutjahr K, Hopfe S, Gösele U, Lee T-H 1997 Layer splitting process in hydrogen-implanted Si, Ge, SiC, and diamond substrates. *Appl. Phys. Lett.* **70**, 1390–2

Trinchi A, Wlodarski W, Li Y X 2004 Hydrogen sensitive Ga_2O_3 Schottky diode sensor based on SiC. *Sens Actuators B* **100**, 94–8

Tsubota T, Tanii S, Ida S, Nagato M, Matsumoto Y 2004 Chemical modification of diamond surface with various carboxylic acids by radical reaction in liquid phase. *Diamond Relat. Mater.* **13**, 1093–7

Uneus L, Ljung P, Mattsson M, Martensson P, Wigren R, Tobias P, Lundström I, Ekedahl L-G, Spetz A L 1999 Measurements with MISiC and MOS sensors in flue gases. *Proc. Eurosensors XIII, The Hague, The Netherlands,* September 1999, pp. 521–4

Vartuli C B, Pearton S J, Lee J W, Abernathy C R, Mackansie J D, Zolper J C, Shul R J, Ren F 1996 Wet chemical etching of AlN and InAlN in KOH solutions. *J. Electrochem. Soc.* **143**, 3681–3

Vescan A, Ebert W, Borst T H, Kohn E 1996 Electrical characterisation of diamond resistors etched by RIE. *Diamond Relat. Mater.* **5**, 747–51

Vescan A, Daumiller I, Gluche P, Ebert W, Kohn E 1998 High temperature, high voltage operation of diamond Schottky diode. *Diamond Relat. Mater.* **7**, 581–4

Wan Q, Zhang N, Wang L, Shen Q, Lin C 2002 Preparation of PZT on diamond by pulsed laser deposition with Al_2O_3 buffer layer. *Thin Solid Films* **415**, 64–7

Wang Y, Bokor J, Lee A 2004 Maskless lithography using drop-on-demand inkjet printing method. Emerging lithographic technologies VIII. *Proc. SPIE* 5378, 628–36

Wenmakers S, Christiaens P, Daenen M, Haenen K, Nesladek M, van de Ven M, Vermeeren V, Michiels L, Ameloot M, Wagner P 2005 DNA attachment to nanocrystalline diamond films. *Phys. Stat. Sol. (a)* **202**, 2212–6

Werner M, Klose S, Szücs F, Moelle Ch, Fecht H J, Johnston C, Chalker I M, Buckley-Golder P R 1997a High temperature Young's modulus of polycrystalline diamond. *Diamond Relat. Mater.* **6**, 344–7

Werner M, Locher R, Kohly W, Holmes D S, Klose S, Fecht H J 1997b The diamond Irvin curve. *Diamond Relat. Mater.* **6**, 308–13

Wild C 1998 CVD diamond for optical windows. *Low-Pressure Synthetic Diamond – Manufacturing and Applications.* Springer, Chap. 10, 189–206

Wild C, Kohl R, Herres N, Müller-Sebert W, Koidl P 1994 Oriented CVD diamond films: Twin formation, structure and morphology. *Diamond Relat. Mater.* **3**, 37–381

Windheim J A, Dreifus D L, Fox B A, Holmes J S, Malta D D M, Plano L S, Stoner B R, Wynands H A 1995 High temperature electronic applications of diamond. *HITEN News(AEA),* No 008, July, pp. 6–10

Wolfenden A 1997 Measurement and analysis of elastic and anelastic properties of alumina and silicon carbide. *J. Mater. Sci.* **32**, 2275–82

Wörner E 1998 Thermal properties and applications of CVD diamond. *Low-Pressure Synthetic Diamond – Manufacturing and Applications.* Springer, Chap. 9.5, pp. 179–184

Wu S, Chen Y-C, Chang Y S 2002 Characterization of AlN films on Y-128 degree $LiNbO_3$ by surface acoustic wave measurements. *Jpn. J. Appl. Phys.* **41**, 4605–8

Xiao X, Birrell J, Gerbi J E, Auciello O, Carlisle J A 2004 Low temperature growth of ultrananocrystalline diamond. *J. Appl. Phys.* **96**, 2232–38

Yang S G, Aslam D M 1996 Single-structure heater and temperature sensor using a p-type polycrystalline diamond resistor. *IEEE Electron Device Lett.* **17**, 250–52

Yang Y, Wang X, Ren C, Xie J, Lu P, Wang W 1999 Diamond surface micromachining technology. *Diamond Relat. Mater.* **8**, 1834–7

Yang Y T, Ekinci K L, Huang X M H, Schiavone L M, Roukes M L, Zorman A, Mehregany M 2001 Monocrystalline silicon carbide nanoelectromechanical systems. *Appl. Phys. Lett.* **78**, 162–4

Yang W, Auchiello O, Butler J E, Cai W, Carlisle J A, Gerbi J E, Gruen D M, Knickerbocker T, Lasseter T L, Russell J L, Smith L M, Hamers R J 2002 DNA-modified nanocrystalline diamond thin-films as stable, biological active substrates. *Nat. Mater.* **1**, 253–7

Young D J, Du J, Zorman C A, Ko W H 2004 High-temperature single-crystal 3C-SiC capacitive pressure sensor. *IEEE Sens. J.* **4**, 464–70

Youtsey C, Adesida I, Bulman G 1997 Highly anisotropic photoenhanced wet etching of n-type GaN. *Appl. Phys. Lett.* **71**, 2151–3

Yun J, Dandy D S 2005 A kinetic model of diamond nucleation and silicon carbide interlayer formation during chemical vapour deposition. *Diamond Relat. Mater.* **14**, 1377–88

Yun F, Chevtchenko S, Moon T-T, Morkoc H 2005 GaN resistive hydrogen gas sensors. *Appl. Phys. Lett.* **87**, 073507-1–3

Zaitsev A M, Burchard M, Meijer J, Stephan A, Burchard B, Fahrner W R, Maresch W 2001 Diamond pressure and temperature sensors for high-pressure high-temperature applications. *Phys. Stat. Sol. (a)* **185**, 59–64

Zeng J, Wang L, Goa J, Song Z, Zhu X, Lin C, Hou L, Liu E 1999 Structural and electrical characteristics of oriented $Pb(Zr_{0.52}Ti_{0.48})O_3$ ferroelectric thin films deposited on diamond substrates by a simple sol-gel process. *J. Crystal Growth* **197**, 874–8

Zhao J S, Park D-Y, Seo M J, Hwang C S, Han Y K, Yang C H, Oh K Y 2004 Metallorganic CVD of high-quality PZT thin films at low temperature with new Zr and Ti precursors having MMP ligands. *J. Electrochem. Soc.* **151**, C283–91

Zhou D, Gruen D M, Qin L C, McCauley T G, Krauss A R 1998 Control of diamond film microstructure by Ar additions to CH_4/H_2 microwave plasmas. *J. Appl. Phys.* **84**, 1981–9

Zhu X, Aslam D M 2006 CVD diamond thin film technology for MEMS packaging. *Diamond Relat. Mater.* **15**, 254–8

Zhu X, Aslam D M, Tang Y, Stark B H, Najafi K 2004 The fabrication of all-diamond packaging panels with built-in interconnects for wireless integrated microsystems. *J. Microelectromech. Syst.* **13**, 396–405

Zhuang D, Edgar J H, Liu L, Lui B, Walker L 2002 Wet chemical etching of AlN single crystals. *MRS Internet J. Nitride Semicond. Res.* **7(4)**, 1–6

Zhuang D, Herro Z G, Schlesser R, Sitar Z 2006 Seeded growth of AlN single crystals by physical vapour transport. *J. Crystal Growth* **287**, 372–5

Ziermann R, von Berg J, Reichert W, Obermeier E, Eikhoff M, Krötz G 1997 A high temperature pressure sensor with β-SiC piezoresistors on SOI substrates. *Proc.1997 Int. Conf. Solid-state Sensors and Actuators,* Chicago, IL, USA, June 1997, pp. 1411–14

Zimmermann T, Neuburger M, Kunze M, Daumiller M I, Denisenko A, Dadgar A, Krost A, Kohn E 2004 P-channel InGaN-HFET structure based on polarization doing. *IEEE Electron. Devices Lett.* **25**, 450–2

Zimmermann T, Janischowsky K, Denisenko A, Kubovic M, Kohn E 2005a Nanocrystalline diamond pn structure grown by hot-filament CVD. *8th Applied Diamond Conference/ NanoCarbon 2005,* Argonne National Laboratories, Argonne, IL, USA, Paper 5.4, May 2005

Zimmermann T, Kubovic M, Denisenko A, Janischowsky K, Williams O A, Gruen D M, Kohn E 2005b Ultra-nano-crystalline/single crystal diamond heterostructure diode. *Diamond Relat. Mater.* **14**, 416–20

Zimmermann T, Neuburger M, Benkart P, Hernandez-Guillen F J, Pietzka C, Kunze M, Daumiller I, Dadgar A, Krost A, Kohn E 2006 Piezoelectric GaN sensor structures. *IEEE Electron Devices Lett.* **27**, 309–12

Biography

Dr. Kohn earned his Ph.D. at the Technical University of Aachen (Germany) in 1975. After many years in industry in Germany, France and the U.S. he joined the University of Ulm as professor and director of the Department of Electron Devices and Circuits in 1989. His research activities include design and technology of advanced electronic device structures, especially field effect transistors, in a number of semiconductors, recently III-Nitride heterostructures and diamond for high temperature and high power applications. This is accompanied by work on advanced packaging technologies and MEMS sensor and actuator devices, especially based on CVD diamond, mainly for RF applications and applications in electrochemistry and life science. He is a member of the IEEE and Electrochemical Society and has served as IEEE distinguished lecturer. Currently he is serving as conference organizer of the 'Diamond' conference series. Dr. Kohn has coordinated joint German and European project groups and has been a partner to US programs. He has been a visiting fellow to the University of Wales, the Norwegian Institute of Technology, Cornell University and the Air Force Research Laboratories in Dayton (Ohio) and has been an Adjunct Professor of Physics at the New Jersey Inst. of Technology, Newark (NJ, USA). He is the director of the Steinbeis Technology Transfer Centre 'Semiconductor Devices'. His activities have led to two spin-off companies, GFD (Gesellschaft für Diamantprodukte) and MicroGaN. He has contributed with more than 300 scientific publications to the field.

1.07 Wet Etching

Kazuo Sato and Mitsuhiro Shikida, Department of Micro-Nano Systems Engineering, Nagoya University, Aichi, Japan

Glossary

Anisotropic etching Etching system where etching front does not advance omni-directionally. There are two types of anisotropic etching by the origin of directionality. One is the directionality caused by the incident angle of etching media, i.e. ions, atoms, and solid particles. Another is the directionality caused by the etched material that has an anisotropic nature like single-crystals. Antonym: **isotropic etching**.

Anodic etching Etching system that provides a positive electric potential to the wafer during etching. When silicon is etched in highly concentrated HF solution under positively biased potential and illumination, the silicon surface shows a variety of morphologies from porous to smooth ones depending on the electric potential and the strength of the illumination.

Anodization When anodic etching is applied to silicon wafer using HF solution as an etchant, silicon surface is etched porously, maintaining the single crystal structure in the remaining material. Such etching system is used for making porous silicon for applications such as photo- and electroluminescent devices, sacrificial layer of microstructures, and so on.

Buffered HF (BHF) Buffered HF is a HF solution diluted with water or NH_4F water solution. When oxide film on a silicon surface is needed to be removed, buffered HF solution is useful, because it selectively dissolves the oxide film at a moderate etch rate.

Diffusion-limited process Chemical etching is a series of process as follows; diffusion of a reactant in the environment, absorption of the reactant to the substrate surface, reaction between the absorbed reactant and substrate material, detachment of the

reaction product, and diffusion of the product to the environment. When the diffusion time is dominant over the total process, i.e. the reaction time is negligible compared to diffusion, the total process is called as diffusion-limited process. In case of diffusion limited process, stirring the solution is effective to enhance the etch rate. Antonym: **Reaction-limited process**.

EDP Mixture of ethylenediamine and pyrocatecol water solution. This is an anisotropic etching solution for single crystal silicon. It has excellent selectivity to oxide mask and also to p^{++} etch stop layer. Disadvantages of EDP solution is that it is highly toxic, and it can easily react with oxygen and quickly ages.

Electropolishing For the purpose of smoothing a silicon surface, photoelectrochemical etching using highly concentrated HF solution is used. Depending on the applied potential, the etched surface morphologies changes from porous to smooth ones. Higher potential is recommended to obtain a bright surface.

Etch stop technique For the purpose of controlling the height of microstructure or membrane thickness in the etching process, etch stop techniques have been developed. In case of anisotropic etching of silicon where uniform thickness reduction is available, p^{++} layer, p-n junction, and SOI wafers are utilized for the etch stop purposes.

HF, HNO_3, CH_3COOH water solution HF, HNO_3, CH_3COOH water solution is an isotropic etching solution for silicon. The system is diffusion-limited process, and thus it is isotropic.

Isopropyl alcohol (IPA) For the purpose of suppressing mask corner undercut during KOH etching of Si(100) wafer, isopropyl alcohol is dissolved to KOH water solution to the saturation level. The corner undercut decreases because that the etch rate ratio of Si (110) to (100) decreases by the addition of IPA to KOH solution. Such solution is also used to get a better smoothness on Si(100), though large hillocks can sometimes grow on Si(100) as a drawback.

Isotropic etching Etching system where etching front advances omni-directionally. Mask undercut advances laterally from the mask edge with the same distance to the etch depth. When the etch depth becomes large, a small sized island mask pattern will disappear from the surface. Antonym: **Anisotropic etching**

KOH Potassium hydroxide (KOH) water solution is an alkaline solution, showing strong anisotropy in etch rate of single crystal silicon. It is easy to handle compared to EDP, showing stronger anisotropy compared to TMAH solution.

Laser beam processing When a laser, typically YAG laser, is irradiated to the wafer before or during etching, one can get a different type of etch profiles than without laser irradiation. This is because of the thermal effects, one is destroying crystalline structure of the substrate material prior to etching, and other is local enhancement of etching at the laser irradiated surface during etching. Laser-enhanced etching is sometimes applied to anisotropic etching of silicon.

Mask corner compensation When fabricating thick micromechanical structures on a wafer by utilizing etching, the mask periphery is severely attacked by side etching. Even when anisotropic chemical etching is applied, mask corners of an island pattern are heavily undercut. One of the methods protecting the mask corner from the undercut is to add an additional masked area to the corner of the island. This additional masking area is called mask corner compensation.

Mechanical dicing Dicing is a process using a rotating grinding wheel to make grooves on a silicon wafer. It is usually used to cut silicon devices into pieces from a wafer. Dicing can be also utilized to make grooves followed by chemical anisotropic etching for making microstructures on a silicon wafer. Laser dicing is an alternative to the mechanical dicing for such purposes.

Open-circuit-potential (OCP) When a silicon substrate is etched in an alkaline solution in an equilibrium state, the substrate stays at a negative potential assuming that the solution is at zero potential. There is no electrical current flowing between the substrate and the solution. The substrate potential under such a condition is called as the open-circuit-potential.

p^{++} layer When silicon is heavily doped by boron at the concentration higher than $1 \times 10^{-19}\,cm^{-3}$, the etch rate of silicon in alkaline solution decreases two orders of magnitude compared to lightly-doped silicon. Because that the boron-doped silicon is p-type, heavily doped silicon region is called p^{++} layer. By utilizing this phenomena, the p^{++} layer is used as an etch stop layer for fabricating membrane structures on a silicon wafer. It is reported that a p^{++} layer doped with Ge or P does not show such a drastic decrease in etch rate observed with B.

Passivation potential When a silicon substrate is etched in an alkaline solution, the electrical potential of the substrate is at negative potential in equilibrium against the solution which is at zero potential. When applying a potential to the silicon towards positive direction, the silicon surface started to be oxidized and etching stops. The critical potential at which the etching stops is called as passivation potential.

p-n Junction A silicon p-n junction is useful for the etch stop technique during alkaline etching. The applied etch-stop system is set such that the n-silicon layer is biased in positive potential, while the p-substrate is exposed to the etchant. Because of the existence of p-n junction, no current flows through the junction even the n-layer is positively biased. Thus the etching advances in the p-substrate as long as the etching front does not meet the junction. However, once the etching front reaches the n-layer, current flows from the silicon to the solution. Thus the etching stops leaving a precisely controlled thickness of n-layer on the wafer.

Porous silicon Photoelectrochemical etching of single crystal silicon using highly concentrated HF solution results in a variety of surface morphologies. When the applied potential is low, porous silicon layer is obtained. Porous surface remain a single crystal. Such a layer is utilized for photo and electroluminescent devices, sacrificial layer of microstructures, and so on.

Reaction-limited process Chemical etching is a series of process as follows; diffusion of a reactant in the environment, absorption of the reactant to the substrate surface, reaction between the absorbed reactant and substrate material, detachment of the reacted product, and diffusion of the product to the environment. When the reaction time is dominant over the total process, i.e. the diffusion time is negligible compared to reaction, the total process is called as reaction-limited process. In the case of reaction limited process, crystalline materials often show anisotropy in the etch rate. Antonym: **Diffusion-limited process**.

Selectiveness Etching utilizes selectiveness in many ways. The etch rate selectivity between the work and etching mask materials is essential for etching process. The other example is the etch-stop techniques described in the text. Anisotropic etching is another case of selectiveness in the process.

Silicon-on-insulator (SOI) The wafer having a single crystalline silicon layer at the top of a wafer electrically insulated from the wafer substrate. Insulator can be a silicon oxide film separating the surface from the substrate. In such a case, the buried oxide film works as an etch stop layer for alkaline etching system.

Step flow model Etching model assuming that the crystal surface is composed of terrace and atomic steps. Terrace surface is chemically stable, while the atomic steps are easy to etch. Etching proceeds mainly as a removal of atoms at the step edges, consequently rows of the atomic step move in lateral direction during etching. This model is developed and established in the field of crystal growth, the reverse process of etching.

Surface micromachining One of the micromachining methods to create movable mechanical structures on a wafer. It is composed of a multilayered film deposition processes each followed by the patterning the film by etching. After completing the layers of deposition process, a certain layer is selectively etched off. As a result, remaining film layer becomes a movable structure partly supported by the substrate, or even fully released from the substrate. This type of machining is in contrast to the bulk micromachining that utilize the substrate material as a structure.

Surfactant During silicon anisotropic etching using alkaline solution, bubbles are generated as an etching product on the silicon surface. It is believed that bubble detachment from the silicon surface is enhanced by the addition of surfactant to the solution, thus results in a smooth surface finish.

TMAH Tetra-methyl-ammonium-hydroxide water solution. It is an alkaline solution, and used as an anisotropic etching solution for silicon. It has high selectivity in etch rate to the silicon oxide film as a mask material. It does not contain alkaline metal ions such as potassium and sodium which are harmful to ICs on the chip. It is also used as a developer for the photo-resist materials in the lithography process.

Wagon wheel Wagon wheel shaped mask pattern used for visualizing anisotropy in etch rate.

Wulff–Jaccodine's construction A graphical method predicting an anisotropically etched profile on a single crystal surface.

1.07.1 Introduction

Wet etching characteristics are classified into two groups, i.e., isotropic and anisotropic systems. The etching mechanism is analyzed as a series of materials transfer and chemical reaction processes. The variation in etching characteristics like isotropy and anisotropy is explained by the difference in the dominant process among series of materials transfer and chemical reaction processes.

Isotropic etching systems and its applications are introduced for silicon, silicon dioxide, and silicon nitride.

Anisotropic etching systems for single-crystal silicon are introduced. Selections of etching solutions and etching mask materials are discussed. Variation in etching shapes obtained by anisotropic etching is described in relation to the etching solution, the wafer orientation, and the etch-mask design. The etch rates are given in detail as a function of crystallographic orientations of silicon. A wide range of practical technologies for controlling etching shapes are introduced and summarized. Finally, advantages and disadvantages of wet and dry etching systems are compared from the engineering point of view.

Origin of etching anisotropy is discussed on the atomistic level. It is shown that the number of dangling bonds of the surface atom is not an appropriate index representing anisotropy in etching.

Etching profile simulation systems supporting process design of single-crystal materials are introduced. Simulation results are demonstrated for silicon and quartz etching.

Anodic etching system that allows electropolishing or porous silicon layers on a silicon wafer is introduced. Applications such as selective removal of microstructures, micro and macroporous silicon are described.

1.07.2 Etching Characteristics – Isotropy and Anisotropy in Etching

Etching is a micromachining process that removes solid material from a substrate surface. Wide ranges of etching media such as liquid, gas, plasma, and ion are utilized. This chapter focuses on wet etching processes that use liquids, i.e., etchants, as the etching media. When the work surface is partially coated with a protective film, etching advances selectively in an area without film coating. The protective films, i.e., etching masks, are made of polymers (photoresist materials), metals, and silicon compounds, which can be selected depending on the resistivity against the etchant.

The aim of an etching process is to selectively remove materials. The key word of etching is selectiveness. The following are examples of how selectiveness plays a key role in etching processes.

The first example is selective removal of material from a work surface only in the area of the mask aperture that is patterned in an arbitrary 2D shape. This is a normal method known as photoetching. Due to the advancement of photolithography technologies, photoresist films as protective films can be well shaped in fine patterns on the order of submicrometers in size. Such fineness of the material removal pattern is the major advantage of the etching process over other material removal processes such as mechanical machining.

The second example is selective removal of a material from other materials. The simplest application is removing an SiO_2 film from the Si surface. Fluoric acid attacks the oxide, but attacks silicon only slightly. An advanced application is to remove a single layer of a material among layered materials that coexist on the same substrate. Even when different materials are layered on the substrate, a single material layer can be selectively removed, as long as an opening is connected to the etchant from the material to be removed. This type of removal is further applied to surface micromachining to release a structural film material partly from the substrate. This allows fabricating freestanding micromechanical structures that move or mechanically oscillate on the substrate.

Thus, selectiveness is the key issue that makes etching processes successful. **Table 1** summarizes liquid-phase etchants for silicon and silicon compounds. An etched recess shows different shapes depending on the difference in etching systems.

The cross sections of typical etched recesses are shown schematically in **Figures 1** and **2**. Etching systems are categorized into two classes depending on the characteristics of the etching profiles. One is isotropic and the other is anisotropic etching. When etching advances normal to solid surfaces during the etching process at any place on the substrate except the masked surface, the etched cross section becomes like the one shown in **Figure 1**. Etching advances under the mask material and results in a round profile at the mask periphery. The etching advancing under

Table 1 Wet etching systems for silicon and silicon compounds

Etched material	*Masking material*	*Etchant*	*Isotropic (iso) and anisotropic (A)*
Si	Au (Cr as base binder)	49wt.% HF + 69wt.% HNO_3 + CH_3COOH (typical volume fraction 2:3:3)	iso
	SiO_2, Si_3N_4	KOH water solution	A
	SiO_2, Si_3N_4	TMAH water solution	A
	Si_3N_4, Au, Cr	EDP water solution	A
SiO_2	Photoresist	49wt.% HF + 50wt.% NH_4F (typical volume fraction 1:6)	iso
Quartz	Au (Cr as base binder)	NH_4HF_2-saturated solution	A
Glass	Photoresist, Cr	HF water solution	iso
Si_3N_4	Photoresist, SiO_2	H_3PO_4 water solution boiling at 180°C	iso

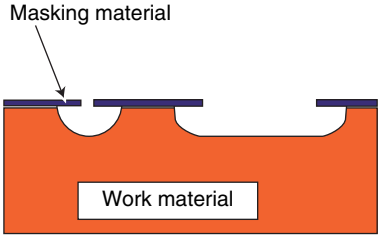


Figure 1 Isotropically etched profiles.

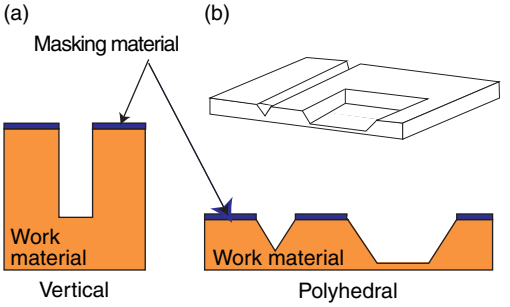


Figure 2 Anisotropically etched profiles. (a) Vertical, (b) polyhedral.

the mask is called mask undercut, or side etching. When a deep isotropic etching is applied to a narrow-patterned mask area, side etching from both sides meet at the center of the masked area, and the mask is finally lost. This type of etching is called isotropic etching.

On the contrary, when material removal takes place selectively in certain directions, the cross sections of the etched recess become like the one shown in **Figure 2**. Such etching is called anisotropic etching. When etching advances only in a single direction perpendicular to the substrate, the etched profile becomes like the one shown in **Figure 2(a)**. When the work material has an anisotropic nature, for instance, being a single crystal, the etched cavity often becomes polyhedral as shown in **Figure 2(b)**.

Unlike electronics circuit fabrication, microelectromechanical systems (MEMS) require that highly 3D structures be fabricated, and the etching depth sometimes reaches tens or hundreds of micrometers. For such fabrication purposes, anisotropic etching is useful, because it can transfer 2D mask pattern deep into the substrate. On the other hand, in isotropic etching, fine pattern of the masking material is easily lost after some depth of etching because of heavy undercut etching that advances beneath the masking material.

Anisotropic etching is further categorized into two depending on the anisotropy. One type utilizes directionality of the etching species physically hitting the surface of the substrate. This can be achieved by dry etching processes that utilize ions accelerated by external electromagnetic or electrostatic forces. Typical processes are reactive ion etching (RIE) and ion beam etching. The etched profile looks like that shown in **Figure 2(a)**. Another type of anisotropic etching utilizes the anisotropy of the substrate material, e.g., single crystals. Both silicon and quartz are available as single-crystal wafers. Etch rates are strongly dependent on the crystallographic orientations; thus the etched cross section shows variations as given in **Figure 2(a)** and **2(b)**, by combinations of wafer orientations and mask designs. The process is dominated by the chemical activeness of the orientations of single crystals. This is further explained in the next two sections.

Single-crystal silicon is wet etched both isotropically and anisotropically depending on the etching systems (**Table 1**). This is explained below.

Chemical wet etching results from a series of processes. The processes are:

(1) Diffusion of etching species in solution approaching the solid surface.
(2) Absorption of the species onto the solid surface.
(3) Reaction of the etching species and the solid atom.
(4) Release of the etched product from the solid surface.
(5) Diffusion of the product into the solution.

These processes are schematically shown in **Figure 3**. Although processes (2), (3), and (4) are not easily detectable as separate, diffusions (1) and (5) are definitely different processes from the rest. The time duration necessary for etching is schematically shown in **Figure 4** for two cases: one for fast reaction cases and another for slow reaction cases.

In the case of fast reactions, the etching species always becomes limited near the solid surface. Materials transfer in the solution is a key to accelerate the etching process. The process is dominated by diffusion. Such a process is called diffusion-limited process as shown in **Figure 4**. The silicon etching process uses a mixture of HF, HNO_3, and CH_3COOH, as given in **Table 1**. Because the chemical reaction is so fast, the surface always suffers from the lack of an etching species. Due to the isotropic nature of diffusion in liquid, the etching profile change becomes isotropic for the diffusion-limited etching process. When one wants to enhance the etching, it is effective to stir the etching solution. This enhances the supply of the etching species.

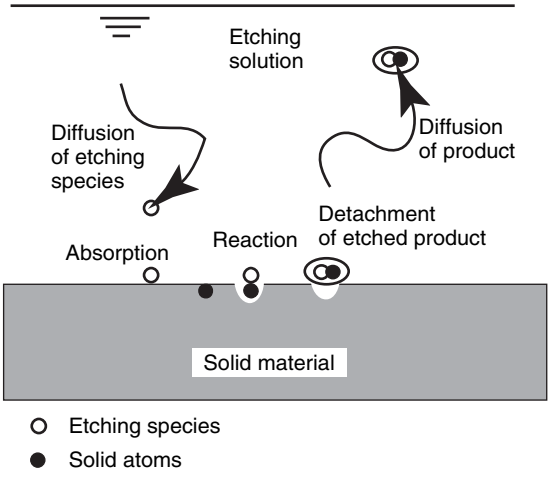


Figure 3 Series of processes that take place during wet etching.

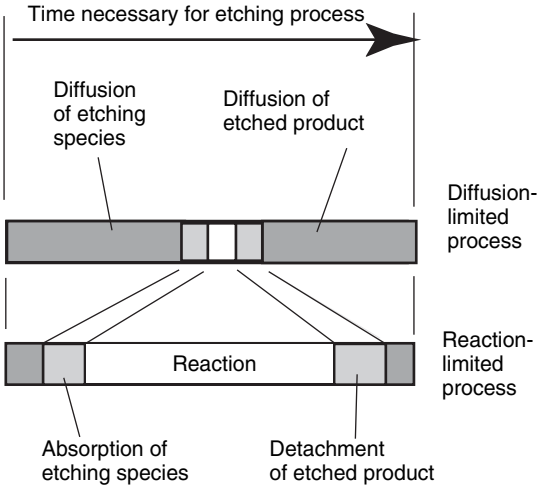


Figure 4 Comparison of diffusion-limited and reaction-limited processes in time domain.

However, it is difficult to achieve uniform flow velocity in omnidirections over the silicon wafer surface. Uniformity in etching profile usually deteriorates because of the stirring motion. An example is the silicon acoustic lens, which has a concave lens profile etched by using this method, as mentioned in Section 1.07.3. Excellent uniformity and sphericity of the lens profile were obtained by stopping the stirring agitation, i.e., by sacrificing etch rates as a trade-off.

On the other hand, in a slow reaction process, i.e., in the case of alkaline etching of silicon using KOH, tetramethylammonium hydroxide (TMAH), and ethylenediamine (EDP) (**Table 1**), the dominant process is the reaction shown in **Figure 4**. Such a system is called reaction-limited process. Now the different reactivities among different crystallographic orientations become obvious. As mentioned in the next two sections, the etch rates for (100) and (111) silicon, using a KOH water solution, differs by nearly 2 orders of magnitude. This is because of the difference in the chemical activeness of the crystal surface having different surface morphologies due to the crystal lattice structure. Even in the case of an isotropic system, such morphological difference exists for single-crystal silicon surfaces, and actually in a $HF/HNO_3/CH_3COOH$ system, a small amount of anisotropy can be observed in the geometry of the etched cavity on a single-crystal silicon, which is discussed in Section 1.07.3. It is understood from **Figure 4** that sometimes the reaction part is not negligible, even in the case of diffusion-limited processes.

1.07.3 Isotropic Etching of Silicon and Related Materials

Isotropic etching is a process in which etching advances in normal to solid surfaces with a uniform etch rate at any place on the substrate except on the masked surface, as has been discussed in Section 1.07.2. Examples of isotropic etching solutions are given in Table 1. The isotropic etching solutions are very reactive chemicals for work materials. Thus, the etching system becomes a diffusion-limited process.

Silicon is isotropically etched with a mixture of HF, HNO_3, and CH_3COOH water solution. Etching advances according to the following series of reactions:

$$HNO_2 + HNO_3 \rightarrow 2HNO_2 + 2OH^- + 2h^+$$

$$Si + 2H_2O + 4h^+ \rightarrow SiO_2 + 4H^+$$

$$SiO_2 + 6HF \rightarrow H_2SiF_6 + 2H_2O$$

The first equation explains that HNO_3 generates OH^- and holes. The second equation is the oxidization of silicon by the holes. In the third equation silicon dioxide is dissolved by HF. The CH_3COOH added in the mixture is used to moderate etching reactions. Depending on the ratio of these three components, the etch rate drastically changes as shown in Figure 5.

In Figure 5, photographs of the top view of the etched cavity starting from a small circular mask aperture are shown. They are of quasi-round profiles, reflecting the isotropic nature of the etching system. The profile of the cavities is the etching fronts deeply undercutting the mask apertures.

When the profile of the cavities etched with different fractions of three components is compared, a cavity etched under the condition A has a slightly square shape representing an anisotropic nature. On the other hand, the one etched under condition C is isotropic, though the bottom of the cavity is not as smooth as the other that provides clear reflection of the illumination from the bottom. Thus, isotropy and surface smoothness are in a trade-off relationship. It is recommended that a mixture of 49 wt.% HF:69 wt.% HNO_3:CH_3COOH with a volume fraction of 2:3:3, as given in Table 1, and an etching temperature at 50°C (Hashimoto *et al.* 1993) for compromising the isotropy and surface smoothness be chosen.

Isotropic etching solution for SiO_2 is HF water solutions. A concentrated 49% HF solution rapidly etches SiO_2 with an etch rate of 1.8 μm min^{-1}. Instead of such a strong etchant, buffered HF (BHF) is usually used to etch an SiO_2 film, which is deposited on silicon wafers by thermal oxidization or CVD processes. Typical components are 49 wt.% HF + 50 wt.% NH_4F with a volume fraction of 1:6

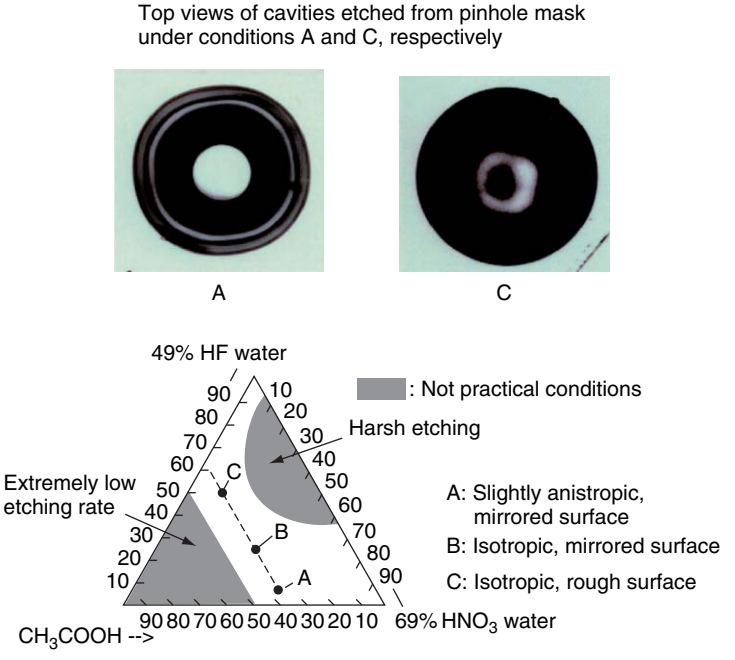


Figure 5 Etching characteristics change based on the change in relative ratio in the HF–HNO_3–CH_3COOH system.

or 49 wt.% HF + 40 wt.% NH_4F with a volume fraction of 1:7. Etch rate is about 0.1 μm min^{-1} at room temperature. This is a mild solution, in contrast to a concentrated 49 wt.% HF solution.

It is widely used for removing SiO_2 films. It is also used as an etchant for releasing polycrystalline silicon microstructures from a silicon substrate during surface micromachining. In such a case, etch rate of the sacrificial layer of SiO_2 can be enhanced by doping P, and in some cases by also doping B, to the oxide film. Phosphosilicate glass (PSG) and boron-doped PSG (BPSG) are removed by the diluted HF solutions. Addition of up to 50 vol.% HCl to the diluted HF is recommended for enhancing the etch rate ratio of the sacrificial glass layers of the structure materials such as polycrystalline silicon and silicon nitride.

Isotropic etching solution for Si_3N_4 is H_3PO_4 water solution at an elevated temperature of 160–180°C. Positive photoresist materials cannot withstand such a harsh condition. Some negative photoresists are used for patterning Si_3N_4 films and an SiO_2 film deposited on the nitride is also used as a mask material against hot H_3PO_4 solution.

Isotropic etching is not widely used in micromachining, because substrates cannot be deeply etched using this system that results in a heavy side etching that deteriorates fine mask patterns. In exceptional cases, such as when fabricating hemispherical concave cavities, an isotropic etching system is useful.

Isotropic etching is used for fabricating an acoustic lens on a silicon substrate (Hashimoto *et al.* 1993). The photographs of the silicon acoustic lens are shown in **Figure 6**. It has a concave profile at the tip whose radius is 55 μm with an excellent sphericity of 0.2 μm with a mirror-finished surface. It is mounted on a scanning acoustic microscopy system, and is used for focusing and receiving ultrasonic vibration at 1 GHz allowing an acoustic image resolution of 1.8 μm. The etchant used is a HF, HNO_3, CH_3COOH water solution. The periphery of the cavity is mechanically ground in a tapered shape using an abrasive grinding method.

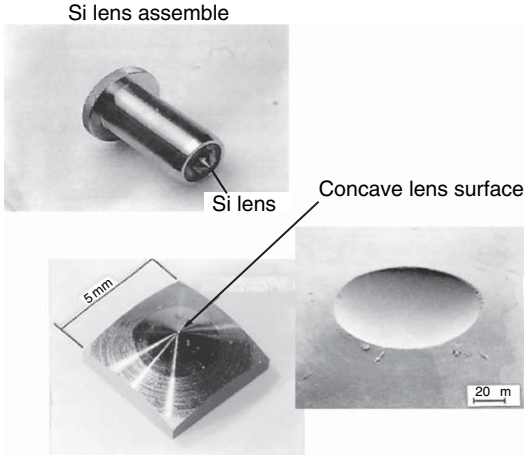


Figure 6 Isotropically etched silicon lens applied to an acoustic microscopy system.

1.07.4 Anisotropic Etching

Anisotropic chemical etching of silicon has been used for fabricating various micromechanical devices such as pressure sensors, acceleration sensors, ink-jet printing heads, and chemical analysis systems. As the (111) orientation of single-crystal silicon is chemically stable and has an extremely low etch rate when compared to other orientations, this type of etching system has been utilized to minimize undercut from a mask pattern edge during deep etching of silicon wafers. Only for such purposes, science in etching mechanism is of less importance than is developing practical technologies. Practical technologies such as etching solutions, mask materials, and etch stop techniques have been developed and widely utilized. However, depending on the increasing demands for dimensional accuracy of fabricated device, for quality of etched surface, and for fabricating various 3D structures, a deeper understanding of etching mechanisms is required.

The aim of this chapter is to give an overview of recent advances in anisotropic etching science as well as to introduce state-of-the-art technologies for MEMS microfabrication.

1.07.4.1 Anisotropic Etching of Silicon

Single-crystal silicon has a diamond lattice structure as shown in **Figure 7**. Each silicon atom has four covalent

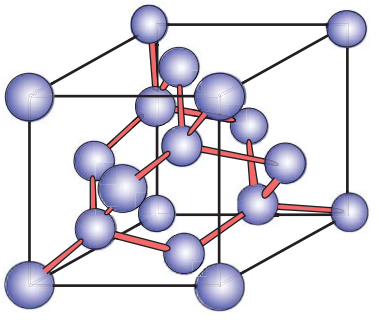


Figure 7 Diamond lattice structure of single-crystal silicon.

bonds. Each bond connects a different pair of atoms. Figure 8 compares three different surfaces cut from the same crystal. They are indicated by using the Miller index as (100), (110), and (111), respectively. When an atom is located on a surface, the bond belonging to the atom loses a neighboring atom. It is known as a dangling bond. The dangling bonds easily react with the etching agent. Though, in many cases, the dangling bonds do not remain free-ended, that is, the surface bonds are reconstructed by combining with each other in high vacuum, or the bonds are terminated with hydrogen atom in water, those bonds on the surface are still a source of surface reactions, i.e., etching.

When the number of dangling bonds on three differently oriented surfaces, (100), (110), and (111), are compared, it is apparent that the (111) surface has the smallest number of bonds (Figure 8). There is only one dangling bond per surface atom for (111), whereas there are two for (100), and one dangling bond plus two surface bonds for (110).

This is a conventional explanation for why (111) is stable against etching. It is partly true when the stability of the surface is discussed qualitatively. However, atom removal from perfectly flat (111) is a minor case that actually occurs on a crystal surface. In most of the cases, a large number of atoms are removed from atomic steps and kinks. Etch rate must be quantitatively discussed based on such dynamic models. It turned out that the conventional static model is insufficient to describe anisotropy in the etch rate for silicon. This is discussed later in this chapter.

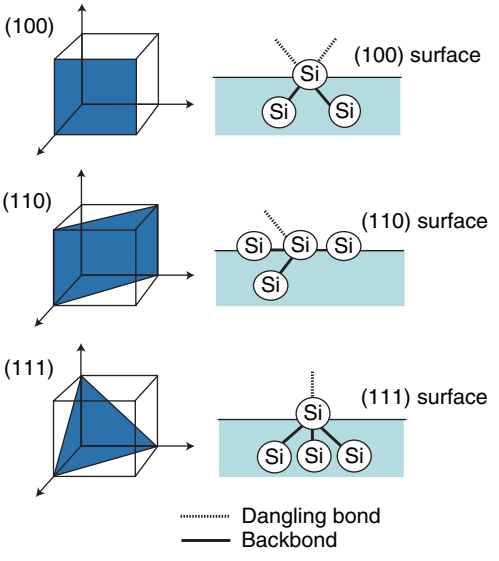


Figure 8 Number of backbonds and dangling bonds of a Si atom on a surface oriented at (100), (110), and (111).

1.07.4.1.1 Etching solutions

Solutions showing orientation dependence on the etch rate of silicon are KOH, TMAH, EDP (ethylenediamine and pyrocatecol), N_2H_4 (hydrazine), CsOH, and NaOH. All are used as water solutions. The chemical reaction for any of these etching solutions is described as follows:

$$Si + 2OH^- + 2H_2O \rightarrow SiO_2(OH)_2^{2-} + 2H_2$$

Silicon reacts with water and an OH^- ion and produces hydroxide ion and hydrogen gas bubbles.

Etching masks are usually made of either SiO_2 or Si_3N_4. The Si_3N_4 mask is excellent in terms of etching selectivity, but when compared to an SiO_2 mask it is not easy to apply on and to remove from the wafer surface. Photosensitive resist films are not durable against etching. Metals such as Ni are sometimes used as a mask material. However, one has to be careful against contamination of solutions with metallic ions, which may deteriorate etching properties such as etched surface roughness and etch rates.

Selection of the etching solutions is made by the following guidelines.

1.07.4.1.1.(i) KOH KOH shows strong anisotropy, and shows large values of etch rate ratios among orientations of about 100 as listed in Table 2 (Shikida *et al.* 2000). It means that high controllability can be expected in etched profiles, while suppressing mask undercut. KOH solutions are less toxic than other etchants. Exhaust solutions after use are easy to process. These are the main reasons why KOH is widely used for fabricating silicon microstructures in industry.

A drawback with KOH is that an SiO_2 mask is significantly etched during a long etching time. Selectivity of Si to SiO_2 mask is about 150 under normal conditions. Another drawback is that KOH etching is incompatible to IC processes because contamination with potassium ion is strictly prohibited in IC processes. Though KOH etching is possibly separated from an IC process, and potassium ion can be removed from the device surface by appropriate rinsing, care must be taken when the microstructure is integrated with electronics circuits.

1.07.4.1.1.(ii) Tetramethylammonium Hydroxide Tetramethylammonium hydroxide (TMAH) is the second major etchant, although it has only recently been introduced (Tabata *et al.* 1991). It is a counterpart of KOH. Etching rate ratios among orientations are

Table 2 Anisotropic etch rates for KOH and TMAH solutions

	Etch rate ($\mu m\,min^{-1}$)		*Etch rate normalized by that for (100)*	
Orientation	*34wt.% KOH (70.9°C)*	*20wt.% TMAH (79.8°C)*	*34wt.% KOH (70.9°C)*	*20wt.% TMAH (79.8°C)*
(100)	0.629	0.603	1.000	1.000
(110)	1.292	1.114	2.054	1.847
(210)	1.237	1.154	1.967	1.914
(211)	0.983	1.132	1.563	1.877
(221)	0.586	1.142	0.932	1.894
(310)	1.079	1.184	1.715	1.964
(311)	1.065	1.223	1.693	2.028
(320)	1.285	1.211	2.043	2.008
(331)	0.845	1.099	1.343	1.823
(530)	1.273	1.097	2.024	1.819
(540)	1.283	1.135	2.040	1.882
(111)[a]	0.009	0.017	0.014	0.027

[a]Etch rates for (111) were too small to be measured from the spherical surface. It was separately measured from groove width increments during deep groove etching on Si (110) wafer (Shikida *et al.* 2000).

less than that of KOH: about one half of KOH for (100)/(111) and (110)/(111) as listed in **Table 2**. TMAH is more expensive than KOH. When the solution is exposed to air containing CO_2 for a long period, the etching characteristics deteriorate.

In spite of the above drawbacks, TMAH has reasons for being accepted by industries. TMAH does not etch nitride mask at all. It attacks SiO_2 mask at an extremely low rate of 15 nm h^{-1} at 90°C (Tabata *et al.* 1991). It does not contain any harmful species to electronics circuits. SiO_2 mask can be easily used for deep silicon etching. One can also design fabrication processes including TMAH etching as an IC-compatible process.

1.07.4.1.1.(iii) Ethylenediamine and others In the early period of micromachining, EDP was widely used. It shows high etching ratios among orientations. A big advantage over KOH solution is that EDP is compatible with IC fabrication process. However, TMAH, EDP, which is also compatible with IC, is used in the micromachining process; this advantage became an absolute reason for selecting this etchant. It is toxic to the human body, causing cancer. Hydrazine is also toxic and inflammable.

1.07.4.1.2 Etching shapes

Etching shapes are categorized into two: one is concave-etched profiles and the other is convex-etched profiles.

Concave-etched profiles on silicon wafers have been fabricated so far for applications in pressure sensors and ink-jet printer head structures. Examples are shown in **Figure 9**. In the case of concave profiles, orientations having an etch rate that is locally minimum appear. Si (111) is the only orientation that has an etch rate of a local minimum for any etching conditions. One can find Si (111) as sidewalls of deep groove on Si (110), and tapered sidewalls on Si (100) in the figure. Depending on the etching conditions, other orientations sometimes show a locally minimum etch rate among other orientations. Si (100) often shows a local minimum for KOH and TMAH water solutions, but does not in cases of EDP, or KOH with isopropanol, or TMAH with a surfactant. In summary, examples of an etched profile obtained by anisotropic etching on an Si (100) wafer are schematically shown in **Figure 10**. The examples of an etched profile obtained on Si (110) wafer are also shown in **Figure 11**. Note that the etching profile is different for different etching solutions. This means that characterization of anisotropy

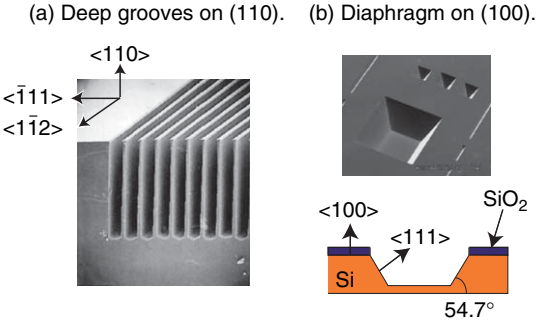


Figure 9 Examples of anisotropically etched microstructures on single-crystal silicon substrates.

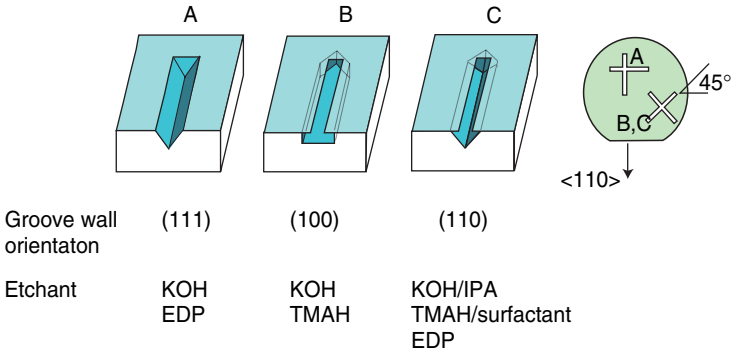


Figure 10 Emerging orientations appearing during groove etching of Si (100).

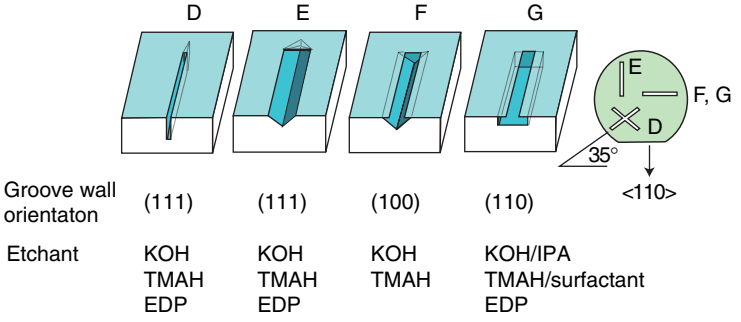


Figure 11 Emerging orientations appearing during groove etching of Si (110).

in etch rate is of importance for fabricating microstructures using wet anisotropic etching.

On the other hand, convex-etched profiles have been introduced for applications as shown in Figure 12. This is an array of silicon needles etched on a Si (100) wafer for the purpose of transdermal drug delivery. The needle profile is defined by fast etching orientations using a KOH solution, i.e., by (110) and others. In case of convex profiles, orientations having an etch rate that is locally maximum appear. It becomes far more difficult to predict and to control etch profiles having large etch rates. Again, characterization of anisotropy in etch rate is of great importance.

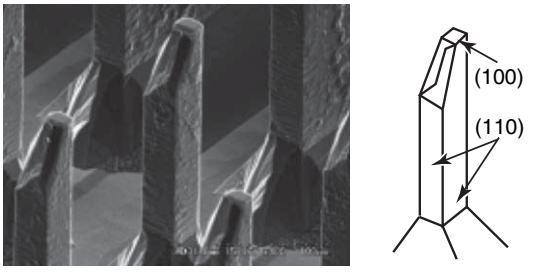


Figure 12 Fast etching orientations emerging as a convex etching profile in case of needle fabrication on Si (100). (Source: Cheng D, Gosálvez M A, Hori T, Sato K, Shikida M 2006 Improvement in smoothness of anisotropically etched silicon surfaces: Effects of surfactant and TMAH concentrations. *Sens. Actuators*, **A125**, 415–21.)

1.07.4.1.3 Masking materials

Most popular etching mask materials for alkaline solutions such as KOH and TMAH are SiO_2, Si_3N_4, and SiN. These materials are deposited on a silicon substrate by various deposition methods. It is required that the deposited films be free from pinholes. Thermal oxidation processes for SiO_2, and high-temperature CVD process for Si_3N_4, are recommended from the above requirements. Low-temperature physical vapor deposition (PVD) and plasma-enhanced chemical vapor deposition (CVD) sometimes cause trouble because the deposited film is not durable against an etchant.

In the case of KOH etching, the etch rate of SiO_2 is not negligible as shown in Figure 13, even if it is a thermally grown oxide. The etching ratio of Si to SiO_2 is about 150 at 70°C. This value is not as high as that of Si to Si_3N_4 or SiN. Deep silicon etching, such as when penetrating whole-wafer thickness about 400 μm, requires a thick layer of SiO_2 of about

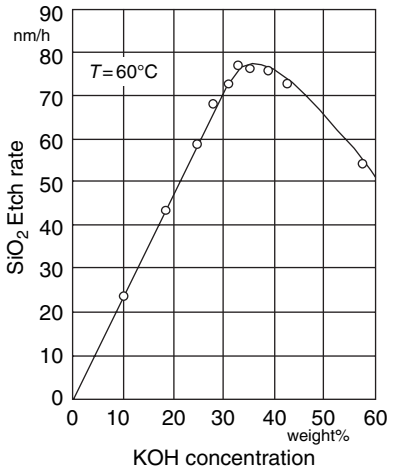


Figure 13 Etch rate of SiO_2 at 60°C as a function of KOH concentration. (Source: Seidel H, Csepregi L, Heuberger A, Baumgärtel H 1990 Anisotropic etching of crystalline silicon in alkaline solutions, I. Orientation dependence and behavior of passivation layers. *J. Electrochem. Soc.* **137-11**, 3612–26.)

2.5–3 μm. This is a disadvantage of SiO_2 – but advantages are easy to grow by thermal oxidation and easy to remove using HF-based solutions. On the other hand, Si_3N_4 is superior in terms of etching selectivity, but is not easy to remove. For the removal of Si_3N_4, a hot H_3PO_4 solution at an elevated temperature of 160–180°C is required. Silicon oxide or some negative photoresists are durable for such harsh etching conditions.

In case of TMAH etching, both SiO_2 and Si_3N_4 are excellent materials in terms of etching selectivity (Tabata *et al.* 1991). Thus, SiO_2 is recommended as a mask material for TMAH etching, because of the easiness of its deposition and removal.

Metallic films such as Ni and Pt are also durable against alkaline etching. However, it is not strongly recommended that metallic masks be used, because metallic ions in alkaline solutions often deteriorate etching qualities such as surface roughness and etch rates. Change in surface roughness as well as in etch rate caused by metallic contamination on the order of ppb level was recently reported by Tanaka *et al.* (2003).

1.07.4.1.4 Anisotropy as a function of orientations

Some work has been reported on the characterization of the orientation dependence in the etching of single-crystal materials. Ueda *et al.* (1986) measured the decrease in the thickness of quartz wafers having different surface orientations. This was the most direct method of evaluating anisotropy in etching. However, a dense net of data as a function of orientation is hard to accumulate because of the numerous experiments required for a single etching condition. Weirauch (1975) measured the undercut length distribution around a circular mask pattern made on a silicon wafer. Seidel *et al.* (1990) and Kendall *et al.* (1982) also measured the undercut length, not of a circular mask but of a sharp vertex of triangular mask pattern. Because a number of triangular mask patterns are densely located in a rotational position with respect to each other on the same wafer, this type of measurement was called the wagon-wheel method. An etched example of the wagon-wheel is shown in **Figure 14**. Though these mask undercut measurements were effective for qualitative visualization of anisotropy, the obtained data did not necessarily show correct etch rates, because the side-etched planes were not always perpendicular to the wafer surface. Measurements of the tilting angle of the wall were necessary for quantitative characterization, as shown in **Figure 15**.

The use of a hemispherical single-crystal silicon specimen, as shown in **Figure 16**, was proposed for the measurement of anisotropy in etch rate by Sato *et al.* (1998). Crystallographic orientations appear on the hemispherical surface, which is mechanically ground, lapped, and polished into a mirror. Etch rates for any orientation can be calculated from the dimensional change before and after etching of the spherical surface. The radius of the hemisphere is 22 mm, which is large enough when compared to the etch depth. This allows interference among

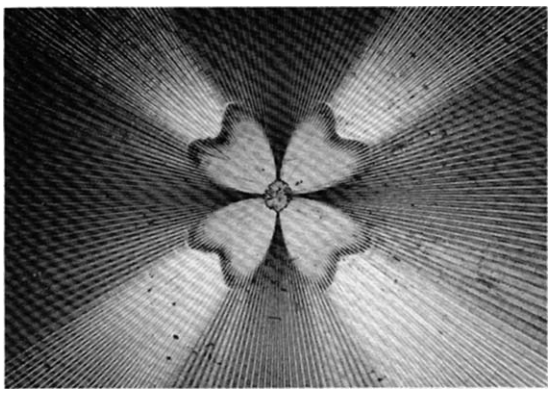


Figure 14 Enlarged top view of the wagon-wheel pattern after etching for visualizing anisotropy as a planer distribution. (Source: Seidel H, Csepregi L, Heuberger A, Baumgärtel H 1990 Anisotropic etching of crystalline silicon in alkaline solutions, I. Orientation dependence and behavior of passivation layers. *J. Electrochem. Soc.* **137-11**, 3612–26.)

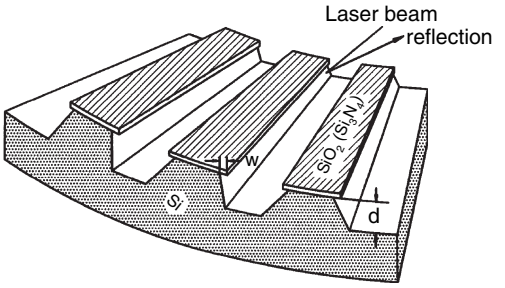


Figure 15 Schematic view of the etched profile under the wagon-wheel mask pattern. (Source: Seidel H, Csepregi L, Heuberger A, Baumgärtel H 1990 Anisotropic etching of crystalline silicon in alkaline solutions, I. Orientation dependence and behavior of passivation layers. *J. Electrochem. Soc.* **137-11**, 3612–26.)

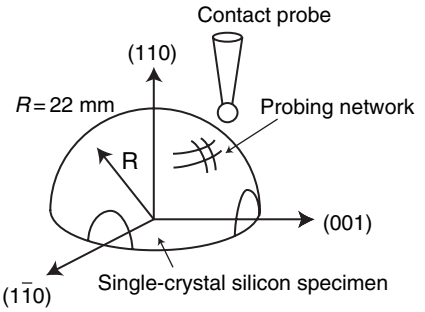


Figure 16 Hemispherical single-crystal specimen for the measurement of etch rate as a function of orientation. (Source: Sato K, Shikida M, Matsushima Y, Yamashiro T, Asaumi K, Iriye Y, Yamamoto M 1998 Characterization of orientation dependent etching properties of single crystal silicon: Effects of KOH concentration. *Sens. Actuators A* **64–1**, 87–93.)

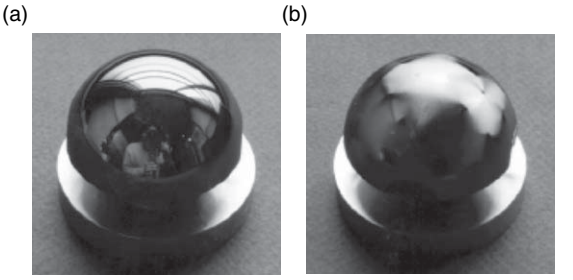


Figure 17 Hemispherical single-crystal silicon specimen (a) before and (b) after etching. (Source: Sato K, Shikida M, Matsushima Y, Yamashiro T, Asaumi K, Iriye Y, Yamamoto M 1998 Characterization of orientation dependent etching properties of single crystal silicon: Effects of KOH concentration. *Sens. Actuators A* **64–1**, 87–93.)

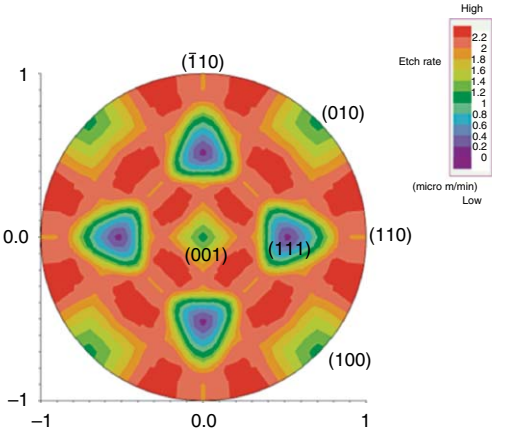


Figure 18 An etch rate contour map as a function of orientation for 40% KOH (34wt.% KOH) at 70°C.

orientations, i.e., the disappearance of some orientations from the surface is negligible.

A comparison of hemispherical specimen surfaces before and after etching is given in **Figure 17**. Though the maximum etching depth was about 150 µm, one can observe anisotropy with naked eyes, i.e., three-time symmetry around (111) and four-time symmetry centered by (100) are visible on the specimen. An etch rate contour map for 40% KOH (34wt.% KOH; commercially available KOH grains contain 85wt.% KOH and 15wt.% water of crystallization) at 70°C is shown in **Figure 18**. The contour map is a projection of the hemispherical specimen surface whose top is orientated to (100).

It is clear that a deep minimum in etch rate is located at (111), shallow local minimum at (100), and relatively high values at (110). There are also saddle points on the map.

However, the contour map pattern is not consistent among different etching conditions. The measured etch rates as a function of orientation are listed in **Table 2** for KOH and TMAH water solutions. The concentrations and etching temperatures are selected so that the etch rate of Si (100) becomes almost equal for KOH and TMAH.

The etch rate contour maps for KOH and TMAH are compared in **Figure 19**. It can be observed that for both etching solutions the local minima in etch rate exist at (111) and also at (100), though the valley of (111) is much deeper than that of (100). The difference is that the triangular contour patterns around (111) are rotated by 60° between KOH and TMAH. This clearly means that anisotropy is not only uniquely determined by silicon crystal lattice structure, but also by the interactions between chemicals in solution and silicon surface.

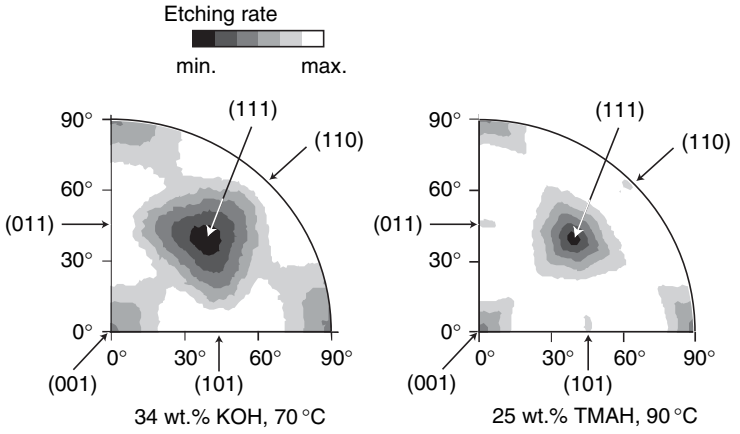


Figure 19 Comparison of etch rate contour map between KOH and TMAH. (Source: Sato K, Masuda T, Shikida M 2003 Difference in activated atomic steps on (111) silicon surface during KOH and TMAH etching. *Sens. Mater.* **15-2**, 93–9.)

1.07.4.1.5 Technologies for controlling etching shapes

There are various recipes that can provide different etching results. When a micromechanical structure is designed, etching profiles can be controlled by introducing the following technologies.

1.07.4.1.5.(i) Etching species and concentrations Etching shape becomes different when the etching solution, i.e., the etching species, and concentrations differ. KOH and EDP showed different groove profiles from each other, as shown in **Figures 10** and **11**.

Deep groove profiles etched on a Si (110) wafer differ also between TMAH and KOH, as graphically analyzed using Wulff–Jaccodine's construction (Jaccodine 1962), as shown in **Figure 20**. The analysis is based on the principle that among a number of

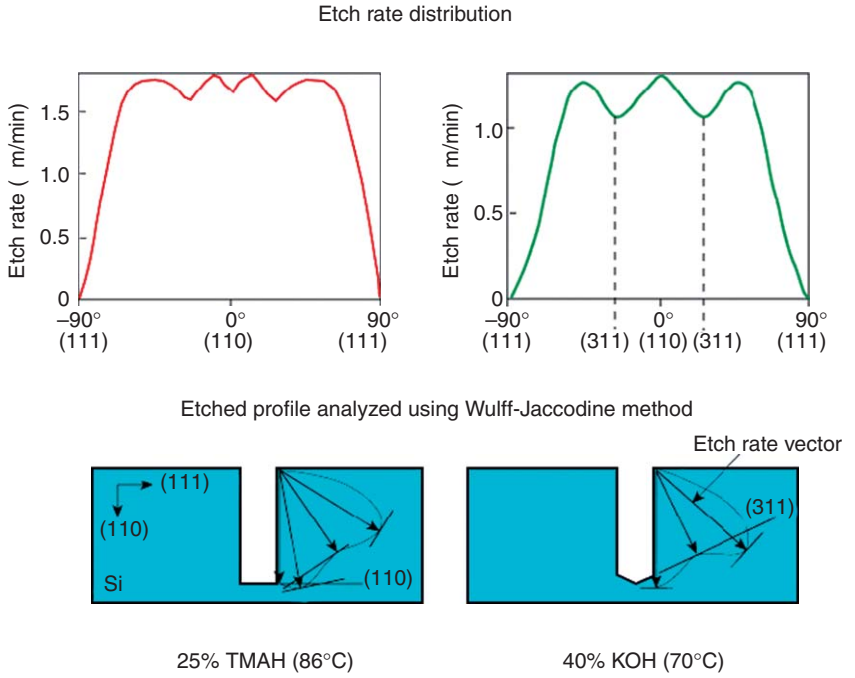


Figure 20 Comparison of etch profiles of deep grooves on (110) etched by using KOH and TMAH, respectively. (Source: Sato K, Shikida M, Matsushima Y, Yamashiro T, Asaumi K, Iriye Y, Yamamoto M 1998 Characterization of orientation dependent etching properties of single crystal silicon: Effects of KOH concentration. *Sens. Actuators A* **64–1**, 87–93.)

potential etching profiles, drawn perpendicular to the etch rate vectors, only a set of profiles that minimize the total surface energy composes a real profile. The results show that the bottom surface of the groove is flat for TMAH and wedge-shaped for KOH. This is because the etch rate vector distribution in the groove cross section differs between two etchants. In case of KOH, a local minimum in etch rate at (311) is deeper than that of TMAH. Thus the (311) orientation appears as the groove profile with KOH, whereas (110) appears with TMAH.

Similarly, a change in solution concentration also results in different etch profiles. Only by the difference in KOH concentration, the groove bottom profile becomes different, as shown in **Figure 21**. The reason is again that the etch rate contour map differs by the difference in KOH concentration (Sato *et al.* 1998). In such cases, it is important to select etching conditions to control etched profiles. An etch rate contour map is necessary for predicting etched profiles.

1.07.4.1.5.(ii) Effects of additives in etching solutions Some additives are used aimed at the following two effects.

(1) Suppressing mask corner undercut without using compensation mask patterns.
(2) Obtaining a smooth etched surface on (110).

It is known that some types of alcohol such as isopropyl alcohol (IPA) added to KOH water solution are effective in minimizing mask corner undercut on (100) silicon. It is also effective in reducing etched surface roughness of the (110) surface. As the (110) plane appears as a result of (100) wafer etching and makes an angle of 45° from the wafer surface, an addition of IPA to the KOH solution was proposed for fabricating a 45° mirror for the purpose of reflecting a light coming out from an optical fiber on the wafer in a direction perpendicular to the wafer surface (Bäcklund and Rosengren 1992).

In order to get a sufficient effect, IPA must be dissolved to a saturation level in KOH water solution. On the other hand, some surfactant added to TMAH shows similar effects, even if the concentration is as low as a few percent. Sekimura (1999) reported the striking effects of a small amount of surfactant polyoxiethylene-alkyl-phenyl-ether (commercial code name is NCW, a product of LION Ltd.) added to the TMAH solution.

Figure 22 shows the effects of surfactant added to TMAH (Sato *et al.* 2001). The rectangular mask corner has been severely undercut by pure TMAH solution, whereas the corner was little etched by TMAH with a surfactant NCW. The undercut at the mask corner depends on the relative etch rate of (110) among other orientations. As shown schematically in **Figure 22** (*top*), the ridge developing from the mask corner is the intersection of two (111) planes. The ridge is on the (110) plane, or more precisely on ($j\,j\,1$), where j is an integer. When the etch rate of (110) or ($j\,j\,1$) decreases when compared to that among other orientations, the mask corner undercut is suppressed. This happens by the addition of NCW, as shown in **Figure 23**. The advantage of surfactant/TMAH over IPA/KOH is that the amount of the additive being effective is as low as

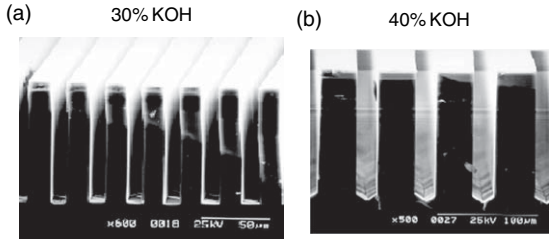


Figure 21 Influence of KOH concentration on the deep groove profile on (110). (Source: Sato K, Shikida M, Matsushima Y, Yamashiro T, Asaumi K, Iriye Y, Yamamoto M 1998 Characterization of orientation dependent etching properties of single crystal silicon: Effects of KOH concentration. *Sens. Actuators A* **64–1**, 87–93.)

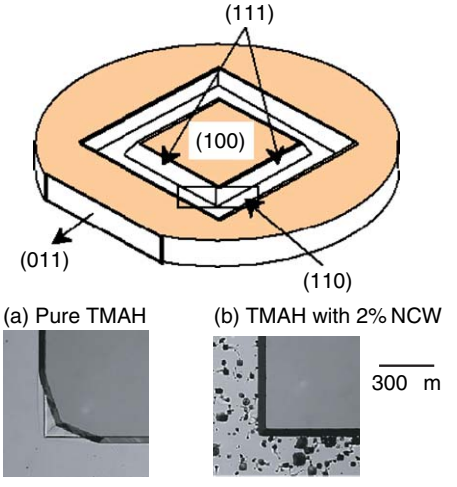


Figure 22 Effects of surfactant NCW added to TMAH for suppressing rectangular mask corner undercut. (Source: Sato K, Uchikawa D, Shikida M 2001 Change in orientation-dependent etching properties of single-crystal silicon caused by a surfactant added to TMAH solution. *Sens. Mater.* **13-5**, 285–91.)

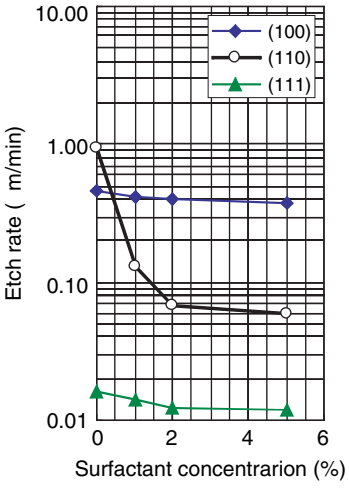


Figure 23 Change in relative etch rate of (110) among other orientations by the addition of surfactant NCW to TMAH. (Source: Sato K, Uchikawa D, Shikida M 2001 Change in orientation-dependent etching properties of single-crystal silicon caused by a surfactant added to TMAH solution. *Sens. Mater.* **13-5**, 285–91.)

2% volume fraction. However, the surfactant does not work as well with KOH as it does with TMAH.

1.07.4.1.5.(iii) Compensation mask patterns

Mask corner undercut during etching of mesa structure on Si (100) is often a problem as mentioned above, and efforts have been done to minimize it. One of the solutions is to use a mask having additional masked areas that can delay the corner undercut. The examples of mask corner compensation pattern are shown in **Figure 24**. One is a square pattern (**Figure 24(a)**) whose center is located at the corner of the original mask, and another is an oblique pattern (**Figure 24(b)**) whose sides are angled at 45° from the <110> directions.

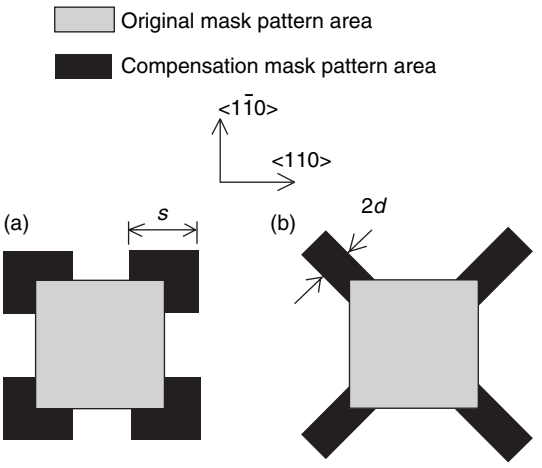


Figure 24 Examples of compensation mask patterns for minimizing mask corner undercut for Si (100) etching: (a) square and (b) 45° oblique.

With a square compensation pattern (**Figure 24(a)**), based on time increments, the underetch advances from the periphery of the small square pattern to the center. Until the etch-front reaches the center of the square, the original mask corner survives. The etch profile change under the square compensation pattern is shown in **Figure 25**. The size of the square pattern s in **Figure 24(a)** needs to be designed based on the etching simulation shown in **Figure 25**, or by experimental trials. Another disadvantage is that the bottom periphery of the etched mesa does not become rectangular, though the masked top of the mesa is perfectly rectangular as shown in **Figure 25**.

In case of 45° oblique pattern (**Figure 25(b)**), one has to use pure KOH or TMAH that has local minimum in etch rate at (100) orientation. One should not use additives that make the local minimum at (110). The reason is that the oblique pattern aims at the etched sidewall along oblique sides to become perpendicular to the wafer surface. Because the sidewalls are oriented to (001), the side etching (undercut) from the oblique sides advance with the same rate as that normal to the wafer surface. The width of the oblique bar should be designed as $2d$, where d is the

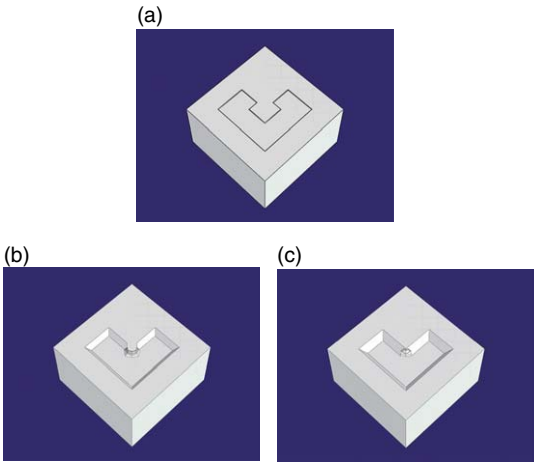


Figure 25 Effects of a square compensation mask pattern suppressing mask corner undercut. (a) Initial mask pattern, (b), (c) etched profiles as time sequence. Simulation results using MICROCAD (FabMeister-ES). (Source: Sato K, Asaumi K, Kobayashi G, Iriye Y, Shikida M 2000 Development of an orientation-dependent anisotropic etching simulation system MICROCAD. *Electron Comm. Jpn. Part 2* **83–4**, 13–22.)

etch depth. When the etch depth reaches the value d, the compensation pattern entirely disappears. An oblique pattern has advantages such as a simple design rule and an excellent finished etch profile. One disadvantage is that when the compensation pattern has to be connected to the neighboring mask pattern because of the lack of space around the island mask pattern, the compensation structure does not disappear easily from the wafer surface. One should carefully estimate the residual structure size in such a case.

1.07.4.1.5.(iv) Multisteps of etching using different mask patterns A single step of etching provides only simple etch profiles composed of (111), (100), or (110), which are orientations existing at a local minimum on a etch rate contour map. When the etching is interrupted and the masked area is replaced with a pattern different from the previous one, different etched profiles become available.

The example is a two-step etching fabricating a round profile at the base of a cantilever on a (100) wafer. The process is shown schematically in Figure 26. The first etching makes a preform of the cantilever much shorter than the final length. The second etching is applied with a mask for the final length of the cantilever structure. When a step profile made by the first etching is totally exposed to etching, the step profile changes during the second etching until it disappears. Just before the disappearance of the step, one can obtain a smooth round profile at the base. The design issues are first and second etch depth and offset length between the first and the second mask. This should be done by an etching simulation system, as described in the following section.

By applying this process, a cantilever beam having a round profile at the base was fabricated, as shown in Figure 27. The round profile was formed both at the top and at both sides of the cantilever base (Koide *et al.* 1992).

It is of course difficult to apply a photolithography process once the wafer surface becomes nonflat by the prior etching process. In order to avoid the difficulty in applying the photoprocess, in particular, the resist coating process, a layered mask process is recommended. Different mask patterns are layered by using different mask materials before silicon etching. When the first etching is finished using the first mask pattern, only the first layer is removed. Then the second mask pattern emerges on the wafer surface, and the second etching is ready to start. The process limitation is that the mask opening pattern must be larger than that of the mask used in the prior etching process to avoid interference between mask patterns. Mask materials should be preferably different from the viewpoint of etching selectivity. However, even with the same material like SiO_2, one can create a layered mask structure by varying the thickness. In the case of the cantilever in Figure 27, three layers of masks of the same material SiO_2 were used for fabricating an acceleration sensor chip (Koide *et al.* 1992).

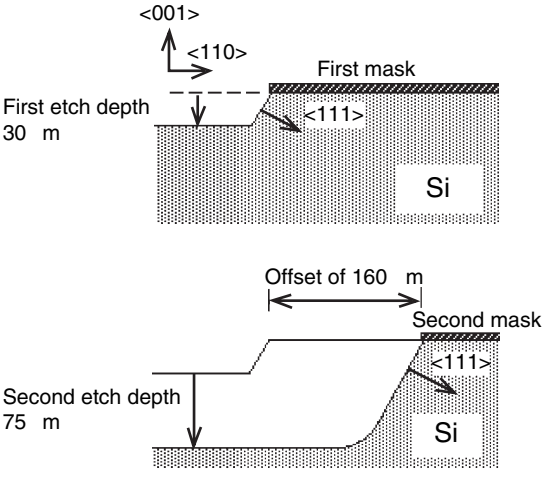


Figure 26 Two-steps etching process fabricating a round profile at the base of a structure extending to <110> direction on a Si (001) wafer.

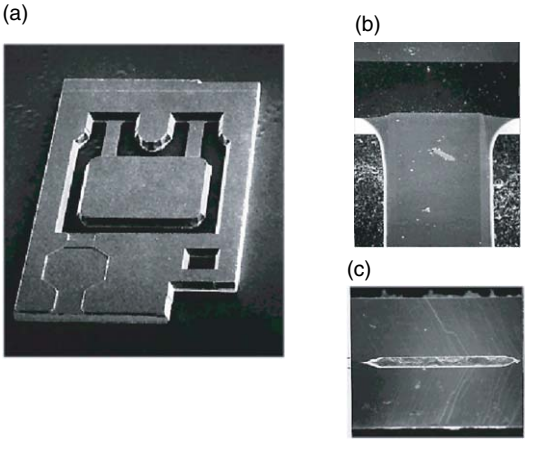


Figure 27 Acceleration sensor structure having a round profile at the base of a cantilever fabricated by using three steps of KOH etching. (a) Total view of the chip, (b) an enlarged view of the cantilever base top, (c) cross section of the beam 15 μm thick.

1.07.4.1.5.(v) Etching from both sides of the wafer Etching profiles show more variations, when the etching is applied from both sides of the wafer. In the previous section, such a process was applied for the acceleration sensor chip in **Figure 27**. Once the etching fronts meet in the middle of the wafer, etching can easily advance in lateral directions as well as in the thickness direction. Thus a 3D microstructure can be fabricated. More examples will be presented in Section 1.07.4.2.

1.07.4.1.5.(vi) Etch Stop techniques An etch stop technique is often required for controlling the thickness of etched structures. The simplest way of controlling etched thickness is by controlling the etching time. This is possible because of the high uniformity in etching depth due to the reaction-limited etching process, in particular, by using KOH solutions under good temperature control in the etching bath. Etch depth variation over the wafer is usually less than 1 μm after hundreds of microns of deep etching. This is an advantage over deep dry etching. However, because of the nonuniformity in initial wafer thickness depending on the wafer suppliers, the resulting membrane thickness may become less uniform than the etch depth.

Besides etching time control, some etch stop techniques are used for the control of the etched depth or etched membrane thickness. They are (a) p^{++} etch stop, (b) electrochemical etch stop using p–n junction, and (c) silicon-on-insulator (SOI) wafer. They are explained in the following section.

1.07.4.1.5.(vi).(a) p^{++} etch stop When silicon is heavily doped with boron to a concentration in the range of 10^{19} (cm^{-3}) or more, the etch rate decreases drastically. The etch rate as a function of doping concentration is shown in **Figure 28** (Seidel *et al.* 1990). When the front side of a wafer is heavily doped with B, and etched from the rear side of the wafer, the p^{++} layer remains as a membrane or cantilever structure, as shown in **Figure 29**. The thickness of the highly doped layer is in the range of a few micrometers. This was utilized from the early stage of silicon micromachining. However, the application was limited because of the following drawbacks.

When an impurity like B is highly doped in the range mentioned above, silicon is no more a semiconductor but becomes a conductor. Electronics circuits cannot be integrated on the same surface. This is a disadvantage for MEMS applications, where integration of mechanical and electrical functions is expected.

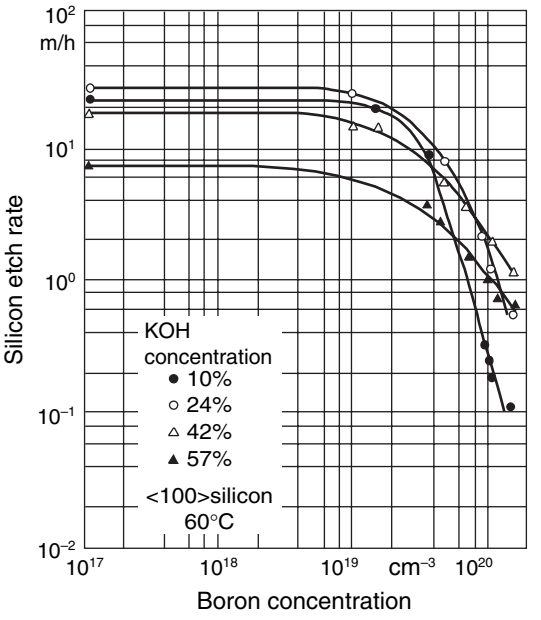


Figure 28 Influence of B-doping concentration on the etch rate of Si (100) using KOH solutions. (Source: Seidel H, Csepregi L, Heuberger A, Baumgärtel H 1990 Anisotropic etching of crystalline silicon in alkaline solutions, II. Influence of dopants. *J. Electrochem. Soc.* **137-11**, 3626–32.)

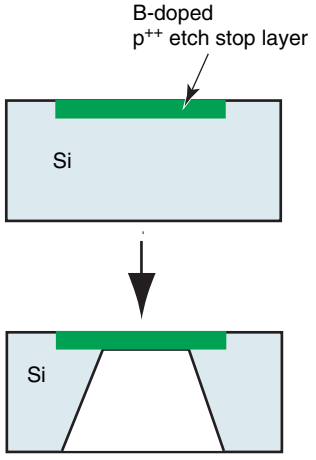


Figure 29 p^{++} etch stop technique applied to fabricate a membrane on a wafer surface.

Highly doped silicon shows high residual stresses. In the case of boron, the residual stress is in the tensile region. Microstructures having internal stresses distributed in the thickness direction show a significant amount of bending or warping deformation. Minimization of the internal stress is thus of

importance. It is reported that germanium as an impurity also shows etch stop properties, whereas the residual stress is compressive. A combination of boron and germanium doping is reported to achieve etch stop and low residual stress at the same time (Seidel *et al.* 1990).

Controllability of an etch stop with p^{++} heavy doping is not perfect. When the doping concentration profile in the depth direction is not sharply defined, e.g., by the annealing process after doping, the thickness of the etch stop layer is not clearly defined.

1.07.4.1.5.(vi).(b) Electrochemical etch stop using p–n junction Contrary to the drawbacks of the p^{++} etch stop technique, the electrochemical method allows a perfect etch stop with an accuracy of about 0.1 μm in thickness. The principle of this technique is as follows. When a wafer is etched in an alkaline solution the electric potential of the wafer is balanced in a range of −1.5 to −1.1 V depending on the dopant type of silicon, illumination, etc. This potential is called open-circuit potential (OCP). When a positive electric potential is applied to a silicon wafer that is higher than that of the etching solution by 2–3 V, silicon oxide is generated on the surface and etching stops. To form the oxide film, electric current is required to flow from the wafer to the solution. For the purpose of achieving the etch stop, p-type substrate having an n-epitaxial layer is placed in the etching solution so that the etching advances from the p-type side, as shown in **Figure 30**. The p-substrate is etched until the etching front reaches the p–n junction, because electric current cannot flow to the solution because of the existence of the p–n junction. Once the etching front meets the n-layer, current start to flow and passivation is completed. The membrane thickness is determined by the n-type layer that can be well controlled by the epitaxial growth process (Kloeck *et al.* 1989).

In order to perform this technique for production, the following practical techniques are needed as engineering skills. Leakage current at the wafer periphery must be strictly avoided. It will cause an etch stop at an earlier stage of etching than expected. Good surface finish and protection from electric shortage are required at the periphery of the wafer, where the p–n junction is exposed to air. Also, the potential of the p-doped substrate should preferably be monitored and controlled for the stabilization of the process. Making an electrical contact with one of the silicon layers needs a specially designed wafer holder equipped with a good sealing against etchant and must be easy to operate.

By applying this technique, Epson has developed an ink-jet head having densely arrayed membrane pumps. They fabricated a Si membrane, 2.15 μm thick, with an accuracy of ±0.35 μm in production, where the deviation 3σ of the thickness over a wafer was reported as only 0.35 μm (Kamisuki *et al.* 2000). The cross-sectional view of the membrane is shown in **Figure 31**. The substrate was Si (110), allowing deep etching for making a dense array of ink reservoir chambers.

1.07.4.1.5.(vi).(c) SOI wafer SOI wafers have been introduced rather recently to the MEMS community due to the increase in the supply boosted by the large scale integration (LSI) applications.

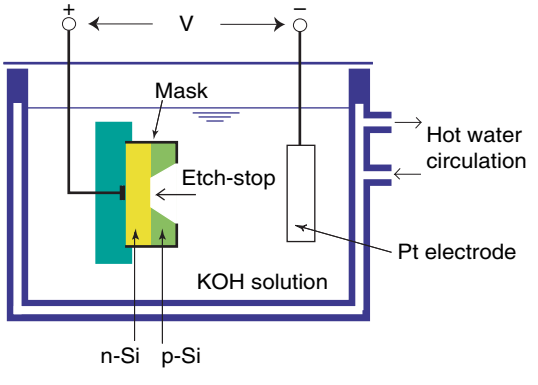


Figure 30 Electrochemical etch stop technique using p–n junction.

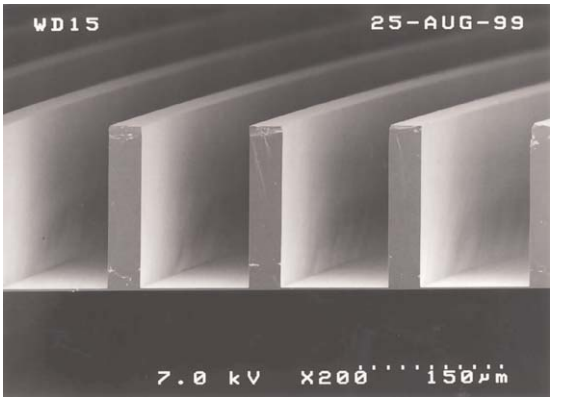


Figure 31 Membrane structure 2.15 μm thick etched on Si (110) fabricated by using electrochemical etch stop technique. It is used in a commercial ink-jet printer head as a membrane pump. Courtesy of SEIKO EPSON Ltd. (Source: Kamisuki S, Fujii M, Takekoshi T, Tezuka C, Atobe M 2000 A high resolution electrostatically-driven commercial inkjet head. *Tech. Dig. 13th IEEE Int. Conf. on Micro Electro Mechanical Systems*, Miyazaki, pp. 793–8.)

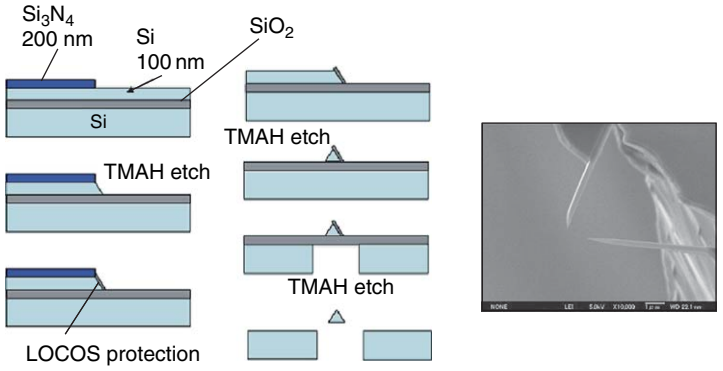


Figure 32 Twin probe of 200 nm diameter fabricated on a silicon-on-insulator (SOI) wafer. (Source: Kakushima K, Mita M, Kobayashi D, Hashiguchi G, Endo J, Wada Y, Fujita H 2001 Micromachined tools for nano technology – Twin nano-probes and nano-scale gap control by integrated microactuators. *Tech. Dig. 14th IEEE Int. Conf. on Micro Electro Mechanical Systems*, Interlaken, pp. 294–7.)

The wafer has a layered structure, i.e., single-crystal silicon layer, buried silicon oxide layer, and silicon substrate.

When etching is applied from the rear side of the substrate, etching stops at the buried oxide layer. Thus the micromechanical structures in the top silicon layer can be easily released from the substrate, after removal of the oxide layer.

Figure 32 shows a fabrication process to make a twin probe on a SOI wafer. The top layer is 200-nm-thick silicon. Probe tips are formed by applying two steps of TMAH etching from the front side. The structure is then released from the substrate by etching applied from the rear side. Thus a very thin probe 200 nm thick was fabricated (Kakushima *et al.* 2001).

1.07.4.1.5.(vii) Combination of wet etching with other processes Limitation in the etched profile using wet process is partly caused by a slow etching orientation, i.e., (111) emerging in the etching holes. This is the reason why a deep narrow hole cannot be etched on Si (110), while a narrow deep groove is possible to etch on (110). If the (111) or any other slow etching orientation is destroyed prior to etching, the limitation in the etched profile can be removed to some extent.

Büttgenbach and coworkers (Alavi *et al.* 1991) proposed the use of laser, and Shikida *et al.* (2004) proposed mechanical machining both for preprocessing the wafer before etching.

1.07.4.1.5.(vii).(a) Combination with laser beam processing Laser beam irradiated on the silicon wafer can remove the material by melting. However, the crystalline structure can be destructed with less energy than that needed for melting. By irradiating a laser beam in an area of mask aperture, single-crystal silicon structure can be destructed to some depth in the colored region, as shown in Figure 33. The wet etching is then applied as usual. The etched results are schematically shown in the figure. The colored regions are easily etched. Thus microstructures can be fabricated deep into the wafer. Note that the etched profiles are well defined by the mask aperture pattern width w, and not by the diameter of the spot size of the laser-influenced area. Also the laser-destructed regions are entirely removed after etching.

1.07.4.1.5.(vii).(b) Combination with mechanical dicing Conventional micromachining for fabricating silicon microstructures basically employs the following three technologies, mask material deposition, patterning the mask materials to define the etching area, and etching to form 3D structures. Such approach has the following two problems:

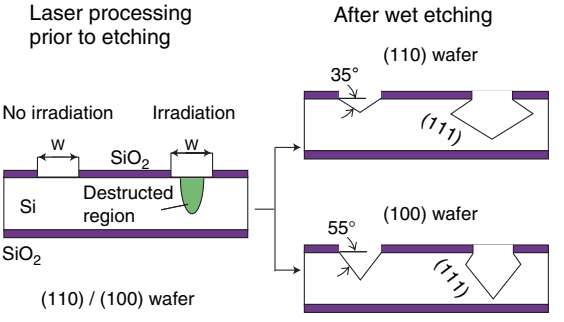


Figure 33 Combined process of laser processing and wet etching.

(1) Requirement of expensive equipment – The process needs an expensive photolithography machine, which is one of the most expensive equipments in MEMS fabrication, to be used to define the etching area on a Si wafer surface. It also requires that an expensive photomask be used in designing the microstructures. The number of masks required depends on what kind of devices MEMS researchers design from Si wafers. At least a few masks are required to fabricate 3D structures.
(2) Difficulty of high aspect ratio microstructure – The two factors, the crystal orientation of the Si wafer and the etch rate diagram, dominate the etched shape in the anisotropic wet etching process (Shikida *et al.* 2002). The wet etching process is intrinsically hard to fabricate high aspect ratio microstructures, because the etched shapes are determined by the crystal orientations of the Si wafer.

To overcome these problems, a different type of approach, which consists of mechanical diamond grinding with a dicing saw and anisotropic wet etching, in microstructure fabrications is proposed (Shikida *et al.* 2004a, b, 2006a, b). Mechanical dicing is used to form a 2.5D structure into a Si wafer instead of using the photolithography process, and anisotropic wet etching is then applied to transform it into 3D structures. The examples of biomedical applications to the microneedle formation are shown in the following.

The process is able to fabricate various shaped needle structures, for example pyramidal-, candle-, and spike-shaped needles, by repeating the use of the dicing and the wet etching process (**Figure 34**). The dicing process basically determines the height and density of the arrayed needle structures, and the succeeding wet etching forms the final needle shape. The process sequences of the needle formations are as follows:

(a) Pyramidal-shaped needle formation – The etching mask is formed on both sides of the silicon wafer, and then grooves on the wafer are made with a dicing saw. The wet etching process is then applied to the diced silicon wafer. The etching proceeds in the direction of the side and bottom of the grooves. The plane located at the saddle point on the etch rate diagram appears during undercutting and defines the etched shapes. As a result, a pyramid-shaped needle is formed by the wet etching process.
(b) Candle-like needle formation – A candle-like needle structure is formed by applying the second dicing process.
(c) Spike-shaped needle formation – A concave shape is formed under the needle tip by the same undercutting principle, and a spike-shaped needle structure is formed during the wet etching process. The groove walls formed at a candle-like formation determine the final height of the spike-shaped needle and the undercutting area.

The proposed principle of combined process allows MEMS society fabricating a variety of shapes in microstructures even without applying expensive photolithographic equipment, by multiplying the number of mechanical dicing and wet etching processes.

1.07.4.1.6 Comparison of wet and dry etching

In the field of MEMS micromachining, wet and dry processes are both useful for fabrication. Pros and cons of those two approaches are compared in **Table 3**.

It is apparent that a dry process like RIE is superior to a wet process for versatile applications, because any 2D mask patterns can be deeply etched in the substrate. Also a dry process has less chance of contamination than a wet process. However, a wet process is advantageous in the following issues.

(1) Controllability and uniformity in etching depth.
(2) Potential for special types of etching profile.
(3) Equipments costs required are of magnitudes lower than for dry etching.

Point (1) is valid for both with and without etch stop layer techniques. Without an etch stop layer, (100) Si etch depth variation on a total wafer is within 1 μm for 200–300 μm deep etching using KOH by a time control combined with monitoring the depth during process. On the contrary, etch depth of deep RIE is not controllable depending on the mask opening size.

The electrochemical etch stop technique mentioned in Section 1.07.4.1.5.(vi) allows excellent control in the etched structure thickness. When the thickness of an epitaxial n-silicon layer is controlled in a range of ±0.05 μm, the variation in the thickness of the membrane obtained is in the same range. On the other hand, a buried oxide layer is used as an RIE etch stop layer of SOI wafer, the thickness variation depends on that of the top silicon layer of the SOI, which at the moment is not superior to the n-epitaxial layer mentioned above.

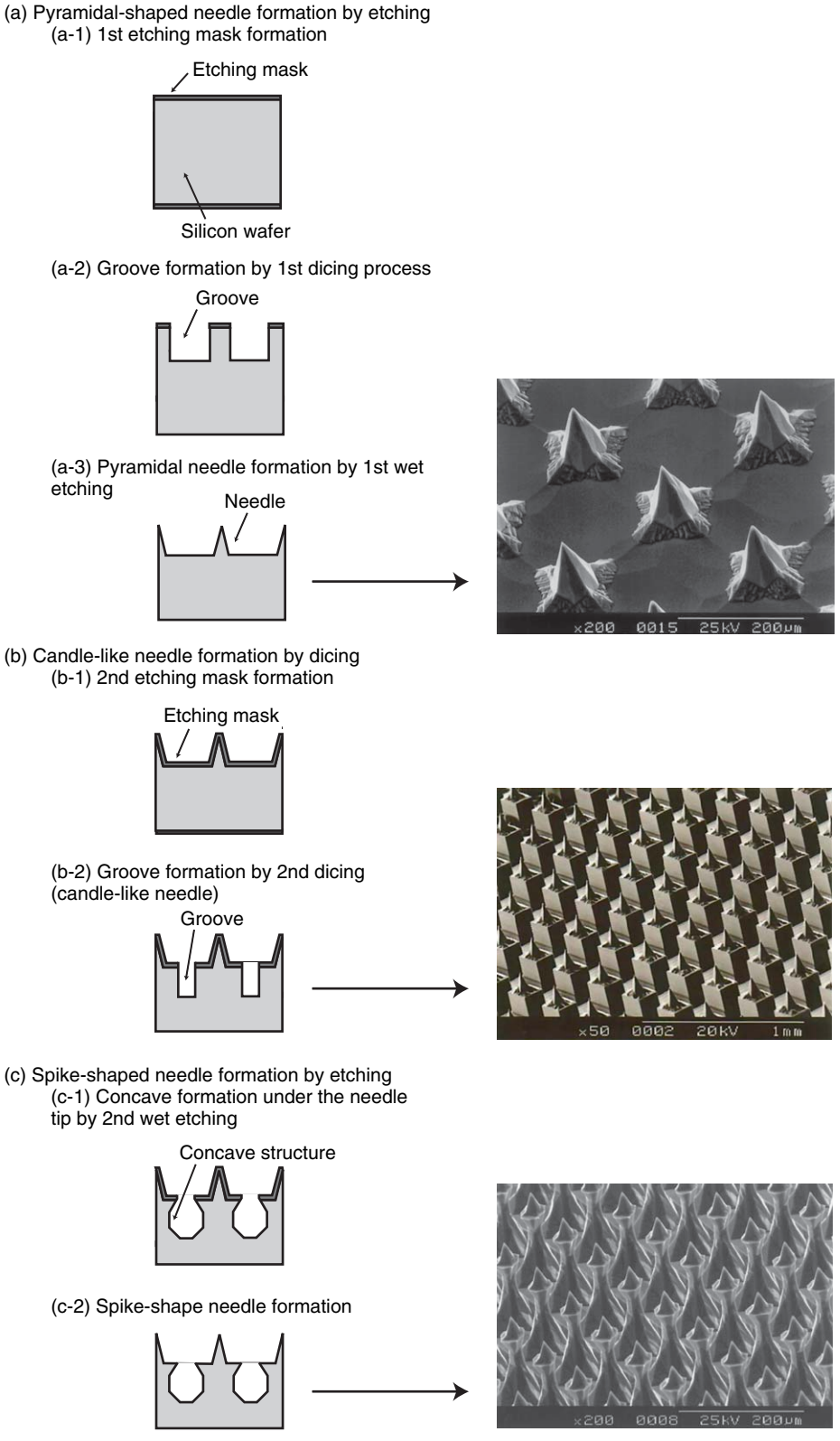


Figure 34 Fabrication of various needle structures by repeating dicing and wet etching processes. (Source: Shikida M, Odagaki M, Todoroki N, Ando M, Ishihara Y, Ando T, Sato K 2004a Non-photolithographic pattern transfer for fabricating arrayed three-dimensional microstructures by chemical anisotropic etching. *Sens. Actuators A***116**, 264–71.)

Table 3 Characteristics of wet and dry processes for anisotropic etching

	Wet etching	*Dry etching*
Fineness of the etched dimension	+	++
Cleanliness of the process, IC-compatible	−	++
Applicability for any 2D mask patterns	− − −	+++
Controllability in etching depth	++	+
Uniformity in etching depth	++	−
Potential for special types of etching profile	++	+
Equipments costs required	+++	− −

Plus (+), superior; minus (−), inferior.

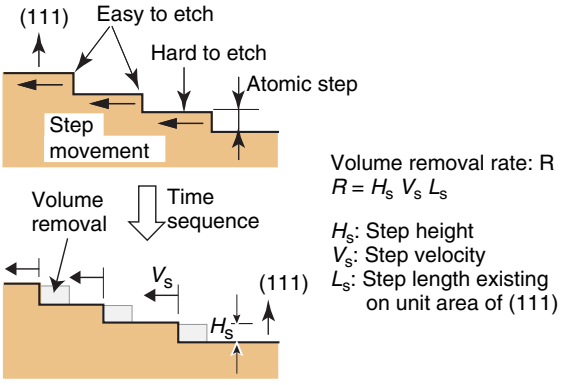


Figure 35 Step flow model for the etching in the vicinity of Si (111). Cross-sectional view of a wafer showing receding motion of atomic steps.

Point (2) includes the fact that an excellent flatness of the polyhedral etched profile can be obtained by the wet processes. Also, it includes a high repeatability in the product dimensions compared to that in dry processes.

1.07.4.2 Etching Model and Simulation Systems

1.07.4.2.1 Step flow model for variable anisotropy

It was discussed in the previous section that the anisotropy in etch rate varies based on a small change of etching conditions. However, the reasons are not clear because of a lack in knowledge on the etching mechanisms. The purpose of this section is to give an etching model for an alkaline solution etching system.

A conventional explanation for why (111) is stable against etching was based on the number of dangling bonds, as already explained using **Figures 7 and 8**. However, atom removal from perfectly flat (111) is a minor case actually occurring on a crystal surface. In most of the cases, a large number of atoms are removed from atomic steps and kinks. This is the weak point of the dangling bond model. The etch rate must be quantitatively discussed based on dynamic models.

This was first pointed out by Elwenspoek (1993), who proposed a step flow model as an etching mechanism of Si (111) and its vicinity. Atom removal starts with a nucleation of a pit and the propagation of atomic steps, as shown in **Figure 35**. Nucleation actually occurs from a perfectly flat (111). However, major atom removal occurs continuously at steps or kinks. This is visible resulting in the movement of steps. This was proved experimentally by Allongue *et al.* (1993) by *in situ* STM (scanning tunneling microscope) observations of Si (111) etching using NaOH solutions. Though anisotropic nature was explained by this model, changes in anisotropy by etching conditions, like differences between KOH and TMAH, have never been discussed in relation to the etching mechanism.

Let us consider why KOH and TMAH show a difference in the etch rate contour map pattern (see **Figure 19**) in the vicinity of (111). Though triangular contour lines surrounding (111) appeared for both etchants, the directions of vertexes of the triangles differed by 60° between the two. Sato *et al.* (2003) observed etch pit growth on Si (111) silicon, i.e., movement of a cluster of atomic steps, and found that the difference in etch rate contour map between KOH and TMAH was due to the difference in movement of atomic steps on Si (111), as discussed below.

1.07.4.2.1.(i).(a) Etch pit observation on Si (111) Sato *et al.* observed shallow etch pit growths on a (111) surface. They introduced defects on the wafer surface by a thermal treatment prior to etching. When the surface is exposed to an etchant, etch pits grew centered by the artificially introduced defects. A single pit was watched with time increments using an optical microscope and a 3D surface profiler. **Figure 36** shows top views of etch pits and those cross-sectional profiles taken by time increments. Etch pits for both KOH and TMAH are basically of hexagonal profiles, and three of the six sides of the hexagon were aligned to $[11\bar{2}]$ and the remaining

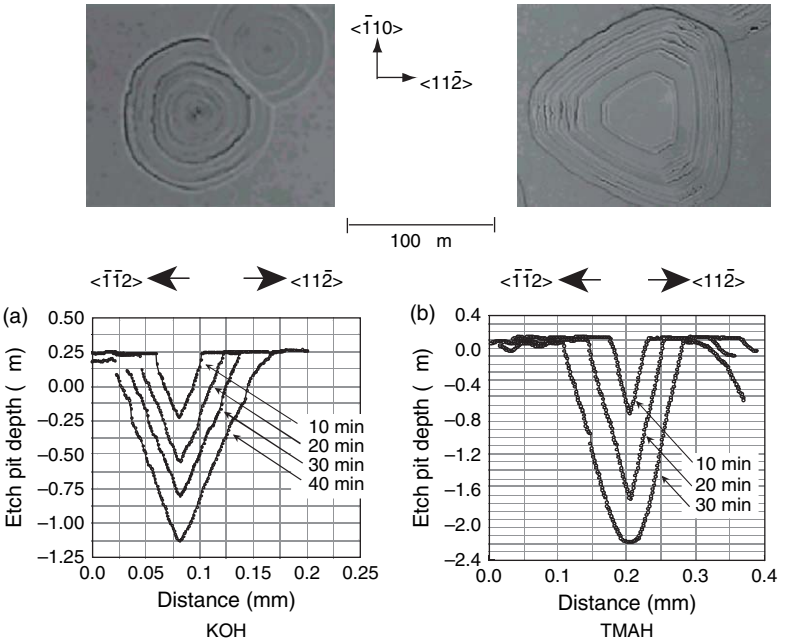


Figure 36 Mirror relationship of etch pit profiles growing on Si (111) during etching in (a) 40% KOH at 70°C, and (b) 25wt.% TMAH at 80°C. In the top are photographs of top views of the pits, and the bottom views are of cross sections measured with time increments. (Source: Sato K, Masuda T, Shikida M 2003 Difference in activated atomic steps on (111) silicon surface during KOH and TMAH etching. *Sens. Mater.* **15-2**, 93–9.)

three to [$\bar{1}\bar{1}$2], respectively. However, the side lengths of those two groups are different. Thus the pit profile becomes triangular in shape. Although both KOH and TMAH etch pits are triangular, note that the orientation of triangle is in mirror relationship between the two etchants.

The etch pit growth in the cross section is shown in the same figure. Both diameter and depth of the pit proportionally increased with an increase of the etching time, though the pits were quite shallow. The increase in the depth saturated when it reached 1–2 μm corresponding to the depth of the defects, while the diameter of the pit kept on increasing in proportion to time. The gradient of the sidewall equaled the ratio of lateral step velocity to the increase in etch pit depth. Lateral step velocity was estimated to be 100 times larger than the increase in pit depth. From an atomic force microscopy (AFM) observation, the step heights were on the order of tens of nanometers. They were regarded as bunched atomic steps. When comparing left- and right-hand sidewalls, the gradients of sidewalls were different. This is because opposite sidewalls are composed of different types of atomic steps as shown in **Figure 37**. A pit nucleated on (111) propagates by a receding movement of atomic steps. It becomes hexagonal

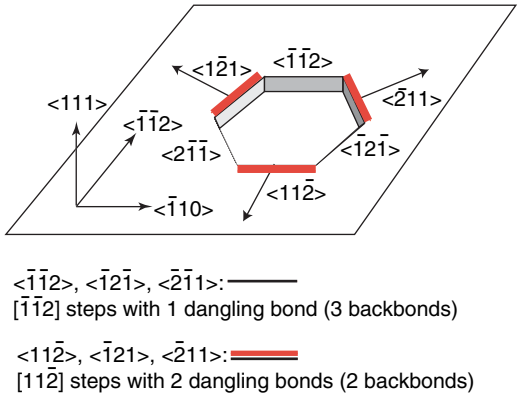


Figure 37 Schematic configuration of a pit on Si (111) composed of two types of stable atomic steps.

profiles composed of stable steps perpendicular to [112]. However, there are two types of steps, one perpendicular to [11$\bar{2}$] and another to [$\bar{1}\bar{1}$2]. They have a different number of dangling bonds and backbonds. Based on the difference in the step velocity of those two types of step, the hexagonal profile becomes triangular. It should be noted that steps having less number of backbonds do not always recede faster than steps having more dangling bonds.

The triangular pit orientation is determined by the relative step velocity of the atomic steps on (111), i.e., by the competition between velocities receding in directions of $[11\bar{2}]$ and $[\bar{1}\bar{1}2]$. The ratio of the velocities for two different steps are actually reversed between KOH and TMAH. The three-backbonded step edge $[11\bar{2}]$ was more stable than the two-backbonded $[\bar{1}\bar{1}2]$ in case of KOH etching. However, in the case of TMAH, two-backbonded steps were more active than three-backbonded steps. As a result, the etch pit profile forms mirror relationships as schematically shown in **Figure 38**.

The step flow model considering step velocities, which is variable, based on the etching conditions, can explain the changeable anisotropy that was observed on the hemispherical specimens discussed in Section 1.07.4.1.4.

Orientations in the vicinity of (111) having a miscut angle can be modeled by a step and terrace model as shown in **Figure 35**. The more the miscut angle, the more the number of atomic steps that exist on a surface in proportion to the miscut angle. Depending on the miscut directions, some surfaces have $[11\bar{2}]$ steps and others have $[\bar{1}\bar{1}2]$ steps. This causes an asymmetric increase in the etch rate by the deviation angle from (111). This results in the triangular contour map pattern on the etch rate diagram shown in **Figure 19**. At the same time, the reason for the triangular contour map patterns rotating 60° between KOH and TMAH has been clarified (Sato *et al.* 2003).

It was further found that the ratio of the step velocities perpendicular to $[11\bar{2}]$ and $[\bar{1}\bar{1}2]$ on Si {111} can also be reversed by a change in solution concentration as well as the difference in etching species such as KOH and TMAH (Cheng *et al.* 2006). The rotation of etch pits is a function of the solution concentration. In case the KOH concentration is extremely high, with values more than 60% KOH, the orientation of the triangular etch pit becomes the same as that of 25wt.% TMAH. On the other hand, in case the TMAH concentration is weak in such a range, i.e., lower that 10wt.% TMAH, the orientation of the triangular etch pit becomes like that of KOH with normal concentrations.

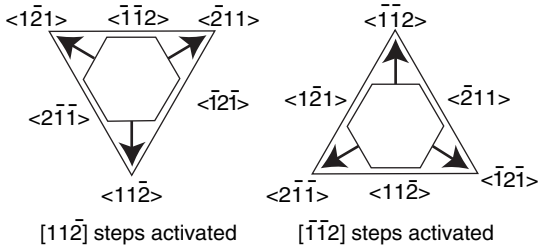


Figure 38 Oriented pit growth on Si (111) governed by the difference in activated step systems.

The critical concentration at which the triangular etch pit rotates is not the same for KOH and TMAH, even if one compares the value in molar concentration. The reason of such reversible anisotropy is considered due to the role of positive ions such as K^+ and TMA^+.

Atom removal from the {111} surface using atomistic Monte Carlo simulation based on first principle calculation (Gosálvez and Nieminen 2003) was recently reported. Simulation results for silicon surfaces having different off-angles in the vicinity of {111} clearly showed the pit nucleation and step propagation processes. Triangular pits growing on {111} were simulated, and the possible rotation in the triangular ridge orientation was predicted as a change of anisotropy in etching.

1.07.4.2.2 Etch profile simulation system

Etching profile simulation is of less importance for making simple MEMS structures such as grooves and diaphragms, which can be fabricated in a single etching step. On the other hand, it is indispensable for designing processes, fabricating microstructures having convex profiles or round profiles composed of many crystallographic orientations.

There are two approaches for the etch profile simulation. One is a macroscopic geometrical approach having knowledge of etch rate as a function of crystallographic orientation. Wulff–Jaccodine's construction (Jaccodine 1962) can be applied to predict the advancement of etching front by time increments, as long as the etch rates for different orientations are known. A few simulation systems have been developed from an early period in MEMS history. Assuming that only three orientations appear as an etch profile, Buser and de Rooij developed a graphical simulation system ASEP (Buser and de Rooij 1991). Sequin (1991) discussed the algorithm of 3D etching profiles appearing at the intersections of different orientations. Than and Büttgenbach (1994) developed a code based on cellular automata considering etch rates for three orientations. Commercially available AnisE (in the system IntelliSuite, a product of IntelliSense) is an etching simulator based on the same principle using cellular automata. SIMODE (Frühauf and Zielke 1996) and MICROCAD (Sato *et al.* 1993) are the systems that use a large number of etch rate data for the analysis, based on experimental characterization of etch rates for a number of crystallographic orientations.

Consequently, the simulated results are satisfactory as long as the database covers the etching condition. The latter will be described in detail later in this section.

Another approach is an atomistic approach considering every atom removal from silicon surface. Palik *et al.* (1985) have proposed a model that Si–Si bonds are broken by the attack of an OH^- ion penetrating from the silicon surface. Moktadir and Camon (1997) simulated atom removal on (111) silicon considering that once the OH is attached to a silicon atom, the strength of the Si backbonds weakens. Gosálvez (2003) calculated backbond weakening based on first principle calculation, and made a simulation code named TAPAS calculating a large number of atoms removed from silicon surfaces by using the Monte Carlo method. It gives qualitative agreements with experimental facts, such as etching profiles and etched surface morphologies that strongly depend on the orientations. He assumed the attachment ratio of OH on the silicon surface as a parameter, although the parameter, of course, needs fitting using experimental data. There is still room for multiscale approaches considering phenomena both in the solid and in the liquid phase, i.e., diffusion in liquid, role of positive ions like K^+, and TMA^+. However, such an atomistic approach is of importance for analyzing etching mechanisms.

Since characterization of anisotropic etch rates is possible by using hemispherical silicon specimen mentioned in Section 1.07.4.1.4, etch rates as a function of orientations for industrial etching conditions were measured (Sato *et al.* 2000, 2001, Shikida *et al.* 2000). Combined with the database, an etching simulation system MICROCAD that predicts etching profiles under a wide range of etching conditions has been developed (Sato *et al.* 2000). The system is linked with the above-mentioned etch rate database ODETTE. The etched profiles are analyzed by using Wulff–Jaccodine's graphical method. Variable parameters in the simulation are surface orientation of the silicon wafer, the shape of the mask pattern, the species of etchant, and its concentration and temperature. It is not restricted for a single etching process, but a series of etching processes using different mask patterns or etching solutions can also be analyzed.

The MICROCAD (current name is FabMeister-ES distributed by Mizuho Information & Research Institute Inc.) allows the designing of process parameters such as size and shape of mask patterns, the number of etching steps, and the etching conditions. Examples of the process design using the system are as follows:

(1) Compensation mask design minimizes undercut etching of the mask.
(2) Successive etching process design provides 3D structures including smooth round etch profiles among polyhedral ones.
(3) Optimization of etching parameters: etching medium, concentration, and temperature are optimized in terms of the desirable etching profile.

Etch profile analysis is robust even when etching is applied deep from both sides to penetrate the wafer. **Figure 39** shows such robustness calculated by comparing experimental and predicted etch profiles. Wafer orientation is (100). Sides of square mask apertures are aligned to (110). Etching fronts first penetrate the wafer thickness, and then they advance laterally. Finally, through-holes are connected in lateral directions. Multistep etching process also can be analyzed with MICROCAD (FabMeister-ES) system.

1.07.4.3 Anisotropic Etching of Quartz

Single-crystal alpha-quartz is a material showing piezoelectric effects, and is applied to oscillators, filters, and some MEMS devices such as rotational angular rate sensors (rate gyro).

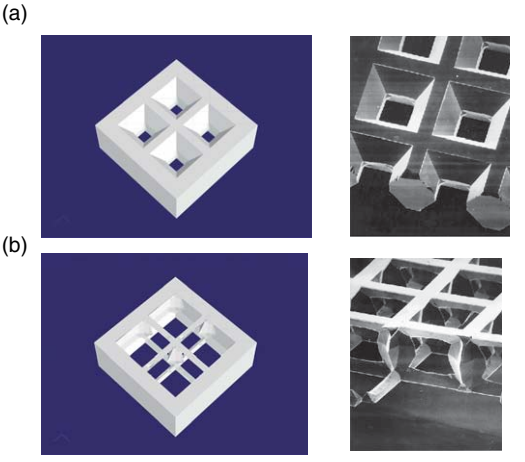


Figure 39 The etch profile simulation results compared to the actual silicon chip. The etching was applied from both sides of the (100) wafer using a KOH solution. (a) Etching fronts meet in the middle of the wafer and (b) further advance in lateral directions resulting in the through-holes also in lateral directions. (Source: Sato K, Asaumi K, Kobayashi G, Iriye Y, Shikida M 2000 Development of an orientation-dependent anisotropic etching simulation system MICROCAD. *Electron Comm. Jpn. Part 2* **83–4**, 13–22.)

Orientation-dependent etching is applicable to quartz, though it has a crystal lattice system quite different from that of silicon. The atomic component is SiO_2. It can be etched with NH_4HF_2 (ammonium bifluoride)-saturated solution. The chemical reaction is given below:

$$SiO_2 + 3NH_4HF_2 \rightarrow SiF_6^{2} + 2H^{+} + 3NH_3 + 2H_2O$$

The mask material is Au on Cr. Cr is a binding layer enhancing adhesion of the Au layer to the quartz wafer. The film thicknesses are typically 300 Å Cr and 1500 Å Au.

Single-crystal alpha-quartz has a hexagonal system as shown in **Figure 40**. Since the crystal system is different from a cubic lattice, expressing the orientations using the normal Miller index is not appropriate in the case of quartz. For example, the crystal properties of the a1- and a2-axis directions are physically identical, but with the Miller index they are expressed as [110] and [100]. Therefore, when expressing a hexagonal crystal system like quartz or Zn, one usually uses a special index, the Miller–Bravais index. As shown in **Figure 40**, there are four axes in a hexagonal system crystal structure: three axes (a1, a2, and a3) on one base plane meet at angles of 120°, and the other axis (Z) is perpendicular to this plane. This crystal structure is represented accurately by the Miller–Bravais index because it has four parameters (h, k, i, l) corresponding to the four axes. Three parameters for a base plane are not independent, and are in a relationship as given below:

$$h + k + i = 0 \qquad [1]$$

Because of this, one can also express the Miller–Bravais index as ($hk\bullet l$).

The principal crystal planes of alpha-quartz after crystal growth are shown in **Figure 41**. Some principal orientations have idiomatic names, as given below:

$$\text{m} : \{10\bar{1}0\}$$
$$\text{R} : (10\bar{1}1), (\bar{1}101), (0\bar{1}11)$$
$$\text{r} : (\bar{1}011), (1\bar{1}01), (01\bar{1}1)$$
$$\text{s} : (1\bar{2}11), (\bar{2}111), (11\bar{2}1)$$
$$\text{x} : (51\bar{6}1), (\bar{6}511), (1\bar{6}51)$$

In the following sections the set of orientations perpendicular to the Z-axis, including the m-planes will be referred to as the equator. The fastest crystal growth has an orientation of (0001), suggesting that etching may be the fastest at this orientation.

In the same manner as for single-crystal silicon, Cheng *et al.* (2005) have characterized the etch rates as a function of orientations by making spherical quartz specimens whose diameter was 33 mm. It has a top oriented to Z-axis (0001), and the equator including m-planes $\{10\bar{1}0\}$. To cover the areas near $\{10\bar{1}0\}$, the specimen, unlike silicon, was designed more as a spherical body than as a hemisphere.

The etch rate contour maps at 70°C and 80°C are shown in **Figure 42**, and etch rates for 13 different crystallographic orientations including Z-, R-, r-, s-, and x-planes are listed in **Table 4**.

It is observed from the contour maps that the etch rate of quartz has threefold symmetry around the (0001) axis as a reflection of the hexagonal system. Although the (0001) orientation is located in a high

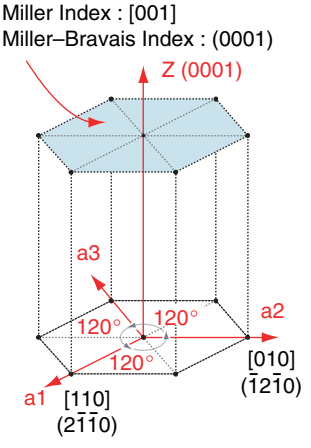


Figure 40 Hexagonal system crystal structure of quartz expressed by the Miller–Bravais index.

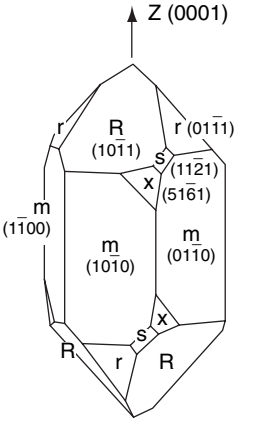


Figure 41 Principal orientations appearing after crystal growth and their Miller–Bravais index.

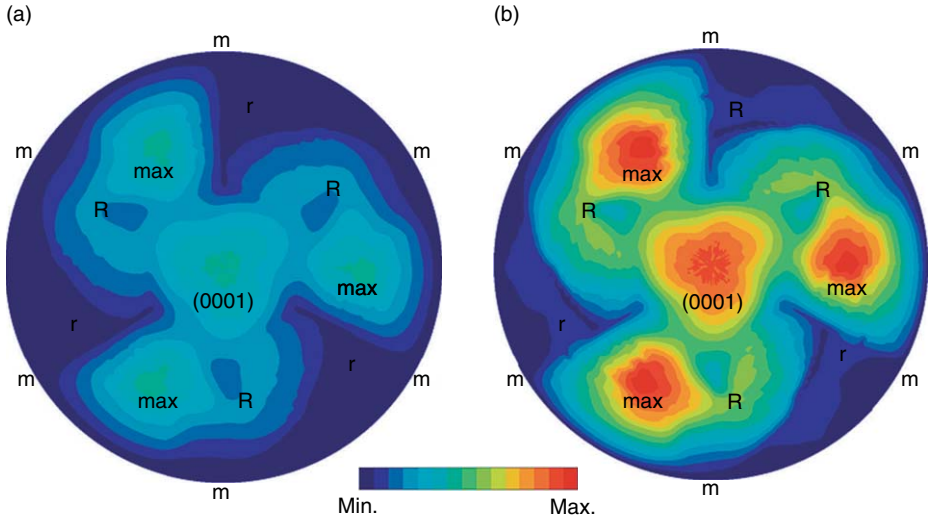


Figure 42 Etch rate contour maps of alpha-quartz at temperatures of 70°C (a) and 80°C (b). The center is oriented to the *Z*-plane (0001), and *m*-planes are on the equator line. (Source: Cheng D, Sato Ka, Shikida M, Ono A, Sato Ke, Asaumi K, Iriye Y 2005 Characterization of orientation-dependent etching properties of quartz: Application to 3-D micromachining simulation system. *Sens. Mater.* **17-4**, 179–86.)

Table 4 Etch rates for 13 different crystallographic orientations of alpha-quartz at etching temperatures of 70°C and 80°C

	Etch rates ($\mu m\ min^{-1}$)	
Crystallographic orientation	*70°C*	*80°C*
(0001) Z	0.571 (1.00)	1.261 (1.00)
$(\bar{1}2\bar{1}2)$	0.565 (0.99)	1.270 (1.01)
$(\bar{1}2\bar{1}0)$	0.013 (0.02)	0.032 (0.03)
$(\bar{1}102)$	0.347 (0.61)	0.755 (0.60)
$(\bar{1}101)$R	0.242 (0.42)	0.608 (0.48)
$(\bar{1}100)$m(Min.)	0.002 (< 0.01)	0.003 (< 0.01)
$(\bar{2}112)$	0.256 (0.45)	0.631 (0.50)
$(\bar{2}111)$s	0.512 (0.90)	1.177 (0.93)
$(\bar{2}110)$	0.043 (0.08)	0.094 (0.08)
$(\bar{5}16\bar{1})$x	0.054 (0.09)	0.116 (0.09)
$(\bar{1}012)$	0.120 (0.21)	0.297 (0.24)
$(\bar{1}011)$r	0.059 (0.10)	0.143 (0.11)
$(\bar{2}42\bar{3})$Max.	0.590 (1.03)	1.432 (1.14)

In parentheses are normalized values relative to (0001) orientation (Cheng *et al*. 2005).

etch rate region, the value was not the highest as speculated from the crystal growth shape. The maximum etch rates exist at these three positions at equivalent crystallographic orientations. The location of maximum is at an angle of 56° measured from the top (0001), and threefold symmetry in a circumferential direction. The Miller–Bravais index of this orientation can be $\{\bar{2}42\bar{3}\}$.

On the other hand, extremely low etch rates were measured along the equator, even as low as close to zero. This means that a (0001) wafer can be etched perpendicular to the wafer surface by wet etching with minimal side etching in the same manner with deep RIE.

Cheng *et al.* (2005) have analyzed the etching profiles on a (0001) quartz wafer using the MICROCAD (FabMeister-ES) system by implementing the etch rate distribution obtained above. **Figure 43** shows the mask pattern and the etched profile obtained by the simulation system. It is clear that the sidewalls of the cavity are steep and mask undercut was quite little in omni directions. This is the reflection of extremely low etch rates including *m*-planes.

An example of a recent application of quartz anisotropic etching is introduced below. Hida *et al.* (2005) have designed a quartz tuning fork to be applied to noncontact mode AFM systems. The tuning fork is a pair of quartz beams oscillating in plane. One of the beams is used as an AFM probe. Quartz AFM probe so far did not have sharp tips integrated in a single body like silicon probes because of the difficulty in micromachining. Hida *et al.* have designed a probe having a sharp tip at the end of the beam as shown in **Figure 44**. By utilizing a MICROCAD (FabMeister-ES) system, they designed an etching process that fabricated a tip and beam on the same body. It became possible to fabricate a micromachined tip structure on a quartz chip, as shown in the figure.

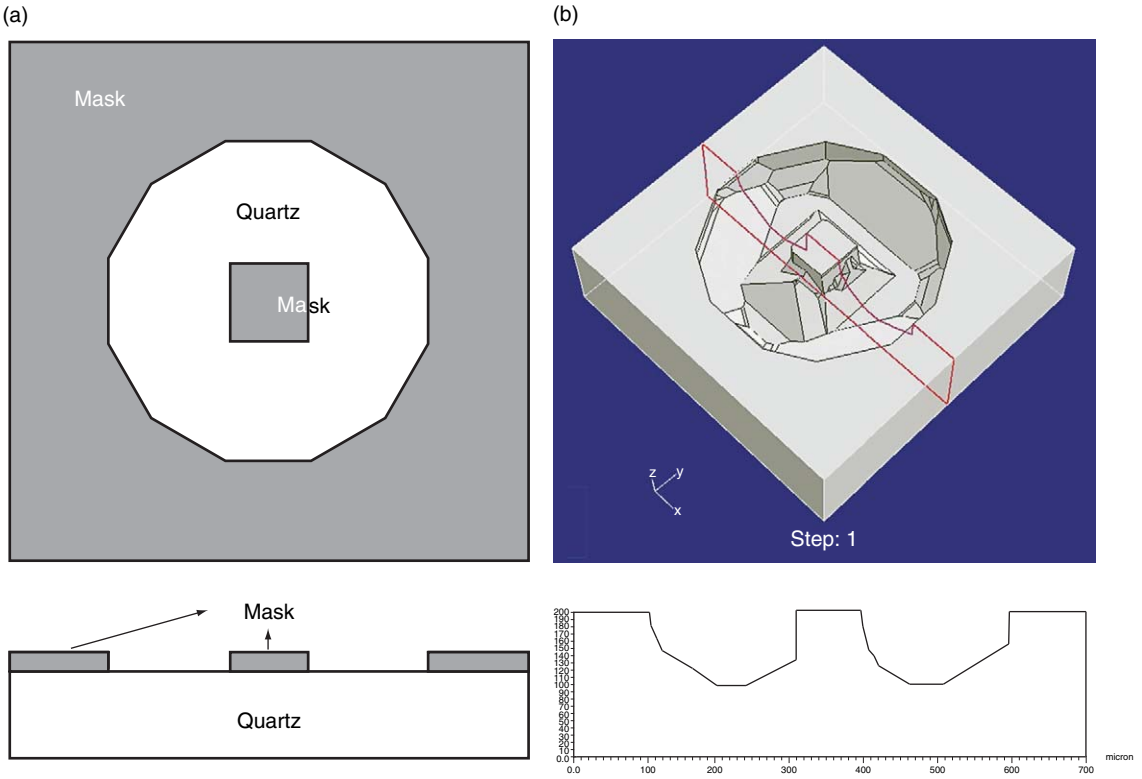


Figure 43 Etch profiles of (0001) quartz wafer analyzed using MICROCAD (FabMeister-ES) system linked with etch rate database of quartz. (Source: Cheng D, Sato Ka, Shikida M, Ono A, Sato Ke, Asaumi K, Iriye Y 2005 Characterization of orientation-dependent etching properties of quartz; application to 3-D micromachining simulation system. *Sens. Mater.* **17-4**, 179–86.)

1.07.5 Anodic Etching of Silicon in HF

Silicon etching is modulated by applying a potential between silicon and the etching solution. Without an external biased potential, silicon wafer during etching is in a state of OCP. When the applied potential between silicon and solution becomes more positive, more holes are supplied to the silicon surface, and the etching speed increases. When the potential reaches the passivation potential, an oxide starts to form at the silicon–solution interface. Thus etching stops in alkaline etching solutions like KOH and TMAH, as mentioned in the previous section. However, in case of HF solutions, the generated oxide is ready to dissolve in HF, and etching advances further. This is called anodic etching, or anodization.

The following two conditions are both indispensable for the silicon to be etched in HF solution, i.e., (1) charges (holes) are supplied in silicon (2) OH^- ions are transported from the solution. Depending on the conditions in the supply of the holes and in the transportation of OH^- ions, etched surface morphologies are variable in HF systems.

1.07.5.1 Electropolishing

For the purpose of smoothening a silicon wafer surface, photoelectrochemical etching using highly concentrated HF solution is used. The etching bath is basically the same as the one shown in **Figure 30**, providing regulated potentials to silicon wafer, though an apparatus illuminating wafer surface is additionally attached to the system.

When the applied potential is low, microporous silicon is formed on the wafer surface, as described in the next section. When the current supplied exceeds a certain value based on the increase in applied potential, the porous film delaminates and bright electropolishing occurs (Lehmann 1993).

The electropolishing advances under a condition in a concentrated HF solution, where the OH^- ion is depleted. The mass transport in the solution limits

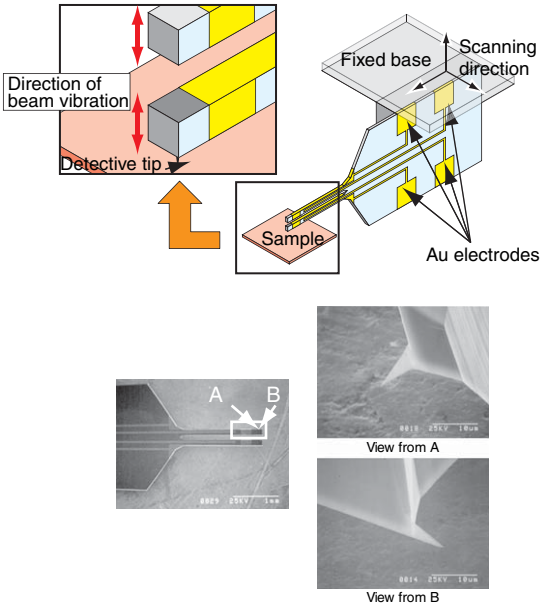


Figure 44 Quartz tuning fork having a sharp tip etched in a single body to be applied to non-contact AFM system. (Source: Hida H, Shikida M, Fukuzawa K, Ono A, Sato Ke, Asaumi K, Iriye Y, Cheng D, Sato Ka 2005 Proposal of new type of micro-machined quartz tuning fork AFM probe. *Proc. IEEE Int. Conf. on Micro Electro Mechanical Systems*, Miami Beach, FL, USA, pp. 323–26.)

the reaction. The hills of rough silicon surface can get more OH^- ions than can the valleys; thus smoothening of the surface occurs.

1.07.5.2 Porous Silicon

Porous silicon formation was first found by Uhlir (1956). Pore size and porosity can widely range from nanometers to tens of microns by etching conditions such as current density, illumination intensity, and silicon types.

Porous silicon has been investigated aiming at following applications.

1. Quantum structures
2. Photo- and electroluminescent devices
3. Oxide films of several microns thick or more, due to the ease of oxidization; sacrificial layer for releasing MEMS structures; dielectric isolation of integrated circuits
4. Cleavage layer in SOI production (ELTRAN method)
5. Mold for tubular MEMS structure

Porous silicon is categorized into two classes, microporous and macroporous silicon, by the differences in size and formation mechanism.

1.07.5.2.1 Microporous silicon

Microporous silicon is formed both on p- and on n-type silicon under low current density in highly concentrated HF. In such a condition, the holes are depleted, while HF is accumulated at the silicon–solution interface. The depletion in holes due to quantum confinement in the fine pore structure is the basis of the formation mechanism. The pore diameter ranges 4–20 nm (Arita and Sunohara 1997, Lehmann 1996).

Such a microporous layer can be thermally oxidized in a short time, allowing a thick layer of oxide film for the purpose of electrical isolation or sacrificial layer for micromachining. Yoshida *et al.* (1992) proposed utilizing p–n junction allowing photo-induced holes to make microporous silicon in the p-layer, eliminating metal wiring for applying positive potential to the p-layer. They have fabricated an n-type silicon beam released from the substrate by removing the sacrificial layer of oxidized microporous silicon. The process is illustrated in **Figure 45**, where the silicon beam structure is of n-type, the sacrificial layer of p-type, and the substrate of n-type.

Porous silicon remains a single crystal. Epitaxial growth is still possible on the porous silicon. This allows the fabrication process of SOI wafer, which is called the ELTRAN method. The process consists of bonding two wafers, one having an epitaxial

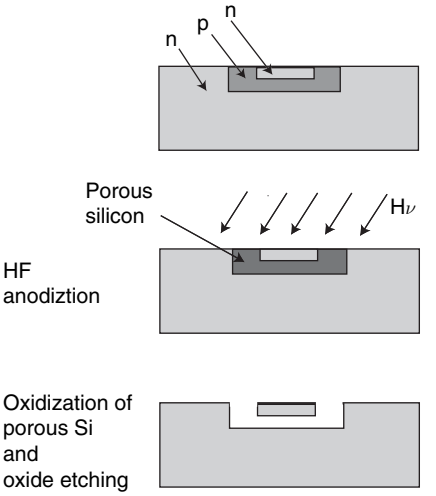


Figure 45 Photo-induced selective anodization of p-silicon for the purpose of sacrificial layer. (Source: Yoshida T, Kudo T, Ikeda K 1992 Photo-induced preferential anodization for fabrication of mono-crystalline micromechanical structures, *Proc. IEEE MEMS '92*, Travemünde, Germany, pp. 56–62.)

layer on porous silicon and the other having oxide on top, followed by cleavage at the porous silicon layer making the bonded wafers apart, with the epitaxial layer remaining on top of the oxidized wafer.

1.07.5.2.2 Macroporous silicon

Pore diameters of up to 10 μm can be fabricated to the depth of hundreds of microns by using a special etching bath illuminating the rear side of a wafer during etching. An example of an etching bath is shown in **Figure 46**.

Macroporous silicon is only available with lightly doped n-type silicon under higher applied potentials than those for microporous silicon in diluted HF (less than 10wt.% HF) solution (Zhang 1991). One of the advantages of this process is that the location of the pore can be determined by giving a seed of pore using photolithographic method. When pyramidal recesses are made on a Si (100) wafer surface using anisotropic chemical etching prior to HF anodic etching, the macropore grows, starting from those recesses in a direction of illumination coming from the rear side of the wafer as shown in **Figure 47**. The reason for such directional pore growth is that carriers are generated in silicon by illumination and holes are localized at the bottom tip of the pores.

As an example of applications of such a macropore having a large structure height, a large number of small glass cells for the purpose of microchemical analysis systems were fabricated by Homma *et al.* (2004). They first made macropores like those in **Figure 47**, then the pore walls were totally oxidized. By removing the silicon substrate from the rear side, transparent cuvettes having a square cross section were fabricated. The diameter of the cuvette was typically 10 μm and the length was more than 100 μm.

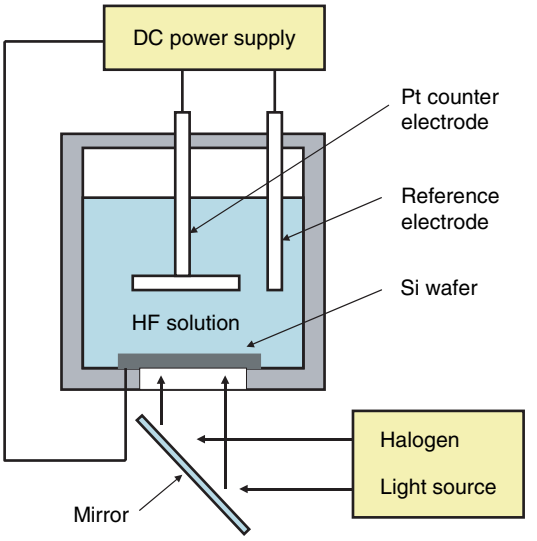


Figure 46 Etching bath having rear-side illumination for macroporous silicon process.

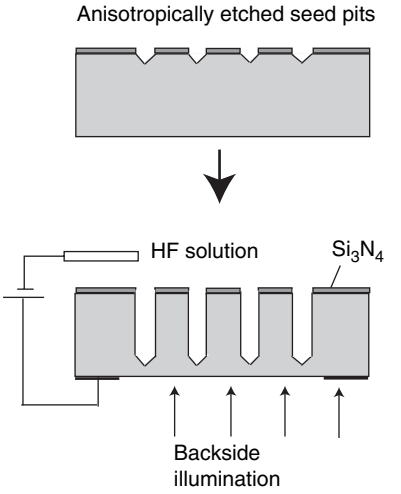


Figure 47 Macroporous silicon process starting from seed pits made by photolithographic means.

References

Alavi M, Büttgenbach S, Schumacher A, Wagner H J 1991 Laser machining of silicon for fabrication of new microstructures. *Tech. Dig. Transducers* **91**, 512–4

Allongue F, Costa-Kieling V, Gericher H 1993 Etching of silicon in NaOH solutions. *J. Electrochem. Soc.* **140–4**, 1009–26

Arita Y, Sunohara Y 1997 Formation and properties of porous silicon film. *J. Electrochem. Soc.* **124**, 285–95

Bäcklund Y, Rosengren L 1992 New shapes in Si (100) using KOH and EDP etches. *J. Micromech. Microeng.* **4**, 75–79

Buser R A, de Rooij, N F 1991 ASEP: A CAD program for silicon anisotropic etching. *Sens. Actuators A* **28**, 71–8

Cheng D, Gosálvez, M A, Hori T, Sato K, Shikida M 2006 Improvement in smoothness of anisotropically etched silicon surfaces: Effects of surfactant and TMAH concentrations. *Sens. Actuators A* **125**, 415–21

Cheng D, Sato Ka, Shikida M, Ono A, Sato Ke, Asaumi K, Iriye Y 2005 Characterization of orientation-dependent etching properties of quartz: Application to 3-D micromachining simulation system. *Sens. Mater.* **17-4**, 179–86

Cheng D, Gosálvez, M A, Shikida M, Sato K 2006 A universal parameter for silicon anisotropic etching in alkaline solutions. *Proc. 19th IEEE Int. Conf. Micro Electro Mechanical Systems,* Istanbul, Turkey, pp. 318–21

Elwenspoek M 1993 On the mechanism of anisotropic etching of silicon. *J. Electrochem. Soc.* **140-4**, 2075–80

Frühauf J, Zielke D 1996 3-Dimensional simulation of the orientation dependent etching. *Tech. Dig. Micro System Technol.* **96**, 747–9

Gosálvez M A 2003 Atomistic modeling of anisotropic etching of crystalline silicon. Dissertations of Laboratory of Physics, Helsinki University of Technology, Dissertation 123, http://lib.hut.fi/Diss/2003/isbn9512267071

Gosálvez M A, Nieminen R M 2003 Surface morphology during anisotropic wet chemical etching of crystalline silicon. *N. J. Phys.* **5** 100.1–28

Hashimoto H, Tanaka S, Sato K, Ishikawa I 1993 Chemical isotropic etching of single-crystal silicon for acoustic lens of scanning acoustic microscope. *Jpn. J. Appl. Phys.* **32**, 2543–6

Hida H, Shikida M, Fukuzawa K, Ono A, Sato Ke, Asaumi K, Iriye Y, Cheng D, Sato Ka 2005 Proposal of new type of micro-machined quartz tuning fork AFM probe. *Proc. IEEE Int. Conf. Micro Electro Mechanical Systems,* Miami Beach, FL, USA, pp. 323–6

Homma T, Sato H, Mori K, Osaka T, Shoji S 2004 High aspect ratio nanovolume glass cell array fabricated by area-selective silicon electrochemical etching process. *Proc. IEEE MEMS-'04,* Maastricht, The Netherlands, pp. 705–8

Jaccodine R J 1962 Use of modified free energy theorem to predict equilibrium growing and etching shapes. *J. Appl. Phys.* **33-8**, 2643–7

Kakushima K, Mita M, Kobayashi D, Hashiguchi G, Endo J, Wada Y, Fujita H 2001 Micromachined tools for nano technology – Twin nano-probes and nano-scale gap control by integrated microactuators. *Tech. Dig. 14th IEEE Int. Conf. Micro Electro Mechanical Systems,* Interlaken, Switzerland, pp. 294–7

Kamisuki S, Fujii M, Takekoshi T, Tezuka C, Atobe M 2000 A high resolution electrostatically-driven commercial inkjet head. *Tech. Dig. 13th IEEE Int. Conf. Micro Electro Mechanical Systems,* Miyazaki, Japan, pp. 793–8

Kendall D L, de Guel G R, Torres-Jacome A 1982 The Wagon Wheel method applied around the [011] zone of silicon. *Electrochem. Soc. Extend. Abstr.* **82-1**, 209–11

Kloeck B, Collins S D, de Rooij N F, Smith R L 1989 Study of electro-chemical etch-stop for high-precision thickness control of silicon membranes. *IEEE Trans. Electron Devices* **36-4**, 663–9

Koide A, Sato K, Suzuki S, Miki M 1992 A multistep anisotropic etching process for producing 3-D Si accelerometers. *Technical Digest, 11th Sensor Symposium,* Tokyo, Japan, pp. 23–6

Lehmann V 1993 The physics of macropore formation in low doped n-type silicon. *J. Electrochem. Soc.* **140**, 2836–43

Lehmann V 1996 Porous silicon – A new material for MEMS. *Proc. IEEE MEMS-'96,* San Diego, CA, USA, pp. 1–6

Moktadir Z, Camon H 1997 Monte Carlo simulation of wet etching of silicon: Investigation of (111) surface properties. *Model. Simul. Mater. Sci. Eng.* **5**, 481–8

Palik E D, Bermudez V M, Glembocki O J, 1985 Ellipsometric study of orientation-dependent etching of silicon in aqueous KOH. *J. Electrochem. Soc.* **132–4**, 871–84

Sato K, Koide A, Tanaka S, 1993 Anisotropic etching simulation system as a design tool for fabrication 3-D microstructures in single crystal silicon. *Proc. 1st IFToMM Int. Symp. Micromechanisms,* Tokyo, Japan, pp. 155–60

Sato K, Asaumi K, Kobayashi G, Iriye Y, Shikida M 2000 Development of an orientation-dependent anisotropic etching simulation system MICROCAD. *Electron Commun. Jpn. Part 2,* **83-4**, 13–22

Sato K, Masuda T, Shikida M 2003 Difference in activated atomic steps on (111) silicon surface during KOH and TMAH etching. *Sens. Mater.* **15-2**, 93–9

Sato K, Shikida M, Matsushima Y, Yamashiro T, Asaumi K, Iriye Y, Yamamoto M 1998 Characterization of orientation dependent etching properties of single crystal silicon: Effects of KOH concentration. *Sens. Actuators A* **64-1**, 87–93

Sato K, Uchikawa D, Shikida M 2001 Change in orientation-dependent etching properties of single-crystal silicon caused by a surfactant added to TMAH solution. *Sens. Mater.* **13-5**, 285–91

Seidel H, Csepregi L, Heuberger A, Baumgärtel H 1990a Anisotropic etching of crystalline silicon in alkaline solutions. I. Orientation dependence and behavior of passivation layers. *J. Electrochem. Soc.* **137-11**, 3612–26

Seidel H, Csepregi L, Heuberger A, Baumgärtel H 1990b Anisotropic etching of crystalline silicon in alkaline solutions. II. Influence of dopants. *J. Electrochem. Soc.* **137-11**, 3626–32

Sekimura M 1999 Anisotropic etching of surfactant-added TMAH solution. *Proc. 12th IEEE Int. Conf. Micro Electro Mechanical Systems,* Orlando, FL, USA, pp. 650–5

Sequin C H 1991 Computer simulation of anisotropic crystal etching. *Technical Digest Transducers '91,* San Francisco, CA, USA, pp. 801–6

Shikida M, Sato K, Tokoro K, Uchikawa D 2000 Differences in anisotropic etching properties of KOH and TMAH. *Sens. Actuators A* **80-2**, 179–88

Shikida M, Nanbara K, Koizumi T, Sasaki H, Odagaki M, Sato K, Ando M, Furuta S, Asaumi K 2002 A model explaining mask-corner undercut phenomena in anisotropic silicon etching: A saddle point in the etching-rate diagram. *Sens. Actuators A* **97-8**, 758–63

Shikida M, Hasada T, Sato K 2006a Fabrication of densely arrayed micro-needles with flow channel by mechanical dicing and anisotropic wet etching. *J. Micromech. Microeng.* **16**, 1740–7

Shikida M, Hasada T, Sato K 2006b Fabrication of hollow needle structure by dicing, wet etching, and metal deposition. *J. Micromech. Microeng.* **16**, 2230–9

Shikida M, Odagaki M, Todoroki N, Ando M, Ishihara Y, Ando T, Sato K 2004a Non-photolithographic pattern transfer for fabricating arrayed three-dimensional microstructures by chemical anisotropic etching. *Sens. Actuators A* **116**, 264–71

Shikida M, Ando M, Ishihara Y, Ando T, Sato K, Asaumi K 2004b Non-photolithographic pattern transfer for fabricating pen-shaped microneedle structures. *J. Micromech. Microeng.* **14**, 1462–7

Tabata O, Asahi R, Funabashi H, Sugiyama S 1991 Anisotropic etching of silicon in $(CH_3)_4NOH$ solutions. *Tech. Dig. Transducers '91, IEEE Int. Conf. Solid-State Sensors and Actuators,* San Franciso, CA, USA, pp. 811–14

Tanaka H, Abe Y, Inoue K, Shikida M, Sato K 2003 Effects of ppb-level metal impurities in aqueous potassium hydroxide solution on the etching of Si{110} and {100}. *Sens. Mater.* **15-1**, 43–51

Than O, Büttgenbach S 1994 Simulation of anisotropic chemical etching of crystalline silicon using a cellular automata model. *Sens. Actuators A* **45**, 885–9

Ueda T, Kohsaka F, Iino T, Yamazaki D 1986 Theory to predict etching shapes in quartz crystal and its application to design devices. *T-SICE* **23-12**, 1233–8 (in Japanese)

Uhlir A 1956 Electrolytic shaping of germanium and silicon. *Bell Syst. Tech. J.* **35**, 333–47

Weirauch D F 1975 Correlation of the anisotropic etching of single-crystal silicon spheres and wafers. *J. Appl. Phys.* **46-4**, 1478–83

Yoshida T, Kudo T, Ikeda K 1992 Photo-induced preferential anodization for fabrication of mono-crystalline micromechanical structures. *Proc. IEEE MEMS '92,* Travemünde, Germany, pp. 56–62

Zhang X G 1991 Mechanism of pore formation on n-type silicon. *J. Electrochem. Soc.* **138**, 3750–6

Biographies

Kazuo Sato received a BS degree in mechanical engineering from Yokohama National University in 1970, and a Ph.D. degree from the University of Tokyo in 1982. He worked with Hitachi Ltd. from 1970 to 1994. He has been a professor at Nagoya University since 1994.

He started MEMS research in 1983, and cochaired the IEEE MEMS-97 workshop. His research interests include chemical anisotropic etching, microsensors, microactuators, characterization of MEMS materials, and micro/nanophysics. He is a Fellow of the Japan Society of Mechanical Engineers (JSME), a senior member of IEE, and a member of JSPE and IEEE.

Mitsuhiro Shikida received his BS and MS degrees in electrical engineering from Seikei University, Tokyo, in 1988 and 1990, respectively. He received a Ph.D. from Nagoya University in 1998. From 1990 to 1995, he worked at Hitachi Ltd., Tokyo. In 1995, he joined the Department of Micro-System Engineering at Nagoya University as a research associate. He was an assistant professor from 1998 to 2004 and has been an associate professor since 2004. He joined the EcoTopia Science Institute at Nagoya University in 2004. His research interests include microfabrication of 3D microstructures for medical applications and micro total analysis systems for biotechnologies.

1.08 Dry Etching

Franz Laermer, Corporate Sector Research and Advance Engineering, Robert Bosch GmbH, Stuttgart, Germany

1.08.1 Introduction

In the early days of microelectromechanical systems (MEMS), wet etching with potassium hydroxide and other alkaline solutions was the most important method of fabricating silicon microstructures (Petersen 1982). In the beginning, these etchants were mainly used to fabricate membrane-type devices, like pressure and mass flow sensors. Due to its crystal orientation-selective etching behavior, with fast etching crystal planes like (100) and slow etching crystal planes like (111), anisotropic wet etching yields cavities, which are limited by walls of (111) crystal orientation. The etch practically stops in the vertical direction also when a cavity reaches a V-form (Bean, 1978, Seidel *et al.* 1990). Its accuracy and repeatability over many wafers processed is due to the anisotropic nature of the etch. Precise V-grooves found some applications in the micro-optic field, mainly for fiber couplers and fiber positioners. Supporting techniques for vertical etch depth control were introduced at that time, e.g., dopant-selective etch stop at heavily boron-doped silicon layers (Palik *et al.* 1982) and potential-selective electrochemical etch stop (Kloek *et al.* 1989) at reversely biased p–n junctions in silicon.

This degree of accuracy is not provided by an isotropic wet etchant, which would etch at the same speed in all directions, including underetching of the mask in the lateral directions. This is one of the reasons why well-known isotropic wet etchants for silicon like HNA (a mixture of hydrofluoric, nitric and acetic acid) have been utilized only as a limited microstructuring technology for MEMS (Robbins and Schwartz 1976). Another reason is that masking of the HNA etch requires thin film layers that are chemically very resistant, e.g., silicon carbide.

Although successful in a number of applications requiring low-cost technologies for cavity etching, the anisotropic etchants of the classical MEMS technology show severe limitations with regard to required masking materials (hard mask materials like silicon oxide, silicon nitride, or carbide needed), their incompatibility with semiconductor manufacturing (especially in the case of monolithic integration of mechanical and electrical functionalities), the limited etching speed, the dependence of the etch on crystal orientation, generally nonvertical sidewalls, and the extremely fast etching of convex corners. For structures of higher complexity, for example, membranes or beams containing extra features like seismic masses, geometries exposing

convex corners during the etch are unavoidable. To protect convex corners from fast etching attack, various corner compensation techniques were suggested (Mayer *et al.* 1990, Menz and Mohr 1997, Sandmaier 1991, Wu and Ko 1989), some of them containing rather complicated sacrificial structures. The complexity of corner compensation together with the dependence on crystal orientation significantly restricts design flexibility and application range of the classical microstructuring technology. People could often not or only under large efforts realize the microstructures they were dreaming of. Clearly a more flexible technology, free from these restrictions, was urgently needed as a key enabler for MEMS devices of enhanced functionality.

The LIGA technique was the first step toward this goal, offering freedom of design for a wide range of microsystems. LIGA stands for the German term Lithographie, Galvanik, Abformung (Ehrfeld 2002, Menz and Mohr 1997) and was developed at the Research Center Karlsruhe, Germany. A thick layer of polymethyl methacrylate (PMMA) resist is exposed through a mask to X-ray radiation from an electron synchrotron and is developed after exposure. The high optical quality of the X-ray beam from the electron synchrotron together with the small absorbance and the dispersion of X-rays within the PMMA film yield extremely precise 3D structures, defined by the 2D patterns used for masking the polymer. Thus, polymer structures of several hundred micrometers in depth and with aspect ratio of 100:1 can be produced. Galvanoforming of the polymer negative form and subsequent demolding yields positive structures in metal, e.g., nickel. Various replication techniques such as molding or hot embossing in polymer, and again galvanoforming of metal, and combinations between different molding and demolding techniques are used for fabricating a large number of metal or plastic parts from a single LIGA masterform. Given the extremely high manufacturing costs for a LIGA masterform, successful replication to produce extremely high volumes was identified as the key to reach acceptable costs at the product level. Although LIGA demonstrated breathtaking microstructures and microstructuring capabilities at that time, the technology has for long been limited to a small number of applications providing sufficiently high market volume, mainly because of the extreme costs related to this technology. However, LIGA showed to the world what MEMS could be if only the right technologies were there. The challenge was to fabricate microstructures similar to LIGA, but with a process technology closer to industrial standards, and at a more reasonable cost level.

1.08.2 Silicon Etching from the Gas Phase

Plasma etching or reactive ion etching (RIE) was already well established in semiconductor technology at that time (Manos and Flamm 1989, Oehrlein 1990, Schwartz and Schaible 1979). The classical tool, which was used in the early RIE work, is the diode reactor shown in **Figure 1**. Detailed descriptions of different plasma reactors and processes can be found in Chapman (1980). The scheme depicts the key elements of the typical RIE configuration: a number of reaction gases provided to the chamber by adjustable mass flow controllers, a grounded showerhead electrode as a gas inlet, a substrate electrode carrying the wafer, and a power supply to the substrate electrode by radio frequency (RF) via a coupling capacitor (which is mostly part of a more complex matching unit to adapt RF impedances). Radio frequencies generally used are in the low-frequency (380 kHz), mid-frequency (2 MHz), or high-frequency (13.56 MHz) regime, in most cases 13.56 MHz is used as a standard for plasma generation.

Pumping produces the low-pressure conditions in the reactor chamber needed for the electrical discharge to form between the powered and the

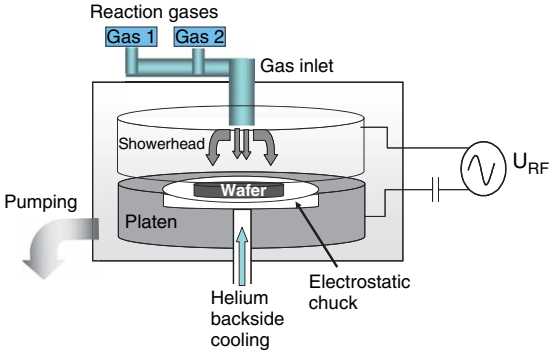


Figure 1 Typical reactive ion etching (RIE) configuration. The RIE reactor is pumped to low-pressure conditions. Reaction gases are supplied to the chamber via the upper showerhead electrode. The lower electrode is powered by radio frequency (RF) to drive the plasma discharge and accelerate the ions toward the substrate.

grounded electrode, which builds up the plasma in an atmosphere of one or more compound gases, at a pressure ranging from 0.1 to 100 Pa.

The coupling capacitor allows the powered electrode to float at a DC potential, which is generated from the applied RF voltage, together with the different mobilities of negative charges (electrons) and positive charges (ions) in the plasma. This is the so-called DC bias voltage, which accelerates the ions toward the powered substrate electrode.

Isotropic, vertical, and all different kinds of tapered profiles could be etched into silicon for applications like trench isolation, contact hole formation, dynamic random access memory (DRAM) trench cells (trench capacitors for memory) and many more, during the manufacturing of electronic circuits. A general overview on RIE can be found in Oehrlein (1990) and a review on silicon RIE is given in Schwartz and Schaible (1979). The capabilities of RIE, mainly its independence on crystal orientation and the potential to fabricate arbitrarily shaped geometries, made plasma etching a promising candidate as a new microstructuring technique for the MEMS field. Early work has demonstrated that micromechanical elements containing shallower pattern features could well be realized using the classical RIE approaches, which were adapted from semiconductor manufacturing technique (Diem *et al.* 1995, Juan *et al.* 1996, Li *et al.* 1995, Sung and Pang 1993). Although still insufficient for deeper structures, and limited by low etching speed and mask selectivities, the intrinsic potential of plasma technology for etching deeper structures at a higher speed was already felt within the MEMS community.

1.08.2.1 Isotropic Etching

With regard to etching, a reactive plasma consists of two relevant kinds of species: the dominating part in numbers, by orders of magnitude, is the chemically active neutral species, the so-called radicals. They are produced by the decomposition of gas molecules into fractional parts after collisions between gas molecules and energetic electrons in the plasma. In most cases, an electron–neutral collision breaks up the neutral molecule into neutral radicals. Being electrically uncharged, radicals reach the substrate with no or only little directionality, bond to the surface, and react more or less readily with the material to be etched to produce more or less volatile compounds. If the reaction takes place instantaneously and the volatility of the reaction products is high enough, the volatile compounds leave the etch surface quickly and take away the material to be etched, yielding isotropic etching behavior. This so-called chemical etch depends much on the properties of the materials involved, yielding high selectivities between different materials and making it in general easy to find appropriate etching masks (e.g., SiO_2 for masking a Si etch).

Fluorine radicals formed in a plasma discharge from fluorine-rich gaseous compounds, such as sulfur hexafluoride (SF_6), carbon tetrafluoride (CF_4), or nitrogen trifluoride (NF_3), show isotropic etching behavior toward silicon. Once adsorbed to the silicon surface, fluorine radicals react immediately with silicon atoms to form silicon fluorides SiF_x ($x = 1,2,3,4$). The reaction products, finally SiF_4, are volatile enough to leave the surface without a need for physical assistance, taking away the silicon material from the etched surface. In contrast to that, etching masks like SiO_2 or photoresist are not or only slowly attacked by fluorine radicals. Since there is no directionality constraint in the mechanism, silicon etching proceeds with no preferential orientation, and the profiles are isotropic. This is an ideal behavior for underetching micromechanical structures in a so-called sacrificial etching or release-etching step for the fabrication of freestanding and moveable elements on top of a wafer. A number of advanced micromachining technologies make use of combinations of anisotropic and subsequent isotropic silicon etching: first etch deep in the vertical direction, then protect the structures and underetch in the lateral direction (de Boer *et al.* 1995, Muenzel 1994). A well-known process example is SCREAM (Shaw *et al.* 1994, Sung and Pang 1993).

For the purpose of silicon sacrificial etching, it is not even necessary to use a plasma as a source of fluorine radicals. A number of gaseous compounds are well known for the ability to etch silicon spontaneously, without a need for plasma excitation. XeF_2 (Chang *et al.* 1995, Flamm 1984), BrF_3 (Wang *et al.* 1997), and ClF_3 (Saito 2002) produce fluorine radicals once adsorbed onto a silicon surface, thus leading to spontaneous isotropic etching of the silicon. The main advantage of plasmaless silicon processing is the very high selectivity toward most nonsilicon materials, because of the pure chemical etch and the lack of physical impact from a plasma discharge, with selectivities, which may well exceed values of 100 000, e.g., for most metals and dielectrics. In a recently published paper, a high selectivity was demonstrated for the first time for selectively etching

a SiGe sacrificial layer with respect to Si functional structures in chlorine trifluoride gas (Leinenbach *et al.* 2007). This finding makes any need for passivation and protection of the functional silicon during the release step obsolete and enables even more challenging MEMS in the future.

1.08.2.2 Anisotropic Etching

The second important kind of species generated from the plasma, less prominent in number but responsible for any directionality and anisotropy of the etch, is the positively charged ions or the so-called cations. They are formed by the ionization of neutral gas molecules by collisions with energetic electrons in the plasma, taking away an electron from the molecule and leaving it back with a positive charge. Being electrically charged, ions can be accelerated toward the substrate by biasing the latter with DC or RF voltage, and by hitting the material in a more or less perpendicular orientation. The ion impact is the only relevant directional contribution provided by the plasma, and all approaches to create anisotropic plasma etching are based on a more or less perpendicular orientation of the ion flow to the surface to be etched. Resulting from their directionality, ions do preferentially hit the bottom of a structure and only to a lesser extent hit their sidewalls. The incoming ions transfer their kinetic energy to the local area of impact. Although the number of ions is small compared to the number of chemically active neutral species, their energetic impact can significantly affect or drive the progress of the etching reaction locally. Ion-assisted directional etching can be caused by one or several of the following mechanisms:

- Induction of chemical reactions between adsorbed radicals from the plasma and the material to be etched by the provision of activation energy with the ion impact, especially if the reactivity of the etching species with the material is low or hindered by an energy barrier.
- Desorption of reaction products from the surface by sputtering action of the incoming ions, especially if the volatility of the reaction products is low and the latter stick to the material surface, forming an inhibiting film there. In this case, ion impact removes the inhibiting film of reaction products from the surface.
- Desorption of other inhibiting films by sputtering action of the incoming ions, e.g., polymer films formed from polymerizing additives to the plasma, or films formed from nonvolatile reaction products between the material to be etched and certain additives to the plasma.

In the case of silicon, typical anisotropic semiconductor etching processes are based on chlorine or bromine, combined with potential additives to the main chemistry for fine-tuning the desired process properties and suppressing the undesired effects. Radicals produced in the plasma from the halogens of lower reactivity, like chlorine and bromine, react slowly with silicon to produce less volatile reaction products, which stick to the surface. Ion impact both activates and induces the formation of silicon chlorides or bromides from silicon and the respective adsorbed halogen radicals, and also removes the inhibitive films of reaction products. In particular, silicon bromides are of very low volatility and need energetic ion bombardment to assist their removal from the silicon surface. In addition to these intrinsic anisotropic properties of the halogens of lower reactivity, oxidation or nitrification of unbombarded silicon surfaces by addition of small amounts of oxygen or nitrogen gas to the plasma can be done to enhance sidewall passivation. As an alternative, sidewall protection is improved by deposition of chlorocarbon or bromocarbon films formed by the addition of chlorocarbon or bromocarbon gases to the plasma, which are ready to polymerize after plasma activation. Ion bombardment breaks up the passivation or the protection of the bombarded surfaces, i.e., the bottom structure, leading to a directional and vertical etching.

Etching based on the halogens of lower reactivity is strongly ion-driven, which gives anisotropy to the process; it, however, limits achievable etch rates (typically below $1\,\mu m\,min^{-1}$), achievable etch depth (a few micrometers), and mask selecitivities (e.g., 20–30:1 for SiO_2 hard mask). Since significant physical impact of the ion flow on the substrate plays a key role in the etch, selectivity as a result of individual chemical properties of the involved materials decreases. There is a fundamental trade-off between increasing the anisotropy by stronger ion bombardment on the one side and maintaining mask selectivity sufficiently high on the other side. In particular, photoresist stability in chlorine- and bromine-based process chemistries is insufficient

for deeper etches, and silicon oxide hard masks must be used in place of or in addition to photoresist (Sung and Pang 1993).

1.08.3 Deep Reactive Ion Etching

Fluorine-based plasma etching processes for silicon are superior in etch rate, achievable etching depth, and mask selectivity, and in particular enable photoresist masking with good selectivity between silicon and photomask removal. However, as already described in this chapter, for the high chemical reactivity and spontaneous etching nature of the fluorine radicals, and for the high volatility of the silicon fluorides as reaction products, the etch should be intrinsically isotropic. Since fluorine radicals do not need ion activation to initiate or enhance their reaction with silicon, nor to remove the reaction products from the silicon surface, anisotropy can be achieved only by the inclusion of sidewall passivation schemes to the process. The existing different DRIE technologies are mainly distinguished by the way sidewall passivation is achieved, which is the key to anisotropy and influences the overall performance of the etch process. In the following, three approaches for deep reactive ion etching (DRIE) are described in detail: the black silicon method, cryogenic dry etching, and the so-called Bosch process with alternating etch and passivation cycles.

1.08.3.1 Black Silicon Method

The black silicon method described in Jansen *et al.* (1995) is a significant improvement over passivation concepts relying on sidewall oxidation in competition with fluorine-based etching of the silicon (Parrens 1981). Fluorine radicals are produced from SF_6 in the plasma discharge and serve as the main etching species, which would readily and isotropically remove silicon wherever it is unprotected and accessible. Oxygen radicals are produced in the plasma in parallel from oxygen gas (O_2) added and serve for oxidation of the silicon surface. Oxidation saturates dangling bonds on the silicon surface and builds up a passivating silicon oxide film, which inhibits fluorine attack of the silicon. The key improvement of the black silicon method consists of the introduction of an oxide scavenging gas, after following a deterministic procedure to detect the best working point of SF_6 to O_2 gas flow ratio, finally to achieve stable anisotropic etching.

To experimentally find out this point of critical balance, the inventors describe a procedure that gave name to the method: the amount of oxygen gas with respect to the amount of SF_6 gas is steadily increased until black silicon (for illustration of black silicon see **Figures 5** and **6**) appears on the etched surface. This blackening of the surface is due to microroughness or microneedles formation, starting at the point of critical passivation. At this point, anisotropy of the etch would be perfect; however, oxide residues at the etch bottom give rise to micromasking effects. To overcome micromasking, a silicon oxide scavenger is added in small amounts to the plasma, e.g., trifluoromethane (CHF_3) gas. The amount of additive is steadily increased until the black silicon surface disappears again. Trifluoromethane is known as an effective compound for ion-assisted etching of silicon oxide, with an oxide removal rate increased by a factor of 5–10 when compared to fluorine radicals (van Roosmalen 1984):

$$2CHF_3 + SiO_2 \rightarrow 2COHF + SiF_4$$

The effect of carbon acting as an attractive reaction partner for oxygen from SiO_2 is a reason for this increased oxide removal rate, compared to fluorine radicals alone.

As a consequence, the ion-bombarded bottom is cleared from the masking oxide and can be etched by the fluorine radicals from SF_6, while the unbombarded sidewall passivation remains unaffected by CHF_3 addition and protects well against fluorine radical attack. Anisotropic etching with vertical sidewalls in silicon can thus be achieved. Photomask selectivity suffers from the presence of both oxygen and fluorine radicals in the plasma, in combination with the energetic ion impact needed to break the bottom silicon oxide. Ion-enhanced photomask erosion by oxygen and fluorine radicals is the undesired consequence. A similar statement holds for SiO_2 mask selectivity, with the silicon oxide mask suffering from the ion-enhanced action of the oxide scavenger together with fluorine radicals.

The black silicon method has potential as a DRIE technique; however, a critical balance has to be maintained between passivants, scavengers, and etchants, which compete with each other in the same process. A high percentage of oxygen flow is needed in the process for effective sidewall passivation, further reducing etch rate and mask selectivity. The process window is narrow.

1.08.3.2 Cryogenic Dry Etching

Cryogenic dry etching is a variation of the passivation technique based on sidewall oxidation. For cryogenic temperatures typically around 175 K (−98°C), which can be realized by liquid nitrogen (LN_2) cooling of the substrates, chemical stability of the inhibiting silicon oxide film in a fluorine-based plasma process is strongly increased, mainly because desorption of the reaction products of silicon oxyfluoride-type from unbombarded surfaces are frozen by the low temperature. Ion bombardment on the etch floor removes the passivating film there by sputtering off most of the chemical compounds. The enhanced oxide film stability at low temperature reduces the flow of oxygen gas required for anisotropic etching to a small percentage of the flow of SF_6, the fluorine radical supply gas. Perfectly vertical etching in silicon can be achieved with a strong overweight of etching fluorine radicals, compared to the passivating oxygen radicals, yielding high etch rates and a relatively wide process window, in comparison to the black silicon method. Impressive results were achieved with cryogenic dry etching in inductively coupled plasma (ICP) tools like the Alcatel reactor, with silicon etch rates reaching 4–6 μm min^{-1}, and mask selectivities of 100:1 for SiO_2 hard masks (Bartha *et al.* 1995). Use of photomasks is difficult, however not impossible if special resist treatment for stress adjustment is performed before the etch, for prevention of cracking and delamination at the low substrate temperatures. The main drawback of the cryogenic etching method is related to the low substrate temperature and the critical dependence of process properties on the latter temperature, given the strong heat impact from the plasma on the substrate and the exothermic etching reaction between silicon and fluorine radicals. In addition, the substrate as the coldest part in the whole configuration acts as a cryopump, freezing out compounds from the plasma, which act as hard-to-remove micromasks on the wafer surface. Micromasking is responsible for the formation of microneedles and micrograss on the silicon surface (black silicon, see **Figures 5** and **6** for illustration), which is observed in most of the deeper cryogenic etches as a disturbing, often unacceptable factor. Nevertheless, cryogenic dry etching has maintained its place as an important silicon microstructuring technique, e.g., in micro-optical applications, where smooth sidewalls on a nanometer scale are of key importance.

1.08.3.3 The Bosch Process

1.08.3.3.1 Process technique

The etching techniques described before make use of relatively hard-to-remove compounds as passivating layers, such as the silicon oxides or oxyfluorides resulting from silicon surface oxidation. Their complete removal from the etch floor without leaving back micromasking residues requires energetic ion impact, eventually in combination with the added scavenging agents. Excessive ion sputtering and the use of scavengers reduce the selectivity toward the masking materials. This is especially true in the case of photoresist masking. Again there is a fundamental trade-off between a clean trench bottom and a high mask selectivity.

To overcome this trade-off and reach both high mask selectivities and a trench bottom free from residues, hard-to-remove passivating compounds must be avoided in the process.

A technique that avoids the formation of such compounds is the deposition of smooth polytetrafluoroethylene (PTFE)- or Teflon®-like films on the silicon surface during etching. Mogab *et al.* (1979) describes the importance of balancing between etching and polymerizing species, and effects of changing this balance for both Si and SiO_2 etching.

Plasma polymerization can be achieved by the generation of $(CF_2)_n$-type radicals in the plasma, starting from precursor gases like trifluoromethane (CHF_3), or preferably hexafluoropropene (C_3F_6) or octafluorocyclobutane (RC318®, C_4F_8), the latter being a nontoxic and stable dimer of tetrafluoroethylene (TFE, C_2F_4). The deposited smooth passivating film consists of a network of long linear $(CF_2)_n$ chains with only little cross-linking, which can be easily removed from the etch floor by low-energetic ion bombardment without leaving back residues. A mixture of SF_6 gas for the fluorine radical supply and C_4F_8 gas for the supply of polymer forming radicals can be used in the plasma to achieve both passivation at the sidewall and etching at the etch floor, yielding anisotropic profiles in silicon. In addition, the presence of fluorocarbons in the process to some extent scavenges the undesired hard passivation from the floor, like silicon oxides from oxygen background gas or oxygen precipitates in the silicon material itself. However, the presence of fluorine radicals as etching species and of Teflon forming monomers as depositing, film-forming species together in the plasma, at the same time, leads to recombination and mutual

extinction of both kinds of species. This makes a mixed process difficult to control for deeper etches and lowers process performance mainly with respect to etch rate. These drawbacks appear more relaxed for plasma sources of very high excitation density like the ICP (see later), but still the potential of the mixed process approach remains limited to shallow trenches with a depth on the order of 10 µm.

The recombination problem was overcome by the patented Bosch process technology (Laermer and Schilp 1992), which is a variation of the Teflon-based sidewall passivation technique. **Figure 2** illustrates the process mechanism. Passivation and etching gases are introduced separately and alternatingly into the process chamber and exposed to a high-density plasma during so-called passivation and etch cycles, respectively. The high-density plasma is generated by microwave surfatron or inductive RF excitation. During each passivation cycle, a thin Teflon-like film is deposited onto the walls of etched structures from C_4F_8 precursor species. Scavenging of some oxides from the silicon etch floor can also happen during or after the Teflon deposition step. During the subsequent etching cycle, part of this film is removed from the coated sidewall by off-vertical ion impact and is driven deeper into the trench, at the same time the trench bottom is cleared from fluorocarbon polymer and etched by fluorine radicals released from SF_6 in the plasma. Switching times between steps normally range from a few seconds up to one minute, depending on the tolerated sidewall roughness. Since the passivating polymer film can be removed by only little ion sputtering impact, mask selectivity reaches very high values, e.g., 150:1 for photoresist and $\gg$300:1 for SiO_2 hard masks. If harder passivating polymers were used, more aggressive ion impact would be required, and as a consequence selectivity toward the masking material would be reduced.

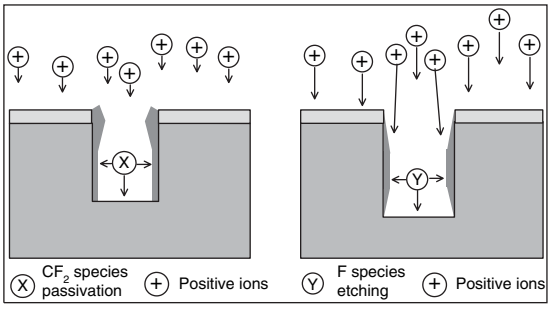


Figure 2 Basic mechanism of the Bosch process. The protecting Teflon®-like film deposited onto the structures during the passivation step is cleared from the trench floor and is driven deeper into the trenches during the (intrinsically isotropic) fluorine-based etching step.

Transport of the sidewall film, i.e., removal and redeposition of Teflon material, makes the etching steps anisotropic, which would otherwise be completely isotropic, in the vicinity of the sidewall, but it does not adversely affect the main etching reaction away from the sidewall. Sidewall roughness is thus reduced despite the discontinuous process. Since plasma polymerization is nonconformal in narrow trenches and film formation preferentially takes place at the trench opening, leaving deeper sidewall regions with less-deposited passivating film, the flow of polymer material along the sidewalls toward the trench floor also provides a more uniform sidewall passivation over the trench depth. **Figure 3** shows typical structures, profile forms, and characteristic sidewall scalloping as the result of standard Bosch DRIE process recipes.

The discontinuous nature of the process overcomes a number of general problems involved with anisotropic etching: firstly recombination loss of reactive species that will not coexist, by gas-phase reactions causing their pairwise extinction, is excluded by separating them in the time domain. This is most important for the volume of the plasma source itself, where the species are generated at high concentration levels and where recombination would be strong. Outside the high-density source area, in a more diluted state, coexistence of species is easier to achieve and some mixing, in the space between the source and the substrate, can be tolerated there. This allows to reduce switching times to below 1 s, e.g., down to 100 ms, keeping passivating and etching species separate in the limited volume of high-density plasma excitation, but overlapping them partly on their way from the source to the substrate. The so-called ultrafast gas-switching conditions yield very smooth sidewalls with hardly any remaining sidewall scalloping at all, as is depicted in the scanning electron microscope (SEM) pictures of **Figure 4**.

Secondly, a vertical etching process having a high mask selectivity bears an inherent risk for micromasking the trench floor. Even minor contamination on the etch surface can lead to formation of residues, microroughness, and even microneedles. Having a sidewall with a periodic nanostructure or ripple on the order of 10 nm in size is an advantage in this respect: micromasks smaller than the characteristic dimension of the scallops are undercut and extinguished from the surface, leaving back no etch residues. Only if the size of a micromask exceeds

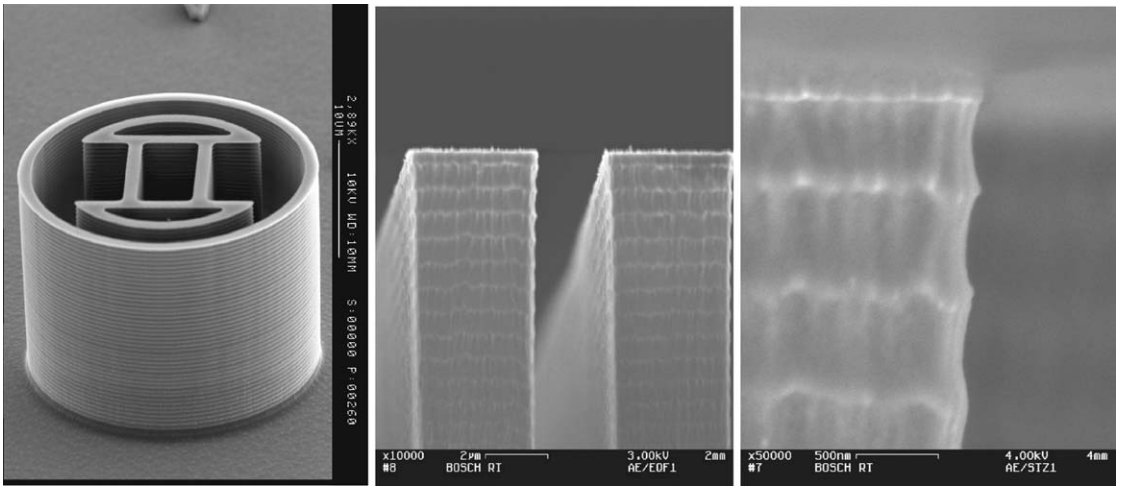


Figure 3 Trench profiles and sidewall structure resulting from a typical Bosch deep reactive ion etching (DRIE) process recipe.

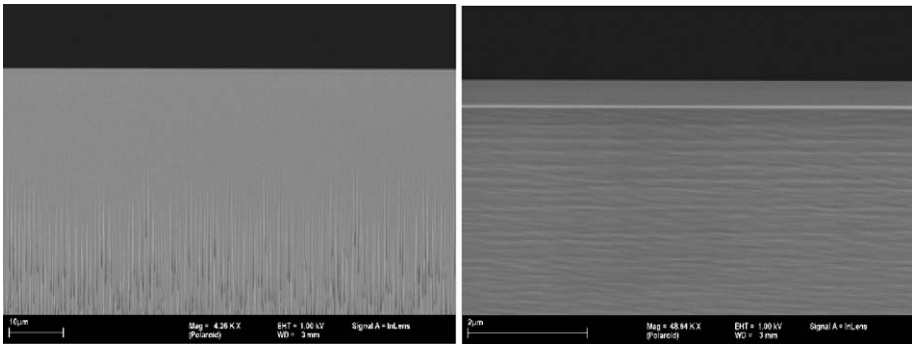


Figure 4 Scanning electron microscope (SEM) pictures of the sidewall resulting from a Bosch process using an ultrafast switching recipe: hardly any horizontal scalloping is visible at the sidewall anymore. Some vertical striations are visible in the lower part of the left-hand side SEM picture, indicating the beginning of microroughness formation at the trench bottom, which does also affect the sidewall.

the critical dimension limit, which the process can tolerate by undercutting, it causes a potential residue problem. This has to be considered in connection with the before-mentioned ultrafast gas switching where the sidewall scalloping is practically gone: the latter process recipes are more critical with respect to the formation of microroughness, resulting from micromasking effects. **Figures 5** and **6** show top views and cross sections (SEM pictures), respectively, of an etching floor with severe microroughness (micrograss or black silicon), resulting from a Bosch DRIE process recipe heavily overloaded with passivation. The excess passivation results in the formation of the characteristic microfeatures, starting from a random distribution of incidental contaminants on the etched surface.

The wide process window, simple control of process characteristics, and the high performance have established the Bosch process as a leading silicon microstructuring technique for MEMS. One of the first papers describing some of the new applications enabled by this technology is given in Klaassen *et al.* (1995). There has been a continuous progress in the technology, with a number of established equipment suppliers offering dedicated equipment tools for a wide range of purposes, from research to mid-volume up to high-volume manufacturing. Starting from a few micrometers per minute in the early days of DRIE, etch rates have meanwhile reached values in excess of 50 µm min^{-1} in silicon, and are even driven further by new requirements coming from through-wafer interconnect and chip-stacking technologies

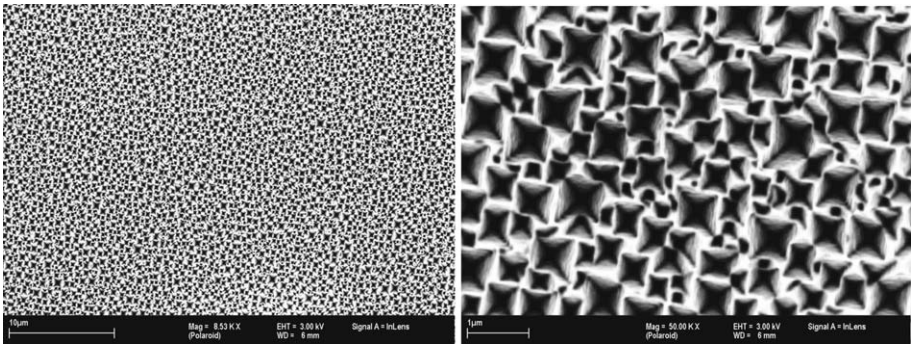


Figure 5 Top-view scanning electron microscope (SEM) pictures of etch bottom with pronounced microroughness (black silicon or micrograss), resulting from a Bosch deep reactive ion etching (DRIE) process heavily overloaded with passivation.

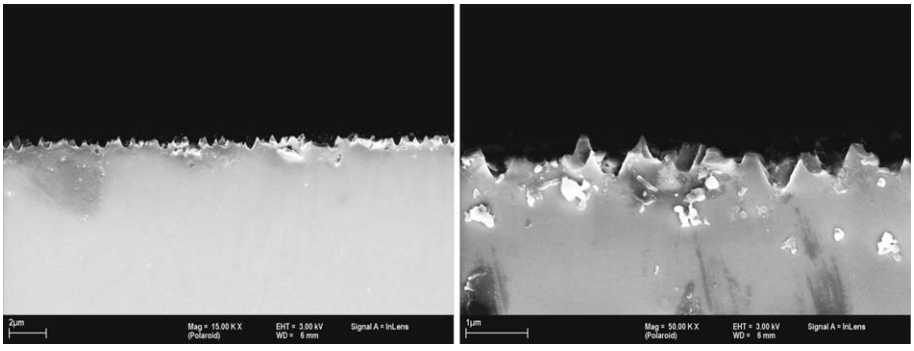


Figure 6 Cross-sectional scanning electron microscope (SEM) pictures of etch bottom with pronounced microroughness (black silicon or micrograss), resulting from a Bosch deep reactive ion etching (DRIE) process heavily overloaded with passivation.

emerging in the semiconductor field. This means that a full silicon wafer can be etched completely from its frontside to the backside within a few minutes. Progress results from the improvement of both process and equipment technology. Having originally started from the microwave surfatron plasma source (a microwave guide in the form of a dielectric tube or a horn, which draws the plasma and the microwave radiation along the guiding structure, see e.g., Laermer and Schilp 1992), the leading equipment for DRIE of today, under industrial conditions, is clearly the ICP source. ICP technology has been developed so far to fulfill all key requirements of the Bosch DRIE approach.

1.08.3.3.2 Plasma equipment for deep reactive ion etching

A high-density plasma tool with so-called remote or decoupled plasma is necessary to make the DRIE technology work, which is mainly based on the chemical activity of the plasma. Chemical activity means firstly to reach sufficiently high concentration of polymerizing radicals for strong sidewall passivation, secondly to reach sufficiently high concentration of fluorine radicals for high etch rates. For a high density of chemically active species, a process pressure regime of typically between 1 and 10 Pa is the most appropriate. Energy of the ions flowing to the wafer surface must be controlled independently of the plasma excitation, through substrate electrode biasing. The latter can be done by a RF or low-frequency bias generator. For a good control of ion acceleration independent of plasma excitation, the plasma potential should be kept sufficiently low, as close as possible to the ground potential. The latter requires decoupling of the plasma excitation zone from the substrate region.

Traditional plasma sources for RIE fail in one or more of these requirements. Diode and triode reactors show insufficient plasma excitation density, below $10^{10}\,\mathrm{cm}^{-3}$. Electron cyclotron resonance (ECR) sources are inappropriate for the very

low-pressure regime (<0.1 Pa) tolerated by the cyclotron resonance mode. ECR is a typical ion current type of source and lacking sufficient chemical activity, mainly for the low-pressure argument. Microwave surfatron sources (a tube or a horn made from dielectric material to guide the microwave field along the interface between the dielectrics and the plasma burning inside the tube or the horn, down to the process chamber, see Laermer and Schilp 1992), helicon sources, and ICP sources open the pressure window up to 10 Pa and even higher pressures. Amongst these alternatives, the ICP emerged as the most appropriate source technology for this kind of processing and as the industry standard. **Figure 7** shows a schematic of a simple configuration of an ICP wafer processing tool.

In principle, an ICP is generated inside a magnetic coil, which is driven by RF power. In most cases, RF of 13.56 MHz is used for driving the ICP, but also 2 MHz and lower frequencies can be found in some places. The coil of one or several turns, depending on the chosen frequency, is wound around a dielectric vessel. By feeding RF power to the coil, a high-frequency oscillating magnetic field is generated inside the coil, with a corresponding oscillating electric field resulting from Maxwell's induction law. The electric field can excite and drive a high-density plasma within the dielectric vessel. Since the plasma is not able to shield itself effectively from this inner driving electric field by the formation of sheaths, extremely high densities of charged particles (electrons and ions $\sim 10^{12}$–10^{14} cm^{-3}) can be provided by the inductive technology. The electrons are heated by the RF electric field and generate the ions and chemically active species by collisions with gas molecules. Before the plasma of charged and neutral species reaches the substrate, the hot electrons need to cool down and as a rule the distributions of ions and neutrals need some kind of readjustment and rearrangement for homogenization of etching results. Mechanical apertures (Laermer *et al.* 1999), gas flow guides, and gas flow redirectors, in combination with balanced excitation of the inductive coil (Becker *et al.* 1999) and magnetic lenses for plasma collimation (Breitschwerdt *et al.* 2000), are useful and found in this place.

RF power levels used today have reached values of 5 kW at 13.56 MHz for the high etch rate conditions, imposing severe challenges on the stability of RF matching mimics (capacitors, variocaps, inductors, BALUN cables, etc.) and on the decoupling, cooling, and homogenization of the plasma at the substrate location. **Figure 8** shows a scheme of a typical advanced remote ICP tool configuration, as an example for illustration.

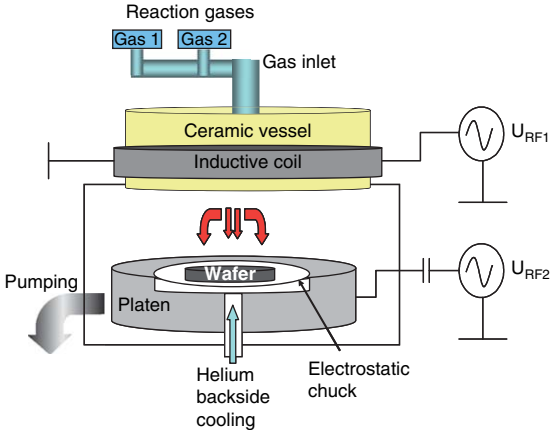


Figure 7 Schematic of an inductively coupled plasma (ICP) tool. Radio frequency (RF) generator 1 powers the inductive coil, wound around the dielectric reactor vessel, to generate a high-density plasma within the reactor. Mostly ceramic materials are used for the vessel. RF generator 2 powers the substrate electrode for biasing and accelerating ions to the wafer. Electrostatic clamping and helium backside pressure serve to cool the wafer during processing.

1.08.4 Special Issues Related to Bosch DRIE

1.08.4.1 Control of RIE Lag

RIE lag is a common and well-understood transport phenomenon in plasma etching, see the review given in Gottscho *et al.* (1992). Etching speed and profile evolution with etching progress depend on the depth:width ratio of the trench, the so-called aspect ratio. In general, the etch in narrower trenches progresses at a lower speed than in wider trenches, i.e., higher aspect ratio trenches lag behind the lower aspect ratio trenches. The amount of the lag effect depends mainly on the mean free path of the gaseous species, which determines their transport in etched trenches, compare, for example, with the conductance model described in Coburn and Winters (1989). In addition, interaction and change of balance between etching and passivating species during their transport in etched trenches are important, which is again related to mean free path and process pressure. Changes in the balance of etching and passivating species with increasing aspect ratio also affect

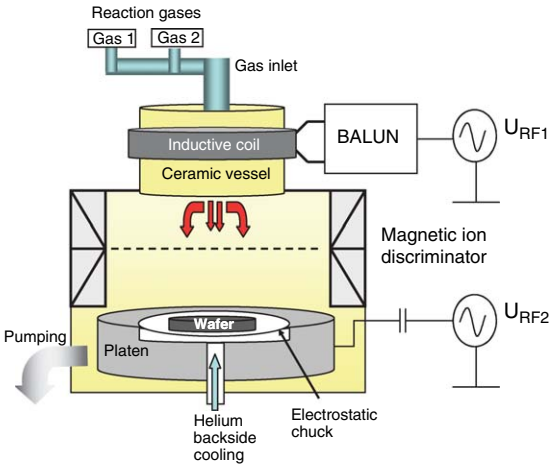


Figure 8 Advanced decoupled high-density plasma tool for deep reactive ion etching (DRIE). Inductive coil symmetrically powered by RF1 from BALUN transformer generates the high-density plasma within the dielectric reactor vessel. Drift zone with magnetic ion discriminator: decoupling of the plasma and homogenization of neutrals flow, and readjustment of spatial ion distribution, e.g., through magnetic lenses, respectively. Application of RF2 power to the substrate electrode for biasing and accelerating the ions toward the wafer substrate. Electrostatic chuck (E-chuck) and He backpressure serve for wafer cooling.

anisotropy and profile form. Figure 9 illustrates a characteristic example for RIE lag in a typical Bosch process.

In a discontinuous process with separate etching and passivating cycles, which can be controlled independently from each other, there is an individually controllable lag effect both for the etching and for the passivating part of the overall process. For the etch step, the higher the aspect ratio of the trench the less the number of ions and fluorine radicals that can reach the trench floor per time unit and etch the silicon, the size of the effect depending mainly on the etch step pressure. For the deposition step, the higher the aspect ratio of the trench the less the polymer formers that can reach the trench floor per time unit and build up an inhibiting film on the floor, the size of the effect depending again mainly on the deposition step pressure. Since a thinner inhibiting film on the etch floor increases the rate for the subsequent etch step, this involves some compensation for the rate decrease with aspect ratio, resulting from the reduced fluorine radical supply. Deposition step lag and etching step lag can compensate each other to reach a lag-free overall process over a wide range of aspect ratios. This is achievable by individual control of the etch step and deposition step parameters, mainly by the pressure values set for the different process steps.

Setting the pressure higher for the individual step increases its lag effect and setting the pressure lower decreases its lag effect. Increasing the deposition step pressure with respect to the etch step pressure, even to values higher than the etch step pressure, reduces the overall RIE lag of the total process and even inverts it at some stage, with higher aspect ratio trenches, i.e., narrower gaps, etching faster than smaller aspect ratio trenches, i.e., wider gaps. However, compensation for RIE lag is always at the expense of a reduced etch rate in the wider trenches, which approaches the low etch rate in the narrow trenches.

On the other hand, running a deposition step at a pressure lower than the etch step pressure will cause a significant overall RIE lag of the total process, however at the benefit of higher overall etching rates in the wider trenches. This is the normal process regime,

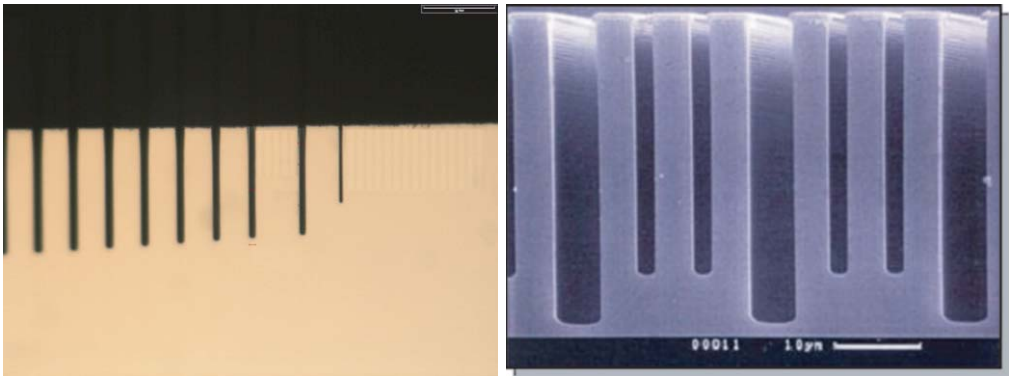


Figure 9 Examples for illustrating the reactive ion etching (RIE) lag effect. Etch rate and etch depth decrease with narrower gaps corresponding to higher aspect ratios.

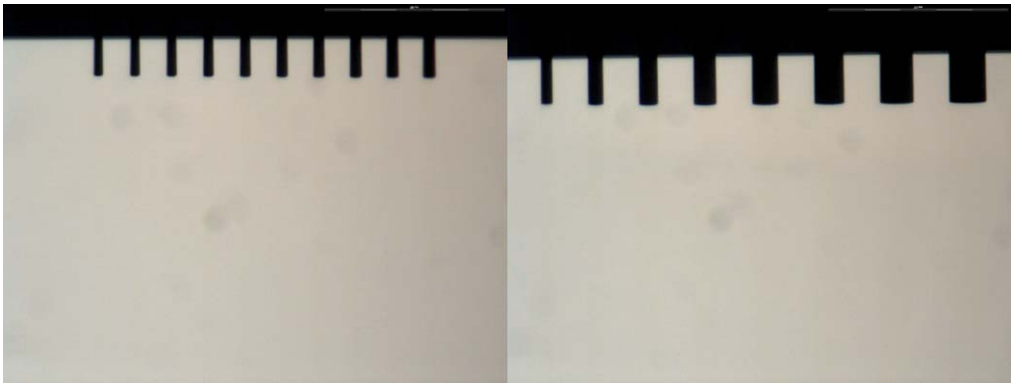


Figure 10 Example to illustrate reduction of reactive ion etching (RIE) lag. Etch depth varies moderately with trench width by a higher-order functional dependence.

which is used in most cases, in the interest of higher rates. **Figure 10** shows some results from a Bosch process modified for reduced RIE lag effect.

Since the stability of the polymer deposited onto the etch floor is an important weight factor in balancing etch and deposition lags, the amount of achievable compensation depends also on the substrate temperature. Lowering the substrate temperature makes the floor polymer more robust and increases the weight of the deposition step in the overall lag balancing. Hence, a substrate temperature of 0°C or even lower, instead of 40°C (under etching conditions, i.e., including heat flow from the plasma to the wafer), was found beneficial to achieve lag-free etching over a wide range of aspect ratios (Becker *et al.* 1998).

1.08.4.2 Parameter Ramping for High Aspect Ratio Etches

As mentioned before, in the fabrication of high aspect ratio trenches, different lag effects for etch and deposition steps yield a changing net etch rate with increasing aspect ratio. In general, for a normal process without the above-described compensation for RIE lag, e.g., for rate arguments, net etch rate lowers with increasing aspect ratio of the trenches. There is a changing balance between the passivating and the etching species, and a trend toward more passivation with the higher aspect ratios is reached during the etch. In addition, the amount of ions reaching the trench floor slowly diminishes due to the aperture effects at the trench opening, in combination with the angular distribution of the ions. As a consequence, profiles tend to develop positive slopes and trenches end up tipping when their aspect ratios typically reach values greater than 20:1.

Given that all trenches have a similar range of widths, the following procedure is adopted. During the progress of the etch, the process recipe is adapted stepwise or continuously (Hopkins *et al.* 1999, Hynes *et al.* 1999), e.g., to reduce the amount of passivation in the process. For example, passivation cycle time or passivation gas flow can be ramped down steadily. As an alternative, etching cycle time or etching gas flow can be ramped up steadily. Although other options like source power, bias power, or pressure ramping do exist, changing cycle times is the most obvious parameter ramping strategy to affect the etch:deposition balance in a predictable and straightforward manner. By using parameter ramping, trenches with aspect ratio >50:1 can be fabricated, while keeping the sidewalls straight and vertical.

1.08.4.3 Notching at Dielectric Interfaces

The notching effect is relevant for all cases where the silicon etch terminates on a dielectric insulating layer. This can be an etch stop for silicon oxide e.g., in surface micromachining, or for a polymer material like photoresist or adhesive foils in through-wafer etching. In surface micromachining, a silicon film is structured on top of a silicon oxide interlayer. This oxide interlayer serves both as a sacrificial layer to be removed later during a so-called release etch in hydrofluoric acid, to produce freestanding silicon structures, and as an electrical insulator to separate different structural parts from each other and from the substrate wafer.

When the dielectric etch stop is reached and an overetch takes place, charging of the trenched pattern features occurs (Arnold and Sawin 1991).

Charging of the dielectric trench floor with respect to the silicon sidewalls is a consequence of the different angular distributions of the predominant negatively charged species, the electrons, with no or only little directionality, and of the positively charged species, the ions, which reach the substrate with an orientation perpendicular to the surface. Thanks to their directionality, a large fraction of the incoming ions reach the dielectric bottom of a high aspect ratio trench, and a part of them get stuck and accumulate there. In contrast to the ion behavior, most of the incoming electrons get trapped by the structural sidewalls. Strong electric fields build up between the positively charged dielectric bottom and the negatively charged silicon sidewalls, redirecting incoming ions to the interface area between silicon and the dielectric layer. **Figure 11** illustrates this charging phenomenon.

Notching is a well-known consequence of the charging effect in high aspect ratio trenches (Hwang and Giapis 1997, Kinoshita *et al.* 1996, Nozawa *et al.* 1994). Where the redirected ions hit the structure sidewalls, deep notches are cut into the silicon near the dielectric interface. **Figure 14(a)** illustrates a typical example after overetching at the dielectric etch stop. The notches affect the mechanical properties of the fabricated silicon structures in an unpredictable way and represent a serious particle source in manufacturing. Many development efforts have been put into the suppression of notching for its strong practical relevance.

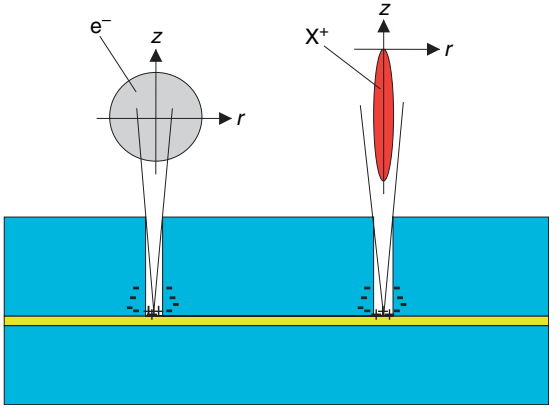


Figure 11 Charging of high aspect ratio trenches. Silicon sidewalls take up electrons with little or no directionality. The dielectric trench floor traps positively charged ions arriving at an orientation perpendicular to the surface. Strong electric fields build up between the accumulated positive charges on the dielectric bottom and the negative charges in the silicon material.

Since charging and the resulting electric fields are the driving forces behind notching, all technical solutions aim at a reduction of charge separation in high aspect ratio trenches. Pulsing the substrate bias as depicted in **Figure 12** gives accumulated charges in the structure a chance to neutralize. Bias breaks with no or little ion impact between the bias bursts with a strong ion impact must be long enough in relation to a characteristic time constant for the leakage currents to discharge. Pulsing is feasible for both RF (13.56 or 2 MHz) and LF (380 kHz) biasing. Low-frequency biasing has some intrinsic advantages with respect to less charging, because both electrons and ions are able to follow the (slowly) alternating electric field, the latter despite their heavy mass. As a consequence, in contrast to high-frequency biasing, the electrons are more directed toward the substrate and reach the dielectric trench bottom more easily. This results in a slower and somewhat reduced buildup of charges at the dielectric interface, which is further reduced by pulsing the bias voltage. The disadvantage of using low-frequency biasing, in contrast to higher frequencies, is that ion energies are less well defined and cover a wide spectrum of energies. The latter makes process control more difficult to achieve. In addition, matching units for low-frequency biasing are not as well established and stable as for biasing at the high-frequency standard of 13.56 MHz. Typical pulse repetition frequencies are in a range of 30–300 Hz, with duty cycles (pulse on per period time) between 5% and 50%.

A more powerful solution for the suppression of notching is described in Laermer (2001), which represents a combination of a sophisticated pulsing scheme for the substrate bias power and pulsing of the plasma source power. **Figure 13** illustrates this

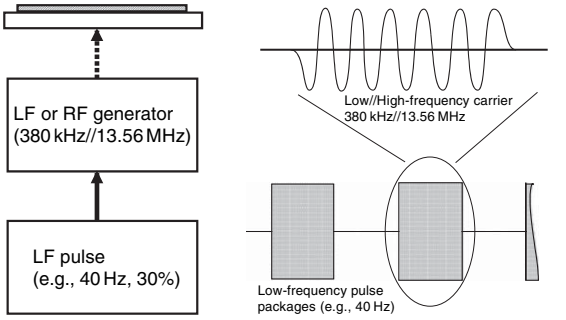


Figure 12 Suppression of notching by pulsed substrate bias power. Radio frequency (RF) substrate biasing (low, mid, or high frequency) is modulated by a train of short pulses, e.g., at 40-Hz repetition rate.

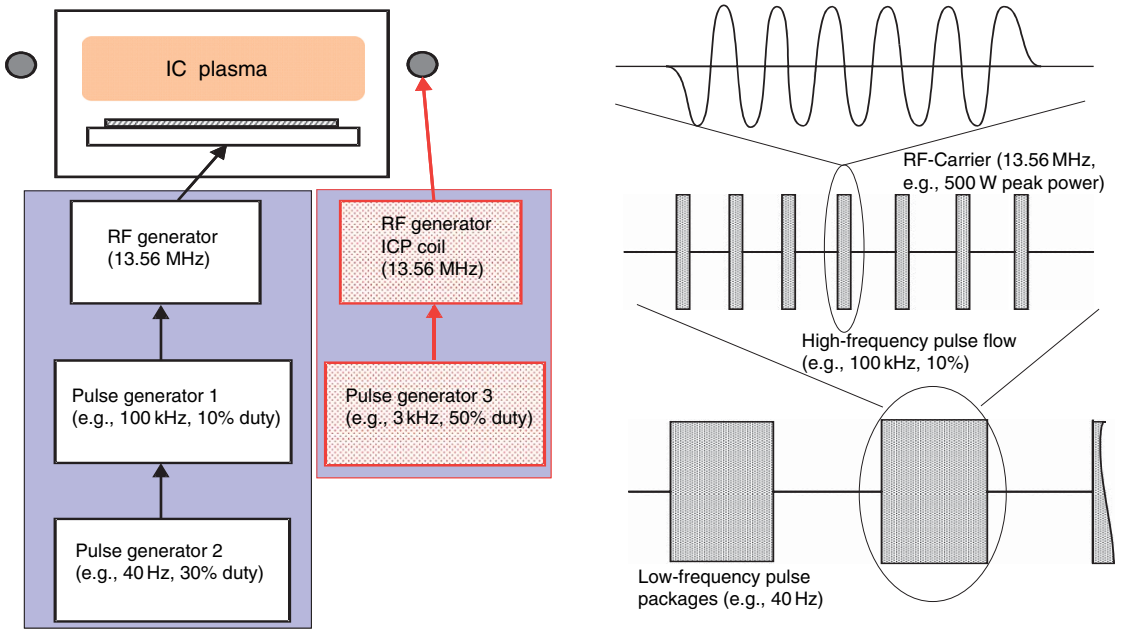


Figure 13 Antinotching scheme for combined pulsing of plasma source power and substrate bias power. *Left*: Hardware configuration for pulsed plasma source and substrate electrode. *Right*: Pulsing scheme at the substrate electrode: radio frequency (RF) carrier (13.56 MHz) is modulated by a fast train of short pulses (100 kHz) combined with a slow pulse modulation (40 Hz).

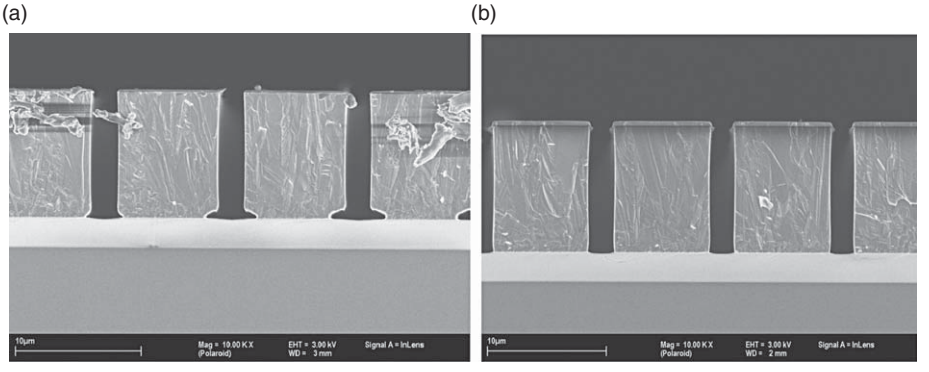


Figure 14 (a) Notching phenomenon from overetch at the dielectric etch stop of SiO_2, without appropriate countermeasures. (b) Notch-free etch stop on SiO_2 after 100% overetch, with the above-described countermeasures taken.

method, both the hardware configuration and the pulse train at the substrate electrode.

Switching off the plasma source power periodically yields peaks of negatively charged ions (anions), which are generated during the afterglow phases when the plasma breaks down and the cooling electrons are caught by neutral gas molecules. In contrast to the undirectional electrons, the anions do reach and effectively neutralize positively charged ions at the dielectric trench floor during substrate bias power breaks, i.e., when the substrate is not repulsive toward negative charges. Typical repetition rates for the plasma power are between 1 and 5 kHz, at a duty cycle of about 50%. Repetition rates for the substrate bias power are, for example, 100 kHz for the fast pulse train and 40 Hz for the slow pulse train. The depicted substrate biasing scheme is an attempt to emulate low-frequency biasing conditions into RF biasing, with higher peak voltage levels and lower average currents to the substrate electrode, and a reduced charge buildup resulting from the fast modulated pulse train at low duty cycle. The goal is

to emulate some of the advantages of using low-frequency biasing, without incorporating the disadvantages. Note that since the carrier frequency is still at 13.56 MHz, a standard RF matching unit can be utilized.

Extremely strong overetch is tolerated without notching using this approach. **Figure 14(b)** illustrates an example of a 100% overetch on silicon dioxide.

1.08.5 Conclusions

The Bosch process has emerged as the standard DRIE technology for microstructuring of silicon MEMS. Performance and flexibility, freedom of design independent of crystal orientation, IC compatibility, and the usability of standard masking materials like photoresist are the relevant success factors. DRIE enables the fabrication of various MEMS structures and fascinating new designs (Senturia 2001). From history, geometries like that were first demonstrated to the MEMS community by the LIGA technique in the 1980s, which attracted a lot of attention and caused enthusiasm for 3D microstructuring at the time. In contrast to LIGA, the DRIE technology is virtually what the MEMS community was dreaming of: reasonably low cost and close to semiconductor industrial standards.

Progress is driven by process and hardware evolution, and new upcoming requirements like through-wafer interconnects requesting even higher rates and accuracy. Developments are nowadays focused into more and more specialized directions. This is strongly supported by an established equipment supplier base defining their individual core competencies in dedicated fields, and covering the full scope of requirements from R&D level into high-volume mass manufacturing.

References

Arnold J C, Sawin H H 1991 Charging of pattern features during plasma etching. *J. Appl. Phys.* **70**, 5314–17

Bartha J W, Greschner J, Puech M, Mauquin P 1995 Low temperature etching of Si in high-density plasma using SF_6/O_2. *J. Microelectron. Eng.* **27**, 453–6

Bean K E 1978 Anisotropic etching of silicon. *IEEE Trans. Electron Devices* **ED-25**, 1185–93

Becker V, Laermer F, Schilp A, Beck T 1999 Plasma etching installation. *US Pat*. 6 531 031

Becker V, Laermer F, Schilp A 1998 Anisotropic plasma etching of trenches in silicon by control of substrate temperature. *German Pat*. DE 19841964, GB-2 341 348

de Boer M, Jansen H, Elwenspoek M 1995 Black silicon method V: A study of the fabricating of moveable structures for micro electromechanical systems. *Proc. Transducers '95,* Stockholm, Sweden, pp. 565–8

Breitschwerdt K, Becker V, Laermer F, Schilp A 2000 Device and method for etching a substrate by using an inductively coupled plasma. *US Pat*. 6 709 546 and 7 094 706

Chang F I, Yeh R, Lin G, Chu P B, Hoffman E, Kruglick E J, Pister K S J, Hecht M H 1995 Gas-phase silicon micromachining with xenon difluoride. *Proc. SPIE Microelectronic Structures and Microelectromechanical Devices for Optical Processing and Multimedia Applications*, Austin, TX, USA, Vol. 2641, pp. 117–28

Chapman B 1980 *Glow Discharge Processes*. John Wiley & Sons, New York

Coburn J W, Winters H F 1989 Conductance considerations in the reactive ion etching of high aspect ratio features. *Appl. Phys. Lett.* **55**, 2730–2

Diem B, Rey P, Renard S, Viollet-Bosson S, Bono H, Michel F, Delaye M T, Delapierre G 1995 SOI 'Simox' from bulk to surface micromachining. *Sens. Actuators* **A46-7**, 8–16

Ehrfeld W 2002 *Handbuch Mikrotechnik*. Carl Hanser Verlag, München/Wien

Flamm D L 1984 Plasmaless dry etching of silicon with fluorine-containing compounds. *J. Appl. Phys.* **56**, 2939–42

Gottscho R A, Jurgensen C W, Vitkavage D J 1992 Microscopic uniformity in plasma etching. *J. Vac. Sci. Technol. B* **10**, 2133–47

Hopkins J, Ashraf H, Bhardwaj J K, Hynes A M, Johnston I, Shepherd J N 1999 The benefits of process parameter ramping during plasma etching of high aspect ratio silicon structures. *Mater. Res. Soc. Symp. Proc.* **546**, 63–8

Hwang G S, Giapis K P 1997 On the origin of the notching effect during etching in uniform high density plasmas. *J. Vac. Sci. Technol. B* **15** 1997, 70–87

Hwang G S, Giapis K P 1998 Mechanism of charging reduction in pulsed plasma etching. *Jpn. J. Appl. Phys.* **37**, 2291–301

Hynes A M, Ashraf H, Bhardwaj J K, Hopkins J, Johnston I, Shepherd J N 1999 Recent advances in silicon etching for MEMS using the ASE-process. *Sens. Actuators A* **74**, 13–17

Jansen H, de Boer M, Legtenberg R, Elwenspoek M 1995 The black silicon method. *J. Micromech. Microeng.* **5**, 115–20; *Proc. Int. Conf. Micro- and Nano-Engineering '94,* Davos, Switzerland, 1995, pp. 475–80, ISSN 0167-9317

Juan W H, Weigold J W, Pang S W 1996 Dry etching and boron diffusion of heavily doped, high aspect ratio Si trenches. *Proc. SPIE* **2879**, 45–55

Kinoshita T, Hane M, McVittie J P 1996 Notching as an example of charging in uniform high density plasmas. *J. Vac. Sci. Technol. B* **14**, 560–5

Klaassen E H, Petersen K, Noworolski J M, Logan J, Maluf N I, Brown J, Storment C, McCulley W, Kovacs G T A 1995 Silicon fusion bonding and deep reactive ion etching: A new technology for microstructures. *Proc. Transducers 1995,* Stockholm, Sweden, June 1995, pp. 556–9

Kloek B, Collins S D, de Rooij N F, Smith R L 1989 Study of electrochemical etch-stop for high precision thickness control of silicon membranes. *IEEE Trans. on Electron. Devices* **36(4)**, 663–9

Laermer F 2001 Process for etching structures in an etching body using a plasma comprising modulating the intensity of the plasma as a function of time. *US Pat. Appl*. 2004/0173571

Laermer F, Schilp A 1992 Method for anisotropically etching silicon. *German Pat. DE*-4241045, *US Pat*. 5 501 893

Laermer F, Schilp A, Funk K, Offenberg M 1999 Bosch deep silicon etching: Improving uniformity and etch rate for advanced MEMS applications. *Proc. IEEE MEMS 1999,* Orlando, FL, USA, pp. 211–16

Leinenbach C, Seidel H, Fuchs T, Kronmueller S, Laermer F 2007 A novel sacrificial layer technology based on highly selective etching of silicon–germanium in ClF_3. *Proc. IEEE MEMS '2007,* Kobe, Japan, pp. 65–8

Li Y X, French P J, Sarro P M, Wolffenbuttel R F 1995 Fabrication of a single crystalline silicon capacitive lateral accelerometer using micromachining based on a single step plasma etching. *Proc. IEEE Electro Mechanical Systems Conference*, Amsterdam, The Netherlands, IEEE, pp. 398–403

Manos D M, Flamm D L 1989 *Plasma Etching: An Introduction*. Academic Press, Boston

Mayer G K, Offereins H L, Sandmaier H, Kuehl K 1990 Fabrication of non-underetched convex corners in anisotropic etching of (100)-silicon in aqueous KOH with respect to novel micromechanic elements. *J. Electrochem. Soc.* **137**, 3947–50

Menz W, Mohr J 1997 *Mikrosystemtechnik für Ingenieure*. VCH-Verlag, Weinheim , ISBN 3-527-30536-X

Mogab C J, Adams A C, Flamm D L 1979 Plasma etching of Si and SiO_2: The effect of oxygen addition to CF_4-plasmas. *J. Appl. Phys.* **49**, 3796–803

Muenzel H, Benz G, Schilp A, Laermer F 1994 Processing silicon in a plasma etch system (fabrication of free-standing silicon structures). *German Pat*. DE4420962, GB2 290 413

Nozawa T, Kinoshita T, Nishizuka T, Narai A, Inoue T, Nakaue A 1994 *Dry Process Symp.* **I-8**, 37–41

Oehrlein G S 1990 Reactive ion etching. In: Rossnagel S M (ed.) *Reactive Ion Etching: Handbook of Plasma Processing Technology.* Noyes, Park Ridge, NJ, pp. 196–206

Palik E D, Faust J W, Gray H F, Green R F 1982 Study of the etch-stop mechanism in silicon. *J. Electrochem. Soc.* **129**, 2051–9

Parrens P 1981 Anisotropic and selective etching of polysilicon using SF_6. *J. Vac. Sci. Technol.* **19**, 1403–7

Petersen K E 1982 Silicon as a mechanical material. *Proc. IEEE* **70(5)**, 420ff

Robbins H, Schwartz B 1976 Chemical etching of silicon. *J. Electrochem. Soc.* **123**, 1903–9

van Roosmalen A J 1984 Review on dry etching of silicon oxide. *Vacuum* **34**, 429–36

Saito Y 2002 Characteristics of plasmaless dry etching of silicon-related materials using chlorine-trifluoride gas. *Sens. Mater.* **14(5)**, 231–7

Sandmaier H, Offereins H L, Kuehl K, Lang W 1991 Corner compensation technique in anisotropic etching of (100)-silicon using aqueous KOH. *Proc. 6th Int. Conf. Solid-State Sensors and Actuators (Transducers '91)*, San Francisco, CA, USA, pp. 456–9

Schwartz G C, Schaible P M 1979 Reactive ion etching of silicon. *J. Vac. Sci. Technol.* **16**, 410–13

Seidel H, Csepregi H, Heuberger A, Baumgartner H 1990 Anisotropic etching of crystalline silicon in alkaline solutions. *J. Electrochem. Soc.* **137**, 3612–32

Senturia S D 2001 *Microsystem Design*, 2nd edn. Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 71

Shaw K A, Zhang Z L, MacDonald N C 1994 SCREAM I: A single mask, single-crystal silicon, reactive ion etching process for microelectromechanical structures. *Sens. Actuators A* **40**, 63–70

Sung K T, Pang S W 1993 *J. Vac. Sci. Technol. A* **11**, 1206–10

Wang X-Q, Xang X, Walsh K, Tai Y-C 1997 Gas-phase silicon etching with bromine trifluoride. *Proc. Transducers '97*, Chicago, IL, USA, Vol. 2, pp. 1505–08

Wu X P, Ko W H 1989 Compensating corner undercutting in anisotropic etching. *Sens. Actuators* **18**, 207–15, ISBN 3-446-21506-9

Biography

Franz Laermer was born in Waldsassen, Germany, in 1960. He studied physics and mathematics at the Technical University Munich, Germany, and physics at the Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, from 1980 to 1986. He obtained a Diploma in Physics from ETH Zurich, Switzerland, in 1986 and a Ph.D. from the Technical University Munich, Germany, in 1989. He joined Corporate Research and Technology Center of Robert Bosch GmbH, Stuttgart, Germany, in 1990. Laermer worked on new microstructuring technologies and new silicon sensors for automotive applications, mainly surface micromachined accelerometers, gyroscopes, and pressure sensors. He is an inventor of the Bosch deep reactive ion etching (DRIE) process and of many patents in the fields of Bosch DRIE, silicon micromachining, and silicon microsensors. Since 2003, he has been a director of an innovation project for new microsystems, targeting new application fields outside the automotive industry. Together with his co-worker Mrs Andrea Urban, he was awarded by the European Patent Office and the European Commission with the prize 'European Inventor of the year 2007 – Industry', for the invention and the sustainable success of the Bosch DRIE process for MEMS.

1.09 Wafer Bonding

Khalil Najafi, Timothy J. Harpster, Hanseup Kim, Jay S. Mitchell, and Warren C. Welch III,
Center for Wireless Integrated Microsystems (WIMS), University of Michigan, Ann Arbor, MI, USA

Published by Elsevier B.V.

Glossary

BCB Benzocyclobutene
CTE Coefficient of Thermal Expansion
CVD Chemical Vapor Deposition
DBD Dielectric Barrier Discharge
DC Direct Current
DI Deionized
DWB Direct Wafer Bonding
DXRL Deep X-Ray Lithography
EMIH Electromagnetic Induction Heating
FIB Focused Ion Beam
GHz Giga Hertz
HF Hydrofluoric acid
IC Integrated Circuit
ICP Inductively Coupled Plasma
IPA Isopropyl Alcohol
LPCVD Low-Pressure Chemical Vapor Deposition
MEMS Microelectromechanical Systems
MPa Megapascals
MWB Mediated Wafer Bonding
PECVD Plasma-Enhanced Chemical Vapor Deposition
PMMA Polymethylmethacrylate
RF Radio Frequency
RTP Rapid Thermal Processing
SAB Surface-Activated Bonding
SEM Scanning Electron Microscope
Slpm Standard Liters Per Minute
SOG Silicon on Glass
SOI Silicon On Insulator
UHV Ultrahigh Vacuum
VCSEL Vertical Cavity Surface-Emitting Laser
W Watts

1.09.1 Introduction

Wafer bonding has been an important and critical technology in the development of micromachined sensors, actuators, microsystems, and their packaging for many decades (Alexe and Gösele 2004, Esashi 1995, Goesele *et al.* 1999, Iyer and Auberton-Herve 2002, Ko *et al.* 1985, Najafi 2001–2006, *Philips J. Res., Spec. Issue Direct Bonding* 49, 1, Schmidt 1998, Tong and Gösele 1999). It is also being used in the fabrication of integrated circuits (ICs) and the formation of composite materials and wafers needed in advanced circuit technologies (Alexe and Gösele 2004). Wafer, or die, bonding refers to the process whereby two or more wafers of similar or dissimilar materials are, often permanently, attached or bonded together. The individual wafers might have already been through previous fabrication steps to form various features on them, or might just be plain wafers.

Wafer bonding has been used for a variety of applications in both microelectromechanical systems (MEMS) and IC areas, including:

- *Sensor/actuator and MEMS fabrication.* The earlier application of wafer bonding in MEMS was for the fabrication of micromachined sensors, specifically pressure sensors (Ko *et al.* 1985), where its primary use was to provide a supporting substrate for the sensor die. It was also used for the fabrication of capacitive pressure sensors where it formed one of the sensing electrodes of a capacitor. The use of bonding for MEMS fabrication has grown significantly over the past 30 years and has become a major technology, which is an indispensable tool in both research and commercial activities.
- *Packaging and assembly.* As important as it has been for MEMS fabrication, wafer/die bonding has also been equally important and critical for packaging and assembly of MEMS and ICs. Whether it was used for the attachment of individual IC die to its package housing or for bonding pressure sensor wafers to a support glass substrate, bonding and die attachments have remained a critical aspect of packaging and assembly in both MEMS and IC industries.
- *Fabrication of 3D or high aspect ratio microstructures and complex microsystems.* Wafer bonding is also now widely used in the fabrication of multilayer 3D and high aspect ratio microstructures and microsystems. In these structures, the creation of large 3D features is possible only using wafer bonding since most planar fabrication techniques used for the fabrication of MEMS are not capable of producing either large and high aspect ratio structures or multilayer structures that have embedded or buried features.
- *IC and semiconductor device fabrication.* Die bonding and attachment have long been used for packaging and assembly of ICs. Recently, wafer bonding has also become a very important part of IC and semiconductor device fabrication. In IC fabrication, the main application of wafer bonding has been for the fabrication of three-dimensionally stacked circuits whereby multiple layers of silicon circuits are directly formed on top of the existing layers using integrated feedthroughs and interconnects. This is an important research area in both academia and industry that could solve the critical challenge of packing more transistors in a given area or volume without having to further reduce feature sizes. In the area of semiconductor device fabrication, wafer bonding has also become an indispensable technology since it allows the bonding of dissimilar materials like silicon and GaAs, thus allowing the fabrication of heterogeneous devices and circuits, which combines the best features of two or more technologies or materials without sacrificing the performance. Thus, some devices can be fabricated on one wafer using a given technology, and other devices on another wafer using another technology, and then the two wafers can be bonded to interconnect the devices and complete the system without disturbing or interfering with the individual processes.
- *Fabrication of composite wafers or materials.* Finally, wafer bonding has become a key technology in fabricating wafers that provide new capabilities and materials for both the IC and MEMS industries. Silicon-on-insulator (SOI) wafers are now routinely fabricated much more economically using wafer bonding rather than using previous methods such as oxygen implantation and silicon recrystallization. In addition, wafers consisting of thin films of a variety of semiconductor materials bonded to a host silicon wafer are also now in widespread use in microelectronics, optoelectronics, and nanotechnologies. In fact, wafer bonding is now considered as one of the most important and critical technologies for the future of the IC industry, and significant resources are dedicated to it to develop its technology into a high-volume and low-cost technology for the IC industry.

This chapter briefly reviews different wafer bonding techniques and discusses their advantages and disadvantages. Although the focus of the chapter focus is on the bonding of two wafers, the same wafer bonding

techniques are also applied for the attachment or bonding of small pieces and die. In fact, many wafer bonding techniques were adapted from earlier techniques developed for die attachment and IC die packaging. Wafer bonding is a very broad field and it is not possible to discuss all of the intricacies and issues involved in all the different bonding techniques in a single chapter. Therefore, this chapter focuses primarily on reviewing each specific bonding technique and its characteristics.

1.09.2 Essential Elements of Wafer Bonding

Wafer bonding involves the attachment, often permanently, of two or more wafers made of similar or dissimilar materials. To achieve this, three fundamental elements are required, as discussed below and illustrated in Figure 1.

1. Wafer preparation
2. External activation
3. Bonding mechanism

- *Wafer preparation.* In almost all wafer bonding technologies, the wafers to be bonded have to be prepared properly. This often involves two steps. The first is the initial cleaning of the wafers to remove organic or inorganic contaminants from the wafers' surfaces, and to remove any particulates that might get lodged in between the two wafers and prevent bonding. The second step involves surface preparation, and this could vary widely from one bonding technique to another. In some techniques, the wafer surfaces have to be devoid of any surface oxide, which requires either physical or chemical removal of the oxide. In others, the surface has to be treated so that it either includes or is free of some chemical species that could help in the bonding process; for example hydrogen plays a key role in direct wafer bonding (DWB) of silicon.
- *External activation.* Once the wafers are cleaned, their surfaces are appropriately prepared, and they are properly aligned and brought together; the environment surrounding the wafers has to be activated such that the bonding process can be initiated and completed. There are a number of different external activation procedures that are used, depending on the type of bonding and the required characteristics. Most wafer bonding processes require heat. Some need an applied electric field, pressure or force, and some use vibration. The role of these external factors is to facilitate and initiate the bonding process. Some bonding techniques differ from others just based on

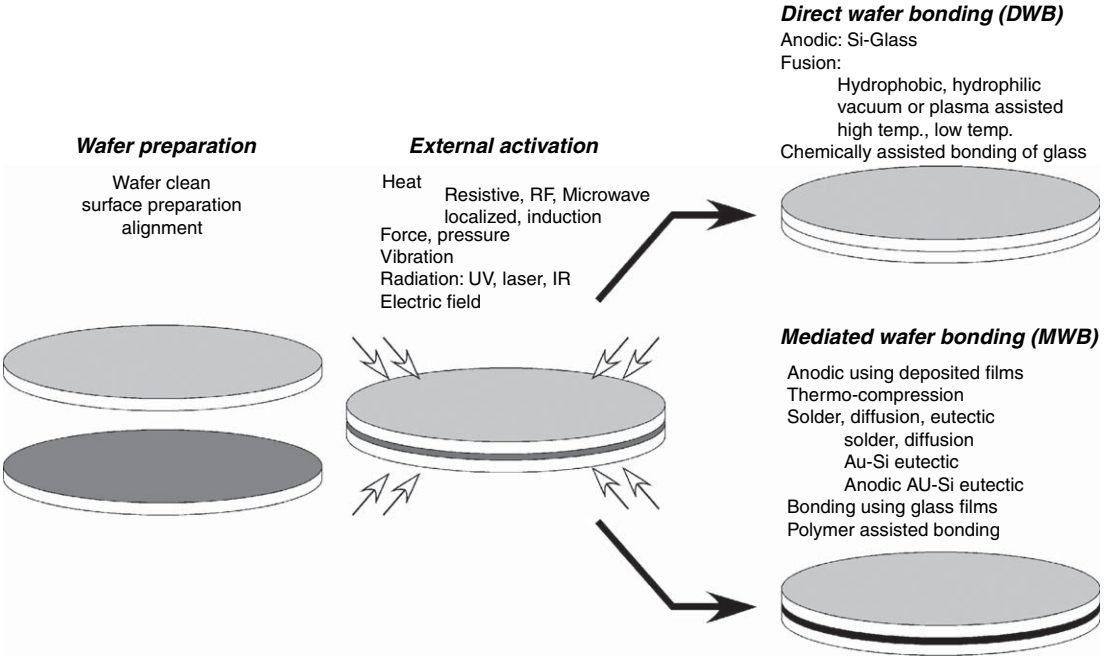


Figure 1 Three essential elements of wafer bonding: wafer preparation, external activation, and wafer bonding.

how this external activation is achieved. For example, heat can be applied to the bond region using a number of different techniques, including resistive heaters and microwave and radio frequency (RF) radiation. Wafer bonding based on all these techniques has been demonstrated and will be discussed.

- *Bonding mechanism.* The primary feature that differentiates different wafer bonding technologies is the intrinsic mechanism involved in the bonding process. The bonding process is initiated and progresses once all the external conditions are in place. For example, in DWB, the two surfaces of a cleaned and polished silicon wafer are brought into intimate contact thus allowing the atoms on the surfaces of these wafers to directly and covalently bond together and form a permanent bond. This chapter uses this distinction to categorize different bonding technologies, as discussed below.

1.09.3 Characteristics Desired in Wafer Bonding

Because of the various applications and materials used, it is difficult to develop a uniform set of requirements for wafer bonding technologies. Different applications have different requirements and demand different specifications for the bond. However, it is instructive to discuss what are typically considered desirable features for a successful wafer bonding technology. These desirable features include the following:

- *Low temperature.* Temperature is perhaps the most important factor in wafer bonding. As will be discussed later, most bonding technologies require the incorporation of some external energy to promote and initiate bonding between two wafers. This is usually done by heating the two wafers. It is very desirable that the bonding temperature be as close to room temperature as possible to reduce any detrimental effects due to such physical phenomena as thermally induced expansion and mechanical drifts.
- *Strong and high-quality bond.* Bond strength and the quality of the bond should be sufficient for the specific application. In most cases the bond should have very high strength not only to survive harsh environments such as external shock and stress due to subsequent processes, but also to minimize any fatigue or long-term drift. In addition to strength, the bonded interface has to be airtight and hermetic as most wafer bonding technologies are used for hermetic and vacuum packaging applications.
- *Low cost.* Needless to say, the cost of wafer bonding should be kept as small as possible. Although many bonding technologies have been, and can be, applied to individual dice or small wafers for many years at very low cost, some are not amenable to low-cost wafer-level processing. That is to say, the wafer bond is often more expensive than the die bond. For any wafer bonding technology to be successful, cost has to be minimized at all cost.
- *Minimum external disturbance.* Wafer bonding also often involves the application of pressure, force, or vibration to further promote interaction between two surfaces and to initiate the bonding process. These external mechanical disturbances should also be minimized in order to minimize the possibility of any damage to the wafers, or the devices that might exist when the wafers are bonded together.
- *Tolerance to particulates, surface preparation, surface topology, or roughness.* One of the most critical challenges in wafer bonding is proper surface preparation. Most bonds cannot be reproducibly and uniformly formed if the surface is not clean, has a thin oxide layer, has any particulates stuck to it, or has topography that could interfere with the bonding. For most bonds to be successful, the two surfaces either need to come into intimate contact or should have sufficient contact to facilitate their mixing and interaction. The role of the surface and surface properties cannot be underestimated.
- *Reproducibility, uniformity, high yield.* Many wafer bonding technologies exist, and many of them have been demonstrated in practice, illustrating the theoretical possibility. But not all wafer bonding technologies are practical or useful. In some cases, the bonds achieved are neither reproducible nor uniform across a given wafer. In other cases, the yield is unpredictable. These are very important characteristics of a successful bonding process and often dictate what the eventual cost will be.
- *Ability to bond patterned wafers.* Increasingly, wafers that either have patterned structures on them or have nonplanarities due to etched regions have to be bonded. Some bonding technologies are more suitable for bonding of patterned wafers than others.
- *Low void/defect density.* Some bonding techniques tend to create voids or defects at or close to the bond interface. Some of these voids are formed as a result of the bonding process itself as gases are trapped in between the two wafers, and others are created due to

thermal mismatch between the two wafers. For some applications, especially in microelectronics devices, it is critical that the density of these voids and defects be kept to a minimum.

- *Compatibility with IC processes.* In some applications the bonding technology has to be compatible with ICs already fabricated on one of the other wafers. This requirement usually imposes a limitation on the maximum temperature to which wafers can be exposed and the chemicals used for forming, cleaning, treating, and activating the bonding surfaces.

Although these features are desirable in most wafer bonding technologies, other requirements may also be needed by some specific applications.

1.09.4 Wafer Bonding Technologies

Many different wafer bonding technologies have been developed during the past few decades, and many of these are in practical applications today. In this chapter these bonding technologies are categorized by the primary mechanism that is used to bond the wafers. The chapter also reviews several bonding technologies that are identified by the different activation processes they utilize in initiating and completing the bonding process. Bonding processes are divided into two broad categories, as illustrated in **Figure 1**. The first requires only two plain wafers without the need for any intermediate layer; this approach is often referred to as DWB. The second requires the use of an intermediate layer between the two wafers, and in this chapter, we refer to this approach as mediated wafer bonding (MWB). DWB approaches can be further categorized into the following techniques:

- Field-assisted anodic bonding
- Fusion bonding
  - Hydrophobic, hydrophilic
  - High temperature, low temperature
  - Vacuum- or plasma-assisted
- Chemically assisted bonding of glass surfaces

MWB approaches can be further categorized into the following techniques:

- Field-assisted anodic using deposited films
- Thermocompression
- Solder, diffusion, and eutectic
- Bonding using glass films
- Polymer-assisted wafer bonding

Bonding technologies can also be categorized based on the external activation technique they utilize, and these are categorized as follows:

- Localized heating
- RF and microwave and electromagnetic heating
- Ultrasonic
- Rapid thermal processing (RTP)
- Laser

Each of these bonding techniques is reviewed with a short discussion of their advantages, capabilities, limitations, and disadvantages.

1.09.4.1 Direct Wafer Bonding

1.09.4.1.1 Field-assisted anodic bonding

One of the most widely used bonding techniques in MEMS is that of bonding between a glass wafer/structure and a polished silicon substrate through a process called electrostatic or anodic bonding (Albaugh 1991, Ando *et al.* 1997, Anthony 1983, Blasquez and Favaro 2001, 2002, Briand *et al.* 2004, Chavan and Wise 2001, Dokmeci *et al.* 1997, Dunn *et al.* 2000, Esashi *et al.* 1992, Fehlner *et al.* 1986, Frye *et al.* 1986, Go and Cho 1998, Hara *et al.* 1999, Harpster and Najafi 2002, Henmi *et al.* 1993, 1994, Kanda *et al.* 1990, Lee *et al.* 2000, Meindl 1980, Pomerantz and Daniel 1968, 1969, Qingfeng *et al.* 2002, Rogers 1992, Rogers and Kowal 1995, Schjolberg-Henriksen *et al.* 2003, Spangler and Wise 1987, 1990, Tatic-Lucic *et al.* 1997, Wallis 1970, Wallis and Pomerantz 1969, Ziaie *et al.* 1996). This technique was first discovered and patented by Pomerantz (Pomerantz and Daniel 1968). This patent describes bonding of insulators to metals and semiconductors, and his work primarily focused on the bonding of metals to ceramics. In a later patent he also described the bonding of liquid metals to ceramics (Pomerantz and Daniel 1969). Later, the MEMS community and industry used the same approach to bond silicon and glass wafers with great success.

This bond can be achieved when two polished wafers of silicon and glass are brought into intimate contact and a high voltage is applied across this sandwich to form a permanent covalent bond, as illustrated in **Figure 2** (Najafi 2001–2006). The glass–silicon sandwich is heated to temperatures of about 300–400°C, which tends to soften the glass and lower its electrical resistivity (the softening point of Pyrex #7740 glass is 821°C and its annealing point is 560°C). Bonding should be achieved below these two temperatures to prevent any permanent distortions in

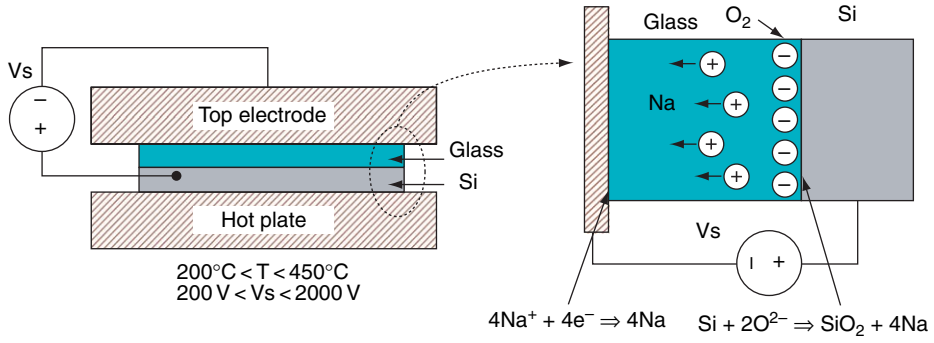


Figure 2 Anodic silicon–glass bonding. (Source: Najafi K 2001–2006 Wafer bonding. Lecture notes for Introduction to MEMS, A course taught at the University of Michigan.)

the glass that could compromise the shape of the final structure being formed. However, the temperature should be high enough to cause slight softening of the glass and to allow it to conform to minor topographical variations that could inhibit intimate contact between the glass and silicon wafers. A DC voltage of 800–1500 V is also applied across the glass–silicon combination for a period anywhere between 5 and 30 min depending on the wafer thickness, the series impedance, the interface planarity, and the bonding equipment. Typically, the wafers are pressed together with a uniform force to make sure that they are in good contact across the wafer. The positive terminal of the voltage source is connected to the silicon wafer, and the negative terminal is connected to the glass wafer. At room temperature, the glass wafer is a relatively good insulator, but at elevated temperatures it becomes a weak conductor. The electric field created within the glass forces the mobile alkali ions (sodium) to be pulled away from the bond interface and to the back surface of the glass where the electrode contact is located. They leave behind fixed positive ionic oxygen charges at the interface between glass and silicon. The applied voltage also generates a very high electrostatic attractive force, which pulls the silicon and the glass into intimate contact. The combination of the good contact and the availability of oxygen and silicon atoms at the interface promotes and forms a strong SiO_2 chemical bond between the two wafers right at the interface. During the bonding process, a current of a few milliamperes per square centimeter flows between the two electrodes, signifying the movement of ions and the completion of the process. Once the bond is complete, the structure is cooled down to room temperature. Figure 3 shows two photographs. The first

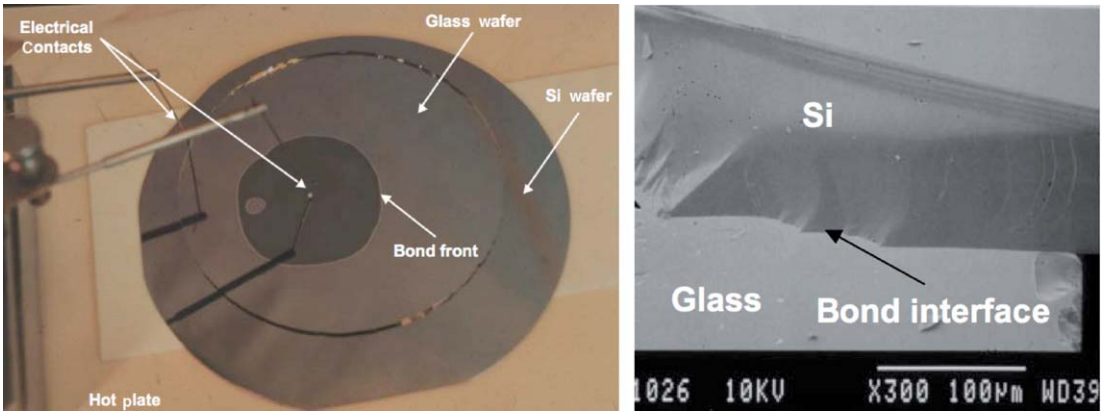


Figure 3 *Left*: View of a silicon–glass bond in progress. The glass wafer is on top of the silicon wafer, and both are on top of a hot plate. Electrodes provide electrical voltage to the two wafers. The dark circle in the middle shows where the two wafers are bonded and as bonding proceeds the circle gets darker. Note that if there are particles between the two wafers, they get trapped and engulfed by the bonded interface. (Photo courtesy: Dr. Leland Chip Spangler, Aspen Technologies.) *Right*: Scanning electron microscope (SEM) view of a broken interface showing the glass and silicon films. The bond is stronger than either silicon or glass.

is the top view of a silicon–glass bond in progress. The glass wafer is on top of the silicon wafer, and both are on top of a hot plate. Electrodes provide electrical voltage to the two wafers. The dark circle in the middle shows where the two wafers are bonded; as bonding proceeds the dark circle enlarges. Note that if there are particles trapped between the two wafers, they get engulfed by the bonded interface. The other photo in **Figure 3** is a scanning electron microscope (SEM) view of a broken interface showing the glass and silicon films. The Si–O bond is stronger than the Si–Si bond and the glass. In order to minimize intrinsic stresses generated within the structure and at the silicon–glass interface, the thermal expansion coefficients of the glass and silicon have to be as nearly matched as possible. Therefore, Pyrex glass code #7740 is often used for silicon–glass bonding applications since this glass has not only an expansion coefficient very close to that of silicon, but also a low softening point that helps lower the bonding temperature. Alignment between the two wafers can be easily achieved using standard wafer bonding equipment.

There are several parameters, such as surface preparation, contact load, temperature, time of applied voltage, and magnitude of the applied voltage, that control the bond quality, which should be monitored during the anodic bonding process. Go and coworkers provide an analysis on glass–Si samples and calculate the effect of contact load, voltage, temperature, and time on bond strength; the results are summarized in **Table 1** (Go and Cho 1998). The data show that the bonding temperature affects bond strength, i.e., higher temperatures result in stronger bonds. The applied voltage and the bonding time contribute to bond strength to a lesser extent. Typical cleaning procedures in bonding include a hot acetone bath followed by an isopropyl alcohol (IPA) rinse and deionized (DI) water rinse.

It is also possible to bond silicon to a number of other types of glasses. More conductive glasses like the Pyrex #7740 glass that have a reasonable content of sodium and boron have lower melting and softening temperatures and are more suitable for anodic bonding. As the glass becomes more pure, its electrical conductivity decreases, which will reduce the electric field strength and lower the electrostatic attraction that exists between the two wafers. Therefore, lower conductivity glass wafers typically need a higher bonding temperature and higher applied voltage to achieve a successful bond. For example, Spangler and Wise reported in 1987 the use of Pyrex #1729 glass (Spangler and Wise 1987, 1990). This high temperature Corning Code #1729 glass is an alkaline-earth aluminosilicate glass that has an anneal point of 853°C and has an excellent match to the thermal expansion coefficient of silicon. For the experiments that Spangler conducted, a bond temperature of 670°C was used along with an applied voltage of 900 V for about 30 min between the silicon and the glass. This glass has a much lower content of sodium and is therefore more compatible with wafers containing circuits.

Table 1 Influence of process parameters on bonding strength

Process parameter	*Contribution to strength (%)*
Bonding temperature	67.7
Applied voltage (V)	18
Bonding time (min)	12
Bonding load (gf cm^{-2})	2.3

Source: Go and Cho (1998).

The anodic silicon–glass bond offers several advantages. It is a relatively low-temperature process, it needs a very simple setup and is extremely reliable and uniform, and the glass can be selected to minimize thermal mismatch between it and the silicon wafer. It is possible to form this bond on more than two wafers and form sandwiched stacks of silicon and glass that are useful for wafer-level packaging or multilayer complex microstructures. The glass substrate is a thermally and electrically insulating substrate and transparent to light, which makes it very suitable for sensing and optical applications. The electrostatic force allows the bond to progress in spite of some local nonplanarities and forces the bond to engulf small particulates that might be trapped between the two wafers. This makes the bonding process quite tolerant to the existence of small particles. Because the silicon and the glass surfaces have to be in contact, nonplanarities and surface topologies must be limited to less than ~500 Å in height, otherwise they will prevent the formation of a good bond. Although most devices and processes can tolerate the large electric field and applied voltage, in some the high electric field could cause a problem. However, techniques have been devised to circumvent this problem by placing appropriate shield metals on the surface of glass to shield the silicon from the high field. There is a substantial body of literature on various aspects of silicon–glass bonding. This technology has been applied to both the fabrication and the packaging of MEMS in many different application fields. Glass to

metal anodic bonds have also been used in MEMS devices and applications (Briand *et al.* 2004).

This technique has also been used for bonding silicon–silicon and glass–glass wafers using sputtered and evaporated thin films of glass or silicon at temperatures as low as 135°C. This approach is discussed in more detail in Section 1.09.4.2.

1.09.4.1.2 Fusion bonding

Fusion bonding is referred to as both DWB and the process of bonding two polished silicon wafers together at a high temperature (Barth 1990, Harendt *et al.* 1990, 1991, 1992a, b, Lasky 1986, Mehra *et al.* 1999, Mitani *et al.* 1991). In this chapter we refer to the bonding of two wafers, silicon and otherwise, as fusion bonding. The name simply refers to the fact that when two surfaces that are polished are brought into intimate contact, they will fuse and will instantly bond. For this to happen the surfaces have to be polished, free of any roughness or damage, free of any polymers, or any other contaminants so that they can be chemically active and form covalent bonds when brought into intimate contact and annealed. Fusion bonding can be used to bond a number of different types of wafers, silicon-to-silicon, oxidized silicon-to-silicon, two oxidized silicon wafers, GaAs-to-silicon, quartz-to-silicon, silicon-to-sapphire, silicon–germanium-to-silicon with oxide, and silicon-to-glass, quartz-to-quartz, and a variety of other materials (Barth 1990, Harendt *et al.* 1990, 1991, 1992a, b, Lasky 1986, Mehra *et al.* 1999, Mitani *et al.* 1991).

In order to achieve a strong bond, the wafer surfaces have to be chemically prepared and activated. This can be achieved in a number of ways, including using wet chemistry, plasma activation, or in ultra-high vacuum (UHV) environments. Fusion bonding can be achieved between two silicon surfaces with either hydrophilic or hydrophobic surfaces, depending on the surface preparation used. These surfaces can also be treated in dry chemistries using plasma activation. In all these cases, after surface preparation, the wafers have to be brought together and then annealed at either high or low temperature. Therefore, fusion bonding techniques can be divided into the following categories:

- Hydrophilic, hydrophobic
- Vacuum- or plasma-assisted
- High temperature, low temperature

Each of these approaches are reviewed briefly. The high-temperature bonding process is sometimes referred to as fusion bonding, and it has become the most widely used silicon–silicon bonding technique in the MEMS community and is reviewed first (Alexe and Gösele 2004, Goesele *et al.* 1999, Tong and Gösele).

When two polished surfaces are brought together, they will tend to stick together with relatively weak intermolecular attraction forces, such as van der Waals' forces or forces associated with hydrogen bridge bonds. If this weakly bonded sandwich is annealed at a high temperature, the bond strength increases significantly due to the formation of strong covalent bonds between atoms on the surfaces of the two wafers. The nature of the bond depends heavily on the surface property and chemistry and on whether bonding is performed in vacuum or in atmosphere. Fusion bonds that are not performed at UHV can be categorized into two broad categories, namely hydrophilic and hydrophobic bonds. When discussing these two methods, the primary focus is on wafer bonding of silicon wafers.

1.09.4.1.2.(i) Hydrophilic fusion bonding In a silicon wafer, when the surface is hydrophilic (meaning it attracts water), as shown in **Figure 4**, strong Si–O–Si bonds can be formed between the two wafers after the wafers are brought into initial contact and annealed at an elevated temperature (Najafi 2001–2006). After a typical wafer cleaning and rinsing process, a thin layer of native silicon oxide is formed on the surfaces of the two wafers, and these oxide surfaces are terminated with OH groups, and hence the surface is hydrophilic. The dried wafers are brought into contact at room temperature and a weak bond is initiated at some point across the interface. The wafers are then heated to anneal the bond. Heating removes the water molecules, leaving behind a Si–O–Si bond, which is very strong. The strong Si–O–Si bonds are shown schematically in **Figure 4**. Obviously in this type of fusion bonding, a thin layer (a few nanometers) of silicon oxide is formed between the two wafers. This thin layer of silicon oxide may not be desirable in applications where a direct contact between the two silicon wafers is needed.

1.09.4.1.2.(ii) Hydrophobic fusion bonding For some applications, the insulating oxide layer at the bond interface can disrupt device operation. For example, if electronic devices such as p–n junctions are produced, good electrical contact is desired between the two surfaces, so wafer bonding should

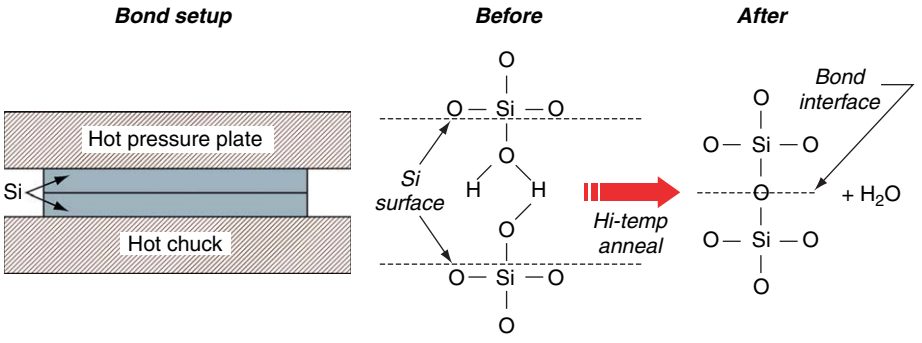


Figure 4 Hydrophilic fusion bonding between two wafers. The surfaces of two silicon wafers are covered with a thin layer of native silicon oxide and are hydrated with OH groups. After contact and high-temperature annealing, Si–O–Si bonds are formed forming a strong bond interface. (Source: Najafi K 2001–2006 Wafer bonding. Lecture notes for Introduction to MEMS, A course taught at the University of Michigan.)

be performed using hydrophobic surfaces that are free of any oxide. The native oxide can be first removed using diluted hydrofluoric acid (HF) or ammonium fluoride, which produce a surface that is temporarily covered with covalently bonded hydrogen. This surface is hydrophobic, and can be easily contaminated by hydrocarbons, which means that the two wafers should be bonded quickly after surface treatment. After the initial bonding, the wafers are again annealed at elevated temperatures whereby the hydrogen leaves the bonds and very strong covalent Si–Si bonds are formed at the interface. This bonding process is illustrated in **Figure 5** (Najafi 2001–2006).

Figure 6 shows a set of photographs showing a fusion bond in its various stages (Harendt *et al.* 1991). The four photographs on the left-hand side show infrared images of two silicon wafers that are bonded using fusion bonding. The bond formation starts on the top left corner where the two wafers are pressed together, and then proceeds until the two wafers are completely bonded. The two photographs on the right show voids that are formed due to the existence of particles between the two wafers. This is a major challenge for fusion bonding in that it is sensitive to particulates. This requires the wafer surfaces to be properly cleaned and to be free of any particulates. In fusion bonding there is no applied electric field, and therefore there is no force to pull the wafers together, as there is in anodic bonding.

The bond strength increases as the annealing temperature increases. The activation temperatures for hydrophilic bonding are lower than for hydrophobic bonding because hydrogen is easier to remove from the interface than are water molecules. The bond energy between silicon wafers at room temperature and with increasing temperatures has been measured for hydrophilic and hydrophobic bonds and has been shown to increase with temperature (*Philips J. Res., Spec. Issue Direct Bonding* **49**, 1). For hydrophilic bonds, the bonding energy increases starting around

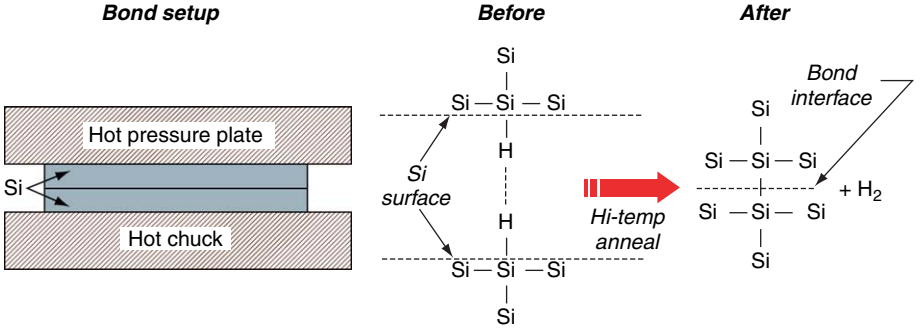


Figure 5 Hydrophobic fusion bonding between two wafers. The surfaces of two silicon wafers are etched to remove any native oxide, and are treated, so the silicon dangling bonds are occupied with hydrogen. After a high-temperature annealing, the hydrogen is removed and the two surfaces permanently bond to form covalent Si–Si bonds. (Source: Najafi K 2001–2006 Wafer bonding. Lecture notes for Introduction to MEMS, A course taught at the University of Michigan.)

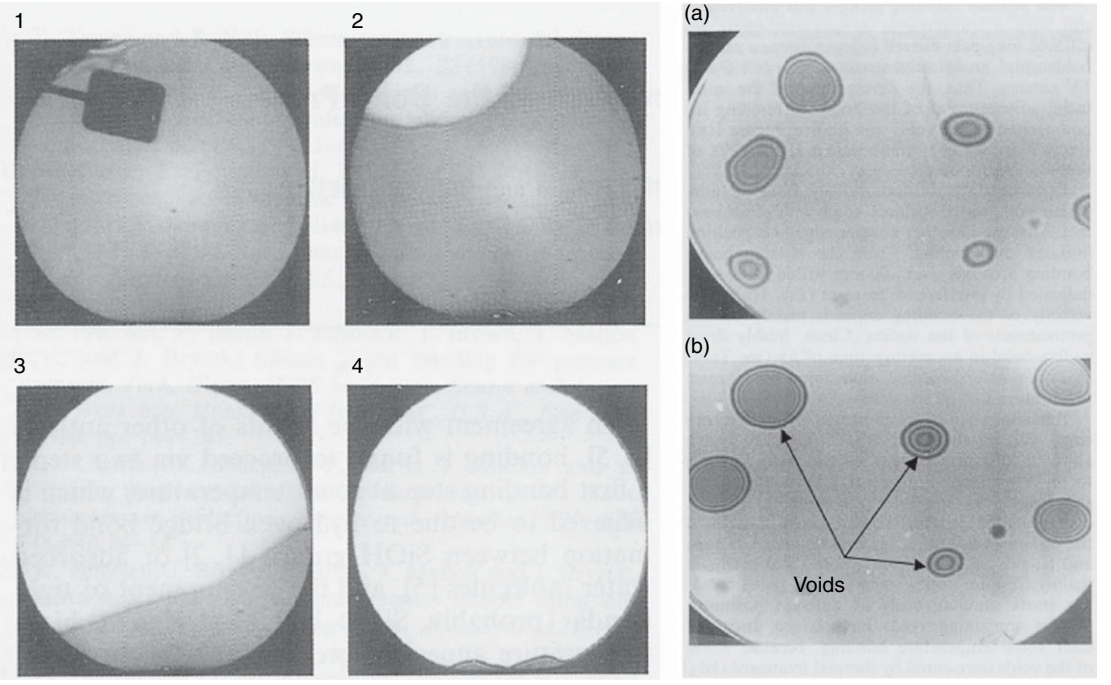


Figure 6 Through-wafer infrared images showing the progression of a typical fusion bond starting from the top corner of the wafer pair. On the right-hand side, the two images show the formation and existence of voids in fusion-bonded wafer. These voids are formed due to particulates or surface irregularities. (Source: Harendt C, Höfflinger B, Graf H G, Penteker E 1991 Silicon direct bonding for sensor applications: Characterization of the bond quality. *Sens. Actuators A Phys*. **25**–**27**, 87–92.)

120°C, and then achieves its highest value when the temperature increases over 900°C. For hydrophobic bonds, the bond energy starts increasing around 300°C, and will be very high at temperatures around and above 700°C. Wafers bonded using the hydrophobic approach are prone to the formation of voids during annealing.

1.09.4.1.2.(iii) Fusion bonding in vacuum As discussed above, the surfaces of wafers have to be prepared to ensure that strong chemical bonding is achieved between atoms on these surfaces. In hydrophobic bonding, especially, the surface is prepared such that after annealing, hydrogen-terminated bonds are removed and replaced with Si–Si bonds. It is also possible to prepare these surfaces in very high-vacuum environments. Wafers are placed inside a high-vacuum chamber and are treated to remove any surface oxide and to expose the silicon atoms on the surface. This treatment could be done by using plasma etching and treatment, by sputtering a thin layer of material from the surface, by ion beam bombardment, or by laser treatment. The high-vacuum environment ensures that the surface atoms are not reoxidized. The wafers are then brought into contact where they bond at room temperature. The bond strength is already high after this room temperature bonding since the surfaces are clean and free of any other species. This technique is referred to as UHV or surface-activated bonding (SAB) (Akatsu *et al.* 1999, Chung *et al.* 1997, Hosoda *et al.* 1998, Suga and Otsuka 2001, Suga *et al.* 1989, 1993, Takagi *et al.* 1996, 1999, Yang *et al.* 1997). This technique has several advantages in that it can be applied to many materials including metals, alloys, and semiconductors. This technique has been of particular interest to the semiconductor industry because it exposes the wafers only to low temperatures during surface treatment, and the bond itself can be formed at room temperature. However, because of the lack of high-temperature annealing, wafer surfaces have to be polished, free of any oxide, and have to be maintained in a very clean environment before they are bonded.

1.09.4.1.2.(iv) Plasma-activated fusion bonding In plasma-activated wafer bonding, the surfaces of the wafers to be bonded are treated and cleaned before bonding by exposing them to plasma (Amirfeiz *et al.* 2000, Benoit *et al.* 2006, Bower and Chin 1997, Bower *et al.* 1993, Eichler *et al.* 2006, Farrens 2005, Farrens *et al.* 1995, Galchev *et al.*

2007, Itoh *et al.* 2003, Morra *et al.* 1990, Suni *et al.* 2002, Wiegand *et al.* 2000). The plasma activation step alters the surface chemistry in a similar manner as the hydrophilic chemical treatment discussed earlier without the wet processing step. The surface chemistry is selected such that the maximum bond strength is obtained at low temperatures. Obviously this process has to be carried out in a vacuum environment and inside a plasma chamber. It is also possible to chemically treat the surfaces using a barrier discharge method where plasma is formed in the atmosphere (Eichler *et al.* 2006, Galchev *et al.* 2007). Plasma-activated bonding has now been applied to various materials, including silicon, compound semiconductors, oxides, and polymers.

In plasma-activated wafer bonding, the wafer surfaces have to be activated by a plasma and cleaned off, and then the two wafers need to be brought into contact to initiate the bonding process. Usually, an annealing step at an elevated temperature of 100–400°C is needed depending on the required bond strength and the materials being bonded. Plasma activation can be performed in a standard plasma chamber, an reactive ion etching (RIE) etcher, or an inductively coupled plasma (ICP) chamber. It can involve plasma of different species, including oxygen, argon, nitrogen, or hydrogen. Oxygen plasmas creates a thin oxide layer on the interface and are used for applications that require electrical isolation at the interface. Hydrogen, nitrogen, or argon plasmas are used in applications where a direct bond interface without the oxide layer is needed. The type of plasma used, its power, and the post-treatment annealing all can have a major influence on bond strength, void or defect density at the interface, and the annealing temperature. Since wafers need to be aligned and then bonded, preferably without breaking vacuum, specialized tools are needed that combine the plasma activation chamber and the bond alignment and annealing into a single machine.

Typically plasma activation is performed in an RIE chamber for a very short duration. Another form of plasma activation, called dielectric barrier discharge (DBD), can also be used to treat surfaces at high pressure, as illustrated in **Figure 7** (Galchev *et al.* 2007). By applying an AC current between two electrodes surrounded by a gas flow, localized atmospheric pressure plasma discharges are created. This is achieved by flowing the desired gas between two electrodes surrounded by a dielectric. This whole electrode assembly is then scanned across the wafer surface, exposing it to the plasma discharges. The voltage can be applied only between the electrodes. In this indirect mode, electronic components on the wafer are protected. Alternatively, the wafer can be grounded, leading to a more energetic direct mode of operation. The interesting characteristics of DBD are that the discharge is too short to generate heat, and so sensitive substrates can be protected, and also unlike low-pressure glow discharge, there is little surface bombardment of the wafer by energetic species. The literature related to low-pressure glow discharge plasma treatment suggests that an optimum exposure exists with respect to power and time of exposure (Weigand *et al.* 2000). In other words, with exposure to plasma, the surface tension rises gradually; however, past a maximum, further exposure to low-pressure plasma reduces the surface tension. During DBD, power, number of treatments, distance between the wafer and the electrodes, and the scan speed can be varied. An optimum set for bonding of silicon wafers has been found to be at 400 W with 50

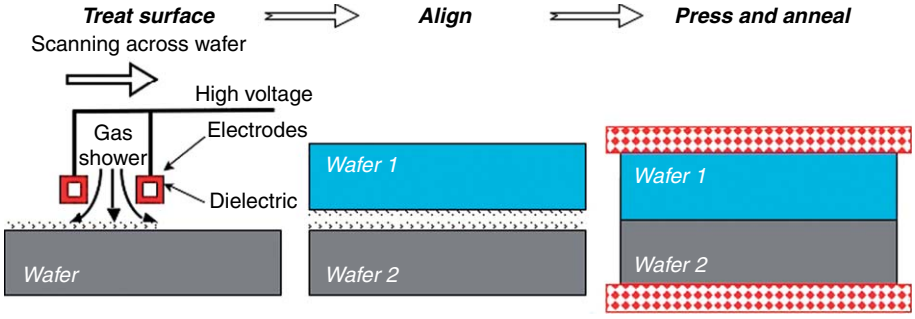


Figure 7 Dielectric barrier discharge (DBD) creates a uniform plasma discharge due to a high-frequency AC voltage applied between two electrodes. The discharge is scanned across the wafer surface. This process does not affect sensitive substrates. (Source: Galchev T, Welch WC III, Najafi K 2007 Low-temperature MEMS process using plasma activated silicon-on-silicon (SOS) bonding. *Proc. IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS Conference)*, Kobe, Japan, January 2007.)

slpm N_2, applied 300 μm above the wafer surface while the electrodes are scanned at 20 mm s^{-1}. After surface treatment, the wafers are moved into the bond alignment chamber where they are aligned and eventually annealed to form the bond. Annealing is typically done at temperatures in the range of 200–400°C. Pressure is also applied during annealing to compensate for any bowing of the wafers.

1.09.4.1.3 Chemically assisted bonding of glass surfaces

In this last category of DWB, two wafers are bonded together by treating their surfaces with wet chemicals and allowing these wet chemicals to stay in the bond region and slowly evaporate off after a specific period of time. This technique is mostly used for bonding oxide surfaces together and has been applied to bonding of quartz, glass, and oxidized silicon surfaces (Mao and Han 2005, Nakanishi *et al.* 1997, 1998). In this process, the surfaces of the two wafers are first cleaned using standard organic and inorganic cleaning processes (such as sulfuric acid and hydrogen peroxide mixtures) and exposed to diluted HF to remove a thin layer of oxide from the surface. The wafers are then placed together and a small amount of diluted HF is dropped in between the two wafers. The HF solution is wicked up into the interface by capillary action (note that the surfaces of the wafers are oxidized and are hydrophilic). The wafers are then pressed together using a high pressure and are allowed to stay together for anywhere between 5 and 24 h. The bonded wafer sandwich is then rinsed and dried. The bond strength has been shown to be as high as a few megapascals. The same technique has been used with chemicals other than HF (Nakanishi *et al.* 1997). Successful bonding has been achieved using only those chemicals that have a reasonable etch rate of SiO_2. This technique is very attractive and has been used extensively for making microstructures for chemical and biological applications where high-temperature processing cannot be used (Mao and Han 2005). Obviously the lack of heat activation reduces the bond strength, but for many of the intended applications the bond strength is more than sufficient.

1.09.4.2 Mediated Wafer Bonding

The second category of wafer bonding involves the use of an intermediate, or glue, layer between two wafers to permit their bonding and permanent attachment. This class of wafer bonding is very broad and has been used in various applications. The glue layer can be metals, semiconductors, inorganic insulators (like glass, oxides, or nitrides), or organic insulators (like polymers, epoxies, and various epoxies and waxes). In this chapter this bonding technique is categorized and each category is reviewed as follows:

- Anodic using deposited films
- Thermocompression
- Solder and eutectic
- Bonding using glass films
- Polymer

1.09.4.2.1 Field-assisted anodic using deposited films

Two wafers of glass or silicon can be bonded using a deposited thin film layer of glass or silicon (Berthold *et al.* 2000, Brooks *et al.* 1972, Choi *et al.* 1996, 1997, Cozma and Puers 1995, Dokmeci 1999, Drost *et al.* 1998, Esashi *et al.* 1990, Krause *et al.* 1995, Liang 1979, Nese and Hanneborg 1993, Plaza *et al.* 1998, Quenzer *et al.* 1996, de Reus and Lindahl 1997, Sassen *et al.* 2000, Weichel and de Reus 1998). As mentioned earlier, anodic bonding has been a very reliable and widely used bonding process for MEMS applications. In its standard form, it requires a glass wafer and a silicon wafer. But in some cases when bonding of either two silicon or glass wafers is desired, an intermediate layer of deposited glass or silicon is used to achieve anodic bonding. The first bonding of two silicon wafers using a deposited glass was demonstrated in 1972 using a thin layer of sputtered glass (Mao and Han 2005). The glass layer had to be a few micrometers thick to produce a reliable bond. Since the deposition rate of sputtered glass was small, a fairly long deposition step was needed. The deposited glass can be patterned using standard photolithography and etching in buffered hydrofluoric acid (BHF)/HF. Because two silicon wafers are used, problems associated with mismatch in thermal expansion coefficient are avoided. In addition, since the glass layer is very thin, the voltage necessary for achieving a good bond is usually not above 50–100 V. Later, other researchers used evaporated glass instead of sputtered glass and achieved successful results (Choi *et al.* 1996, 1997, de Reus and Lindahl 1997, Weichel and de Reus 1998). Deposition rates of evaporated glass are much higher but the deposition conditions have to be controlled more carefully to produce a deposited layer of glass that is uniform and homogeneous in its composition. **Figure 8** shows the

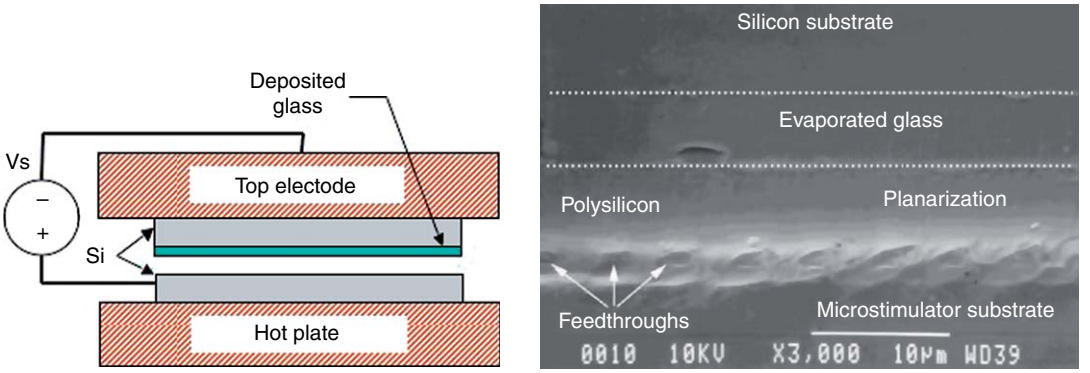


Figure 8 Si–Si anodic wafer bonding using deposited glass. *Left*: The basic setup for anodic bonding. (Source: Najafi K 2001–2006 Wafer bonding. Lecture notes for Introduction to MEMS, A course taught at the University of Michigan.) *Right:* The cross section of a silicon wafer bonded to a layer of evaporated glass on top of a planarized polysilicon film. The bond interfaces between the top silicon wafer and the bottom polysilicon layer and the evaporated glass layer are shown. (Source: Dokmeci M 1999 Hermetic glass–silicon micropackages and feedthroughs for neural prostheses. Doctoral dissertation, University of Michigan.)

bonding process and a photograph of a bonded silicon wafer pair formed using evaporated glass (Dokmeci 1999, Najafi 2001–2006). The glass was evaporated on top of a polished polysilicon layer that covered a series of electrical feedthroughs. The second silicon wafer is then electrostatically bonded to the bottom glass film (Dokmeci 1999).

Anodic bonding process can also be used for the bonding of two glass wafers with an intermediate deposited layer of silicon. The deposited silicon can be either sputtered or deposited using plasma-enhanced chemical vapor deposition (PECVD). These low-temperature processes ensure that the glass is not stressed too much during the deposition process. The deposited silicon has to have a smooth surface to form a good bond. It is also possible to bond two glass wafers using an intermediate silicon wafer. The technology of glass–silicon–glass bonding has been extensively used for making various 3D MEMS structures and for packaging of MEMS.

1.09.4.2.2 Thermocompression bonding

Metals have long been used as a glue layer for wafer bonding and die attachment. The most familiar of metal joining techniques is soldering, which is discussed in Section 1.09.3.2.3.(i). It is also possible to bond two wafers covered with a layer of metal using thermocompression (Condra *et al.* 1975, Furman and Mita 1992, Kim *et al.* 1995, Taklo *et al.* 2004, Tsau 2003, Tsau *et al.* 2004). This process is very similar to silicon–silicon fusion bonding in that it requires two clean and chemically active metal surfaces that are free of any particulates or oxide. Once these surfaces are brought together, pressed, and heated, they can permanently bond, as illustrated in **Figure 9** (Najafi 2001–2006). The heating of the surfaces softens them and provides sufficient energy to form a covalent bond at the interface. Almost any metal can be bonded via thermocompression. For metals such as Al or Cu the bonding surfaces must be treated to remove any oxide that could impede bond formation.

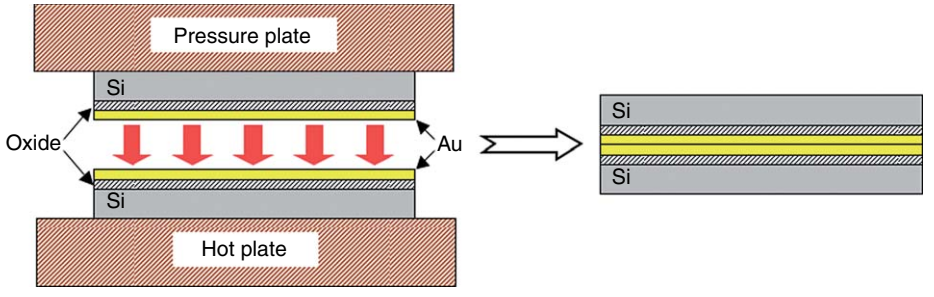


Figure 9 Thermocompression bonding of two wafers covered with a metal layer, like gold. (Source: Najafi K 2001–2006 Wafer bonding. Lecture notes for Introduction to MEMS, A course taught at the University of Michigan.)

Although it is possible to bond practically any two metals, thermocompression bonding of gold surfaces has received the most attention. This is primarily due to the inertness of gold, which makes its surface resistant to oxidation, and due to its softness, which enables it to deform under the applied pressure and form a strong bond. Thermocompression bonding has been a key technology for wire bonding in the IC industry and has also been used for die attach. However, it has only been recently used for full wafer bonding for MEMS applications. The challenge in full wafer bonding is achieving a uniform and reproducible bond without very high external forces or temperatures. The success of a gold thermocompression bond relies on many factors, the two most important of which are pressure and temperature. Pressures on the order of at least a few megapascals are needed and temperature has to be around 350°C–400°C. In most applications, the temperature should be reduced as much as possible, which then necessitates a higher pressure (Tsau 2003, Tsau *et al.* 2004).

Tsau and coworkers have described the process issues and techniques required for successful full wafer bonding in detail (Tsau 2003, Tsau *et al.* 2004). Their findings were the following:

(1) The gold film should be deposited on a diffusion barrier film such as oxide to prevent its reaction with the underlying silicon film.
(2) The gold surface should be as smooth as possible (rougher surfaces requires higher pressure and higher bonding temperature).
(3) The surface has to be treated to remove organics; UV exposure and cleaning have been shown to be effective.
(4) Bonding temperature and pressure of 260°C and 120 MPa, respectively, are recommended to ensure the level of repeatability and yield desired in manufacturing environments. Since the pressure may damage the underlying thin films or other structural materials, a tradeoff between pressure and temperature can be made.
(5) The bond time was not shown to significantly affect the bond quality. Given the preference for faster throughput, the bond time can be as short as 2 min.

Thermocompression bonding is quite attractive because it is relatively simple and requires only the deposition of a layer of gold on the surfaces to be bonded. The surface has to be clean and flat, and there are no complications due to metallurgical reactions as is sometimes the case for solder-based wafer bonding. The drawbacks of thermocompression bonding are the high pressures required and the need for very clean and polished surfaces for achieving a uniform and reproducible bond across a large wafer.

1.09.4.2.3 Solder, diffusion, and eutectic wafer bonding

In addition to thermocompression, metals have been used extensively for wafer bonding in the form of solder. Solders are formed either by using a mixture of various suitable metals or by forming a eutectic of a metal and a semiconductor, such as gold and silicon. These two approaches are discussed below.

1.09.4.2.3.(i) Solder wafer bonding Solder is a low melting point metal alloy (<450°C), typically containing tin or indium, which is used to join two metal surfaces together (Bosco and Zok 2004, Chuang and Lee 2002a, b, Gale 2003, Goyal *et al.* 2003, 2004, Humpston *et al.* 1994, Kim *et al.* 2006, Lugscheider *et al.* 2003, Park *et al.* 2001, Rogge *et al.* 1998, Singh *et al.* 1999, Sommadossi *et al.* 2003, Sparks *et al.* 2001, Welch and Najafi 2005, Welch *et al.* 2005a, b, c). It acts as a metal glue that adheres to each of the metal surfaces on two wafers and holds them together. A typical soldering process starts with three separate materials in contact: material A, the solder alloy, and material B as illustrated in **Figure 10** (Najafi 2001–2006, Welch and Najafi 2005, Welch *et al.* 2005a, b).

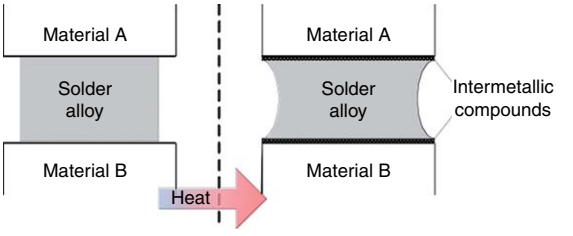


Figure 10 Illustration of the solder bonding concept. (Sources: Najafi K 2001–2006 Wafer bonding. Lecture notes for Introduction to MEMS, A course taught at the University of Michigan; Welch W C III, Najafi K 2005 Transfer of metal MEMS packages using a wafer-level solder sacrificial layer. *18th IEEE Int. Conf. Micro Electro Mechanical Systems*, Miami Beach, FL, USA, January 30–February 3, 2005, pp. 584–7; Welch W C III, Chae J, Najafi K 2005b Transfer of metal MEMS packages using a wafer-level solder transfer technique. *IEEE Trans. Adv. Packaging (IEEE Trans. Components Packaging Manufact. Technol. Part B Adv. Packaging)* **28(4)**, 643–9; Welch W C III, Chae J, Lee S, Najafi K 2005c Transient liquid phase (TLP) bonding for MEMS packaging applications. *Tech. Dig., IEEE Int. Conf. Solid-State Sensors, Actuators, and Microsystems, Transducers 2005*, Seoul, Korea, June 2005.)

When the whole assembly is heated, the solder melts, and in its molten state it wets the surfaces of materials A and B and interacts with them to form intermetallic compounds. After the solder wets both surfaces, the assembly is cooled down and the solder bond is complete. The low melting point alloy still exists in the solder joint, and so the service temperature of the assembly cannot exceed the bonding temperature (i.e., the melting temperature of the solder) or the solder will melt again and compromise the integrity of the joint.

The choice of the solder alloy is an important part of the solder process design. It determines the bonding and service temperature, the mechanical strength of the joint, the filling characteristics, and the material compatibility. There are several different solder alloys to choose from that suit different types of applications (**Table 2**). There is a large pressure in the industry to eliminate solders containing lead in order to comply with pending worldwide legislation.

The key to understanding the behavior of solder lies in two concepts: wetting and intermetallic formation. To understand the behavior of solder, it is necessary to understand how it behaves as a liquid. Most of the action in a soldering process occurs during its liquid state. The fluid behavior of the solder is dependent on the surface energies of the materials present in the solder joint. Each material has a certain surface energy, which is determined by the free energy from the unsatisfied bonds of the atoms on the surface of the material. This surface energy creates a surface tension force that keeps the liquid together. The definition of "to wet" for soldering purposes is a contact angle $<90°$, which means that the droplet has spread over the surface of the solid. The lower the contact angle, the greater the spreading effect. This contact angle and spreading effect during bonding have a great impact on many of the attributes of the final solder joint.

Table 2 Several widely used low melting point solder alloys

Composition (wt.%)						
Ag	*Bi*	*In*	*Pb*	*Sn*	*Au*	*Eutectic (°C)*
–	49.0	21.0	18.0	12.0	–	57
–	57.0	–	–	43.0	–	139
3.0	–	97.0	–	–	–	144
–	–	99.4	–	–	0.5	156
–	–	–	38.0	62.0	–	183
3.5	–	–	–	96.5	–	221
–	–	–	–	20	80	278

During the soldering process, the solder alloy interacts with the materials it is in contact with and forms new phases. The new phases consist of one of the metals in the solder alloy and the parent materials. For instance, when tin-based solders are used to bond gold surfaces together, the tin reacts with the gold to form Au_x-Sn_y phases. The new phases may form a new interface between the molten solder and the parent metal or they may diffuse into the molten solder. This intermetallic formation process consumes the parent metal. The consumption rate depends on the solder alloy, the parent metals, the temperature, and the amount of intermetallic phases already formed. The erosion rate and the presence of new phases in the solder joint must be considered during the design of the soldering process. The new phases can add desirable mechanical properties, such as ductility in the case of Ag–Sn, or they can make the joint more brittle in the case of Au–Sn. The erosion rate can cause problems when soldering to thin films, especially to thin films of gold.

Traditional solder bonding consists of a molten filler metal wetting two surfaces, then solidifying to form a strong metallic bond. For a successful solder bond, it is imperative that the solder wets the two surfaces and that the erosion of parent metals by the solder be accounted for in the design. As mentioned earlier, good wetting occurs when the surface tension of a solid is higher than the surface tension of the liquid in contact with it. Many pure metals have a high surface tension, but their oxides do not. Most metals react with the oxygen in the atmosphere to form a very thin native oxide film on their surface that prevents good wetting. Fluxes are highly reactive materials that are used in a soldering process to ensure good wetting. During the soldering process, they remove this native oxide and other contaminants from the bonding surface. Fluxes are a good way to achieve proper wetting and a reliable solder joint, but they are not compatible with MEMS processing. The liquid flux can cause stiction issues with released MEMS devices. Other alternative flux-free soldering processes must be used for MEMS packaging. There are several fluxless soldering processes. One approach, plasma-assisted dry soldering, uses a plasma pretreatment to alter the chemistry of the metal's native oxide to make it solderable without a flux. This approach is reliable, but it requires specialized equipment. Another fluxless soldering technique first removes the metal oxide with a chemical treatment, such as a light acid etch, and then coats the metal with a thin layer of gold, using

physical vapor deposition to electroplating, to prevent further oxidation. The gold layer does not oxidize, and many solder alloys consume it rapidly and reveal the clean metal surface underneath. Some process development is necessary to choose a compatible metal oxide etch. Also, care must be taken to limit the regrowth of the native oxide between the etch step and the gold deposition. It is also important to limit the amount of gold used to prevent any negative mechanical effects from the gold intermetallics. For reliable soldering to thin films, several different materials need to form the thin film stack. Each material performs a specific function: usually gold is the topmost layer to ensure proper wetting; below it there is usually a refractory metal, like nickel, that reacts with the solder slowly. The nickel prevents the solder from consuming the entire thin film during the process and revealing the underlying substrate to the molten solder. The erosion of parent metallizations and the formation of intermetallics are used in a relatively new technique called diffusion soldering to form bonds that are completely made up of intermetallics when they are finished.

Another type of solder bond, called diffusion soldering, combines the advantages of diffusion bonding and solder bonding. It has the planarization capabilities of a traditional solder bond, but it can survive much higher service temperatures than its formation temperature like a diffusion bond (Gale 2003, Humpston *et al.* 1994). It starts in much the same way a solder bond does. Two parent metals sandwich a low melting point interlayer. The bond formation process involves four stages (Welch and Najafi 2005, Welch *et al.* 2005b, c) (**Figure 11**). First, the two parent metals and the low melting point interlayer are brought into contact. Heat is then applied and the liquid interlayer melts and begins to react with the parent metals and forms intermetallic compounds. Third, the joint undergoes isothermal solidification as the last of the liquid interlayer is consumed in the formation of intermetallic compounds. After the consumption of the interlayer, the remelting temperature of the joint increases from the melting point of the interlayer to the melting point of the intermetallic compound. Last, further heating increases the quality of the bond as the intermetallic compounds diffuse away from the interface leaving behind the continuous pure parent metal.

Several factors need to be considered when designing a diffusion-soldered bond. The material system, the thickness of the parent and interlayer metals, the heating rate, and the pressure applied determines the final quality of the bond. The materials used to create the diffusion-soldered bond determines the bond formation temperature and the maximum temperature it can survive after completion. The bond formation temperature is just above the melting point of the interlayer and is thus dictated by the choice of this layer. Indium and tin are the most widely used interlayer metals because of their low melting points (156°C and 232°C, respectively) and high intermetallic reaction rates. For temperature-sensitive steps, indium is a better choice because it has a lower melting temperature, but it is more expensive. The choice of the parent metals

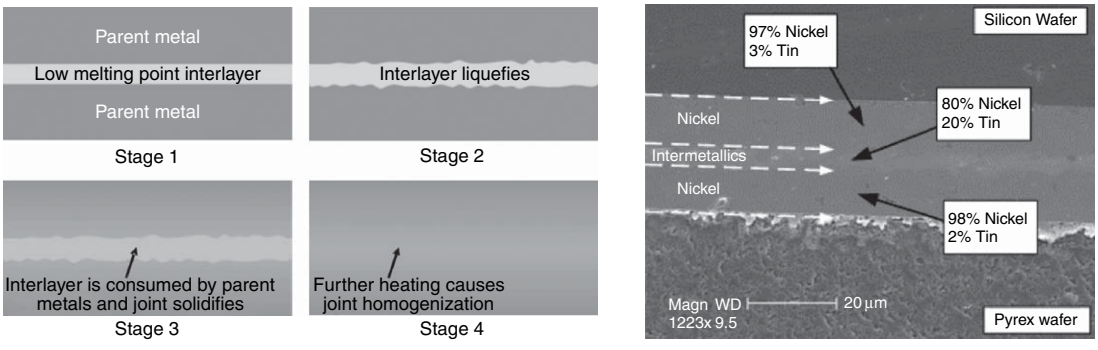


Figure 11 *Left*: The four stages of diffusion soldering. *Right*: Cross-sectional scanning electron microscope (SEM) view of a Ni–Sn–Ni diffusion solder bond. (Sources: Welch W C III, Najafi K 2005 Transfer of metal MEMS packages using a wafer-level solder sacrificial layer. *18th IEEE Int. Conf. Micro Electro Mechanical Systems*, Miami Beach, FL, USA, January 30–February 3, 2005, pp. 584–7; Welch W C III, Chae J, Najafi K 2005b Transfer of metal MEMS packages using a wafer-level solder transfer technique. *IEEE Trans. Adv. Packaging (IEEE Trans. Components Packaging Manufact. Technol. Part B Adv. Packaging)* **28(4)**, 643–9; Welch W C III, Chae J, Lee S, Najafi K 2005c Transient liquid phase (TLP) bonding for MEMS packaging applications. *Tech. Dig., IEEE Int. Conf. Solid-State Sensors, Actuators, and Microsystems, Transducers 2005*, Seoul, Korea, June 2005.).

determines the final melting temperature of the joint. Material systems like silver–indium and silver–tin offer the greatest increase in remelting temperature due to the high melting temperature of the silver intermetallics. The increase in melting temperature is 350°C for silver–tin and 705°C for silver–indium. The thickness of the metal layers and the heating rate are the most important factors in the formation of void-free bonds. A critical interlayer thickness exists that determines whether the final bond will be void-free or not (Welch and Najafi 2005, Welch *et al.* 2005a, b). If the interlayer is thicker than the critical value, the final joint will be void-free, if not, there is a chance of void formation during bonding. The critical thickness depends mostly on the heating ramp rate during bonding and the interaction rate between the interlayer and the parent materials. As the joint is heated, the interlayer begins to react with the parent metals and forms intermetallic compounds even before it melts. If the heating rate is too slow, this interaction completely consumes the interlayer before the melting point is reached. In this case, the interlayer never reaches its liquid phase and cannot fill any gaps in the joint between the parent metals, leaving voids in the final bond. A faster heating rate or a greater interlayer thickness increases the amount of the liquid phase during bonding and increases the planarization capabilities of the diffusion soldered bond.

Wafer bonding using solder is a very attractive technology. It can be done using a number of different material sets, offers a large range of bonding temperatures, the solder flows, and can therefore accommodate wafers that have large nonplanar features on their surfaces, including microfabricated structures or feedthroughs for packaging applications, and it can be used with various wafer materials. The major challenges in wafer bonding using solders are materials interactions and achieving good wetting without the use of flux. Wafers to be bonded must have suitable under-solder materials that will not only ensure that the solder is properly formed but also that its adhesion to the two wafer surfaces is adequate for the given application. Solder bonding has been used for die attach and for many electronics applications, and it is now being more extensively used in MEMS and wafer bonding applications as well.

1.09.4.2.3.(ii) Silicon–gold eutectic bonding

Gold–silicon eutectic formation occurs at 363°C for 19 at.% silicon and 81 at.% gold composition. The eutectic can be used to bond two wafers or for hermetic and vacuum packaging. To form the eutectic, silicon and gold have to be provided to the interface where two wafers are bonded. Typically, the gold is deposited on one of the wafers to a desired thickness, and the silicon is provided either from the bulk of one of the wafers or from thin films deposited on one or both wafers. When the two wafers are brought into contact and heated above 363°C, the films will react to initiate the eutectic liquid formation. Upon cooling, the bond consists of Au–Si hypereutectic phase and represents a typical strong diffusional bond.

Au–Si eutectic bonding has been investigated in depth in the past (Adachi 2002, Adams *et al.* 2003, Bruson and Gerl 1982, Chang *et al.* 1988, 2000a, Christensen and Bouwstra 1996, Dalton 1956, Harpster 2005, Harpster and Najafi 2003, Johnson and Johnson 1987, Johnson *et al.* 1987, Kato 1987, 1989, Mei *et al.* 2002, Mitchell *et al.* 2005a, b, Tiensuu *et al.* 1994, Valero 1984, Wang and Lee 1992, Wolffenbuttel 1997, Wolffenbuttel and Wise 1994, Xiao *et al.* 1998). Researchers have investigated such parameters as silicide formation, epitaxy, and fractal growth, and surface oxide formation for Au–Si films formed on the surface of a Si wafer. Others have studied Au–Si eutectic bonding for its application in chip bonding applications, applied it for wafer bonding applications, and even used it in the alignment of optical fibers.

The gold–silicon eutectic process has to be carried out carefully in order to ensure a good and solid bond between the eutectic and the underlying substrate. Mei and coworkers characterized the bonding and the adhesion quality of the gold–silicon eutectic on several different materials, including Si/Ti/Au to Au/Ti/Si; Si/Ti/Au to Si; Si/Ti/Au to polySi/Si; Si/Ti/Au to Oxide/Si; and Si/Ti/Au to nitride/Si. Of these, the bond quality and the uniformity between Au–Au, Au–Si, and Au–PolySi was the best. In particular, it was noted that the bond quality was at its best when the gold and the silicon were supplied from two wafers, instead of both from the same wafer. Test wafers where the Au–Si eutectic was bonded to a substrate covered with either oxide or nitride produced very nonuniform and poor bonds. This is illustrated schematically in **Figure 12(a)** (Mei *et al.* 2002, Mitchell *et al.* 2005a, b). The figure shows the cross-sectional view of a gold–silicon eutectic bonding two silicon wafers together. The gold in the eutectic mix is revealed after the silicon is etched away in a selective silicon etchant. The figure also shows the uniform mixture of the gold and silicon films.

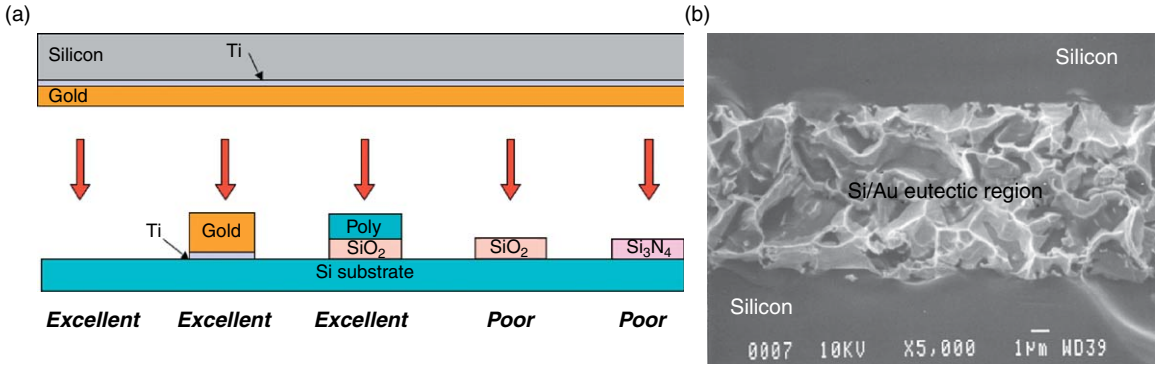


Figure 12 (a) The gold–silicon eutectic bond quality for different surfaces and layers. (Source: Najafi K 2001–2006 Wafer bonding. Lecture notes for Introduction to MEMS, A course taught at the University of Michigan.) (b) The cross-sectional view of a gold–silicon eutectic film bonding two silicon wafers together is also shown on the right. The gold in the eutectic mix is revealed after the silicon is etched away in a selective silicon etchant. The figure shows the uniform mixture of the gold and silicon films. (Sources: Mei Y, Lahiji G R, Najafi K, A 2002 Robust gold–silicon eutectic wafer bonding technology for vacuum packaging. *Solid-State Sensor, Actuator and Microsystem Workshop*, Hilton Head, SC, USA, pp. 772–4; Mitchell J S, Lahiji G R, Najafi K 2005a Encapsulation of vacuum sensors on a wafer level package using a gold–silicon eutectic. *Tech. Dig., IEEE Int. Conf. Solid-State Sensors, Actuators, and Microsystems, Transducers '05*, Seoul, Korea, June 2005; Mitchell J S, Lahiji G R, Najafi K 2005b Reliability and characterization of micro-packages in a wafer level Au–Si eutectic vacuum bonding process. *Proc., ASME/Pacific Rim Tech. Conf. Exhibit. on Integration and Packaging of MEMS, NEMS, and Electronic Systems, InterPack05*, San Francisco, CA, USA, July 2005.)

Mitchell *et al.* (2005a, b) have further characterized the process and used gold–silicon eutectic bonding for wafer-level vacuum packaging. In their baseline bonding process, they used one wafer with a layer of electroplated gold and another wafer with a thin layer of polysilicon deposited on top of a silicon–nitride layer. This process that produced uniform and reproducible bonds is schematically shown in **Figure 13(a)**. The 4″ diameter wafers were first held at 345°C under a vacuum pressure of 1 μtorr in order to allow for outgassing from the inner surface of the cavity. The wafers were then pressed together with a pressure of 1 MPa above the eutectic temperature for 40 min. Bonds were formed at

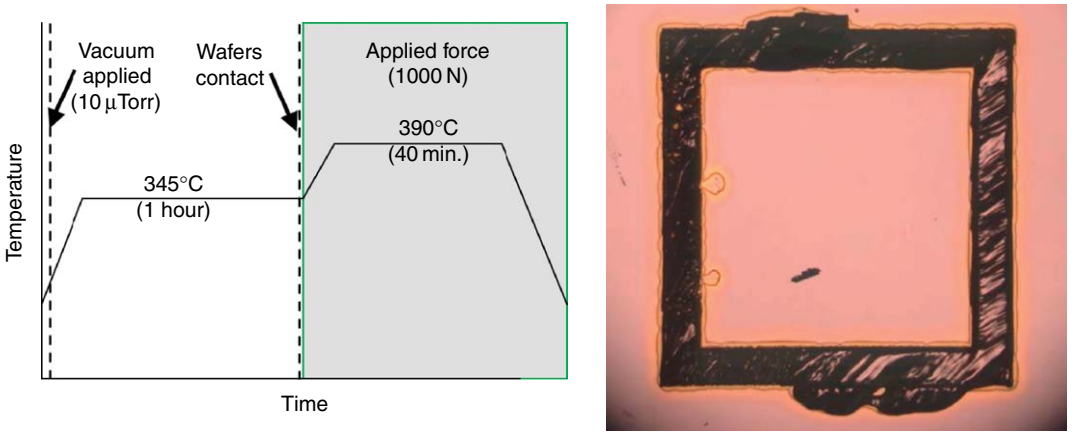


Figure 13 (a) The process recipe for successful and uniform bonding of two silicon wafers together. (Source: Mitchell J S, Lahiji G R, Najafi K 2005b Reliability and characterization of micro-packages in a wafer level Au–Si eutectic vacuum bonding process. *Proc., ASME/Pacific Rim Tech. Conf. Exhibit. Integration and Packaging of MEMS, NEMS, and Electronic Systems, InterPack05*, San Francisco, CA, USA, July 2005.) (b) Photograph of a bond ring used for vacuum packaging after the bond is broken, leaving pieces of silicon on the opposite wafer and illustrating the very uniform and strong nature of the gold–silicon eutectic bond. (Source: Mitchell J S, Lahiji G R, Najafi K 2005a Encapsulation of vacuum sensors on a wafer level package using a gold–silicon eutectic. *Tech. Dig., IEEE Int. Conf. Solid-State Sensors, Actuators, and Microsystems, Transducers '05*, Seoul, Korea, June 2005.)

temperatures ranging from below 390°C to 450°C and with bond pressures from 1 to 10 MPa (lower bond pressures were not investigated). One key parameter in this bonding process involved applying pressure before going above the eutectic temperature. It was also important to choose the correct thickness of the gold bond ring on the cap wafer. Using gold of a thickness <2.5 μm resulted in nonuniform bonds. That is, a strong bond was formed on some areas of the wafer and not on others. Another important factor for obtaining a good bond was the thickness of the polysilicon layer on the device wafer. In one set of experiments, 4- to 8-μm-thick gold bond rings were used on the cap wafer to bond to 0.8 and 2.2 μm of polysilicon on the device wafer. With very little force, caps were pried off fracturing in the polysilicon. In this work they were not able to determine why bonding to films of this thickness were problematic. Bonding to polysilicon films thinner than 0.5 μm alleviated this problem, resulting in consistently strong bonds.

1.09.4.2.3.(iii) Field-assisted anodic silicon–gold eutectic bonding As mentioned above, silicon–gold eutectic bonding requires that the surfaces be prepared and cleaned well. In addition, it is critical that during the bonding process, the gold and the silicon are supplied from two wafers, so that the interdiffusion of the two materials creates an intimate contact with both surfaces. It has also been mentioned that the bond quality of Si–Au eutectic to silicon oxide, nitride, or glass is typically not very good and nonuniform. Harpster and coworkers have developed a new bonding technique based on both Si–Au eutectic and anodic bonding (Harpster 2005, Harpster and Najafi 2003). It is found that by applying an anodic bias, as is typically done in direct Si–glass wafer bonding, in a standard Si–Au eutectic bond, a significant improvement in bond quality, uniformity, and reproducibility can be obtained. This new technique makes the following three very significant contributions to wafer bonding:

(1) The electrostatic forces ensure intimate contact between the Si–Au solder and glass interface uniformly over large areas, which is absolutely necessary for wafer-level bonding and hermetic sealing.
(2) A roadblock in obtaining hermetic and vacuum sealing using electrostatic methods is planarization of the bonding interfaces. The fluid Si–Au eutectic solder ensures instant and contiguous planarization over nonplanar substructures such as electrical feedthroughs.
(3) It is speculated that all the bonding materials are biocompatible, which open the arena for biomedical applications.

The bonding process is illustrated in **Figure 14(a)** (Harpster 2005, Harpster and Najafi 2003). A Ti/Au layer (200/10 000 Å) is evaporated onto the silicon wafer. A glass wafer and the gold surface, on the Si wafer, are kept at a fixed separation distance using spacers as typically found in wafer bonding equipment. Control of the diffusion of the Au and Si is critical to the formation of a hermetic bond and therefore the temperature of the wafers throughout

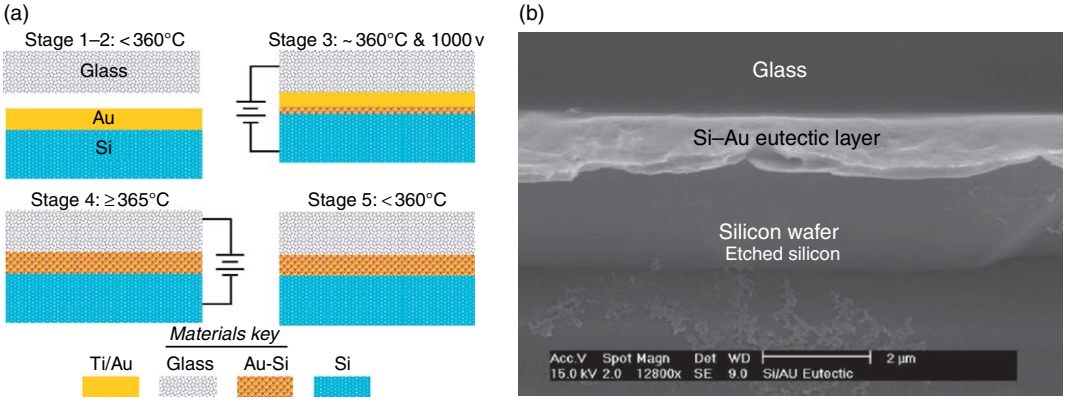


Figure 14 (a) Stages of the liquid Au–Si alloy formation and bonding process. (Source: Harpster T J 2005 Hermetic packaging and bonding technologies. Doctoral dissertation, University of Michigan; reproduced with permission.) (b) Scanning electron microscope (SEM) of bonded cross section. (Source: Harpster T J, Najafi K 2003 Field-assisted bonding of glass to Si–Au eutectic solder for packaging applications. *Proceedings, International IEEE Microelectromechanical Systems Conference (MEMS Conf.),* Kyoto, Japan, January 2003.)

this bonding process is a critical control parameter. In stage 1 the wafers must be heated to a temperature below the eutectic point of the Au–Si solder and allowed to come to equilibrium. A set temperature of 250°C is used prior to ramping to 345°C to reduce the overshoot on the stage-2 heating cycle. Reducing the ramp rate can also minimize the overshoot. In stage 2 the wafers are heated to a temperature below the eutectic temperature of the Au–Si alloy and the chamber is evacuated to ~10 μtorr. The transition to the eutectic temperature must be accurately controlled. Therefore to reduce the uncertainty of time at which the wafer reaches the eutectic temperature and to maximize the temperature uniformity across the wafer, the temperature step from the soak to bond temperature is minimized. The wafers are held at 345°C to allow the wafers and bonder to reach thermal equilibrium before the critical transition to the stage 3.

In stage 3, the wafers are brought into contact, and a force of 1000 N is applied. After a 3-min wait for the upper and lower wafers to reach thermal equilibrium a voltage of 1000 V is applied. The applied force is primarily used for overcoming wafer bow and for deforming the solder after its transition to the liquid state and thereby planarizing over micromachined structures. The wafers are then immediately heated to 365°C whereby the formation of a liquid-phase eutectic initiates at the Au–Ti–Si interface. The liquid Au–Si alloy layer increases in thickness or volume depending on the contour, as it consumes the Au and underlying Si and Ti layers. After the Au layer is consumed and correspondingly the eutectic composition ceases to grow, the Au diffusion into the underlying silicon will continue thereby leaving a Si-rich solder and as a result solid Si phases will crystallize. Solid phases in the liquid layer will likely increase the effective viscosity of the melt and may impede planarization of the glass surface, thus bond quality and uniformity will be compromised. Therefore, time and temperature above the eutectic temperature are critical. The diffusion rate should be kept low to ensure homogeneous mixing of Au and Si and to allow time for structures to be planarized before the solid Si phases form. Therefore, the bonding temperature must be kept near the eutectic temperature and a bias must be applied prior to a eutectic formation at the glass surface because this ensures that the electrostatic forces pull the liquid solder into contact with the glass before any solid phases can form.

When the Au–Si–Ti solder reaches the glass surface, the electric field at the glass–solder interface, due to the anodic reaction, pulls the solder into intimate contact with the glass surface. There are three competing reactions that play a role in forming the solder-to-glass bond:

(1) A Au–glass eutectic reaction (Katz *et al.* 1991)
(2) A glass–Si anodic reaction (as in a typical glass–Si anodic bond)
(3) Wetting of the solder to glass being promoted by Ti in the solder (Humpston 2004).

In stage 4 the wafers are held at 365°C for ~20 min to allow the liquid alloy to form and the glass–solder interface to bond. In typical glass–Si anodic bonding, a general rule of thumb is that the bond time is determined by the decay time of the anodic current. One of the proposed mechanisms for the glass–solder bond is the anodic oxidation of Si at the glass–solder interface; therefore, this rule of thumb is followed.

In stage 5 the wafers are cooled to ~220°C at a maximum cooling rate of 19.5°C min^{-1} (0.33°C s^{-1}) to ensure that the Au–Si alloy solder is solidified before the applied bias and contact pressure are removed. The resulting bond between the glass, thin-film solder, and the silicon wafer is robust and uniform and cannot be separated by mechanical prying. A section of the bonded wafer is cleaved for cross-sectional viewing. An SEM photograph of the bonded interface is shown in **Figure 14(b)**.

1.09.4.2.4 Bonding using glass films

In addition to metal solders, glass solders or films can also be used for wafer bonding. The advantage of these seals is that they are insulating and can be applied to insulate electrical feedthroughs for MEMS packaging (Cook and Liniger 1999, Deligeorgis *et al.* 2002, Dragoi *et al.* 1999, Drost *et al.* 1998, Ebert and Bagdahn 2004, Esashi *et al.* 1990, Kikuchi *et al.* 1997, Liu *et al.* 2003, Sparks *et al.* 2003, 2004). There are two common methods for wafer bonding using glass seals. The first method uses lead borate or solder glasses as an adhesive between wafers to be bonded, and the second method using spin-on-glasses that are deposited in liquid form and then cured and baked to form a final solid film.

In the first method, the glass compounds are manufactured as powder called frit and mixed into screen printable pastes. High temperatures are used to soften the glass frit, and pressure is used to ensure intimate contact forming a hermetic seal. The frit is then cooled

whereby it solidifies forming a permanent bond. The frit is made up of a base glass, a refractory filler, and a vehicle. The base glass forms a majority of the frit. The filler reduces the coefficient of thermal expansion (CTE), enhances the mechanical strength and the crack resistance of the frit. The vehicle is made up of a solvent that provides fluidity for screen printing and an organic binder. Many different types of glass frits are commercially available with different melting points, CTEs, organic binders, and screen printing properties. Almost all the low melting point formulations of glass frit contain some lead, which is a major drawback. Solder glasses are characterized as either vitreous or devitrifying, depending on whether they crystallize at a temperature within the sealing range. Although devitrifying glasses generally have higher strength, vitreous glasses soften at lower temperatures and thus are more attractive for MEMS device wafer bonding. Among the lowest temperature glasses are lead borates, generally of high lead oxide content with additions of B_2O_3, ZnO_2, SiO_2, and Al_2O_3. The eutectic of the $PbOB_2O_3$ system occurs at 490°C (87.5% PbO/12.5% B_2O_3). Phosphorus doping and the addition of other materials, such as low CTE refractory filler materials, are used to obtain acceptable CTE values, improve flow characteristics, increase strength, and lower the softening temperature. Glass compositions that allow bonding below 400°C without contaminating the integrated device are available. Frit glass is screen printed on to wafers about 10–25 μm thick. This thick layer is advantageous for mitigating defects such as particulates, planarizing over wafer topologies such as feedthroughs, and providing a standoff for moving MEMS structures. Glass frit is widely used for hermetic sealing of MEMS devices such as accelerometers and gyroscopes. Integrated Sensing Systems (ISSYS) has used this technology together with its NanoGetter technology to fabricate wafer-level vacuum packages for MEMS using glass frit bonding (Sparks *et al.* 2003, 2004). ISSYS has demonstrated that the frit is stable and can hold very high levels of vacuum for extended periods of time for a variety of devices and applications.

Drawbacks of glass frit bonding include relatively large footprint (>150 μm) due to limitations of screen print patterning. Additionally, outgassing the organic paste constituents require the use of getters for vacuum sealing applications.

Silicon on glass (SOG) is another glass seal wafer bonding technique whereby silanol (Si–OH) and methyl (CH_3) polymers are dissolved in an alcohol/acetone solvent to form a low-viscosity spin-on material often referred to as a sol–gel (Cook and Liniger 1999, Liu *et al.* 2003). Similar to glass frit bonding, polishing and etching to reduce roughness of the wafer surfaces before bonding is not critical because SOG can fill and smooth out the surface topology. SOG thicknesses are limited to 0.5–1 μm due to induced stress from the resulting glass film. The wafer pairs are bonded at low temperature (250°–400°C); however typically a secondary high-temperature step (~1100°C) is employed to improve the dielectric breakdown voltages. Advantages of SOG wafer bonding include controllable wafer gap for capacitive devices, hermetic sealing capability, ability to seal over limited topology (~<1 μm), and the ability to bond wafers of different materials. SOG patterning has been demonstrated using high-energy lithography techniques such as electron scanning, focused ion beam (FIB), and deep X-ray lithography (DXRL). SOG has been demonstrated in bond combinations of Si, Si with oxide, silicon with dielectrics, and even III–V materials. SOG with alkali components has also been used for field-assisted wafer bonding. The glass is cured prior to bonding thereby reducing glass outgassing during anodic bonding, thus high-vacuum packaging can be achieved. This method has been demonstrated for Si-to-Si and Si-to-GaAs wafers. Liu *et al.* (2003) uses a 500-nm SOG layer to form a low-temperature (~500°C) wafer bonds for optically pumped vertical cavity surface-emitting lasers (VCSELs). Thicker layers were demonstrated but they cracked under the mechanical-induced stresses from dicing. Historically, the high-temperature step used for curing has limited the use of SOG for many commercial applications. More recently, SOG films have been used for anodic bonding of silicon-to-silicon wafers. This is a useful technology since it allows thick layers of glass to be deposited to planarize over feedthroughs and then use anodic bonding to bond the second wafer on top of the first.

1.09.4.2.5 Polymer-assisted wafer bonding

Polymers are extensively used for bonding wafers together (Ageoroges and Ye 2002, Bayrashev and Ziaie 2002, Bilenberg *et al.* 2004, Chou 2001, Chou and Najafi 2001, Dragoi *et al.* 2003, Eaton *et al.* 1994, Folsom *et al.* 1999, Han *et al.* 2000, Hansford *et al.* 1998, Huang *et al.* 2004, Jeon *et al.* 2002, Kim 2006, Kim and

Najafi 2003, 2005, Lan *et al.* 1997, Lee *et al.* 1997, 2003, Meng *et al.* 2000, Mizuno *et al.* 2005, Niklaus *et al.* 2000, 2001, Noh and Hesketh 2000, Noh *et al.* 2003, 2004, Oberhammer *et al.* 2003, Oh *et al.* 2002, Owen and Smith 1995, Pantelidis *et al.* 1999, Smith 1993, Su and Lin 2001, Takahashi *et al.* 2000, Xu *et al.* 2003, Yang *et al.* 1999). Frequently they are used for temporary bonding of wafers where they simply play the role of glue. Therefore, in these cases the polymer need not be stable or provide long-term adhesion, it just has to keep the wafers together for a short while. More recently, however, polymers have been increasingly used for permanent bonding of wafers. They have to satisfy many more requirements and have to be stable and reliable. The primary advantage of polymer-assisted wafer bonding is the low bonding temperature and the ease with which polymers can be applied on a wafer. Other advantages include low cost, sealing over large topographies, bonding of dissimilar materials, and transparency to visible light. Although polymer bonding is used for many commercial devices, polymers are susceptible to water uptake and gas diffusion, which leads to delamination at the bond interfaces. Additionally, polymer bonding is not compatible with vacuum packaging because of outgassing during the curing step.

Various polymers have been used for wafer bonding, including parylene (Kim and Najafi 2003), SU8 (Bilenberg *et al.* 2004), benzocyclobutene (BCB) (Chou and Najafi 2001, Niklaus *et al.* 2000, 2001), CYTOP (fluorocarbon polymer) (Han *et al.* 2000, Oh *et al.* 2002), polymethylmethacrylate (PMMA) (Lee *et al.* 2003), polyimide (Bayrashev and Ziaie 2002), Mylar (Su and Lin 2001), many different epoxies and waxes, and liquid crystal polymers (LCP). Some of these polymers and their primary features are listed in **Table 3**. Polymers are usually deposited on bonding surfaces as a viscous liquid through spin-on coating, and then cured at a moderate temperature (around 150°C) to form a solid bonding layer. These conventional polymers, however, go through a phase transition from a liquid (before curing) to a solid (after curing). Thus, they have problems, as bonding agents, with voids caused by the curing process, have difficulty in aligning fine features before curing, and cannot be easily spread uniformly on a wafer surface with tall structures or deep etched regions.

Of these polymers, BCB and parylene are discussed briefly since these are the two polymers that produce the least amount of outgassing and void formation. BCB is a spin-on liquid polymer and is

Table 3 Low-temperature wafer bonding using adhesive polymer layers

Polymer type	*Parylene*	*SU8*	*BCB*	*CYTOP*	*PMMA*	*Polyimide*	*MYLAR*
Reference	Kim and Najafi (2003, 2005), Noh *et al.* (2004)	Bilenberg *et al.* (2004)	Chou (2001), Chou and Najafi (2001), Niklas *et al.* (2001)	Han *et al.* (2000), Oh *et al.* (2002)	Lee *et al.* (2003)	Bayrashev and Ziaie (2002)	Su and Lin (2001)
Bonding conditions							
Curing (temperature)	Not required	90°C	65°C	90°C	N/A	90°C	N/A
Temperature (°C)	<230	120	65	160	92	280	140
Layer thickness (μm)	<1	10	1–12	1.4	N/A	2–20	15
Vacuum	Yes/No	Yes	Yes	N/A	No	Yes	No
Force (stress)	800 N	2000 N	800 N	4 MPa	3 ~ 4 MPa	N/A	0.4 MPa
Bonding time	30 min	10 min	20 min	30 min	2 min	7 min	0.25 s
Bond strength (MPa)	>3.6	<16	N/A	<4.3	>0.96	>1.5	N/A
Long term (2 years)	>3.6 MPa						
Chemical endurance	>1.15 MPa						

Sources: Ageoroges and Ye (2002), Lee *et al.* (2003).

photodefinable, whereas Parylene is evaporated down as a dry layer over highly nonplanar surfaces. These are used as examples to show how polymer bonding can be done.

BCB is an epoxy-based polymer. It was primarily developed as a low dielectric constant material for planarization of ICs (Lan *et al.* 1997, Lee *et al.* 1997, Takahashi *et al.* 2000). When fully cured, BCB shows low volume shrinkage and high resistance to most wet chemicals (Folsom *et al.* 1999). Some of the properties of BCB are listed in **Table 4**. Void-free thermal compressive bonding by using this material was reported by Niklaus *et al.* (2000, 2001). BCB bonding strength is larger than other polymer-based materials (**Table 3**). Since the cured BCB is inert to most acids, bases, and solvents, it is an excellent bonding material for applications that require postbond wafer processing. BCB can be spun onto a wafer, exposed, and then developed to form specific patterns. After pattern definition, BCB can be partially cured to remove any residual solvent or directly bonded with a companion wafer by applying pressure and temperature. Compared to other bonding techniques, BCB bonding is a simpler process since it does not need any special setup. The cured BCB thickness can vary from ~1 μm to more than 10 μm depending on the spin speed and BCB-type used.

After patterning and precuring, the wafer coated with BCB is bonded to another wafer, and is then postbaked and cured to achieve the final strong bond between the two wafers. The bonding was performed under an applied pressure of 150 kPa at 250°C for 30 min, at which point BCB is more than 95% cured. The curing time increases exponentially in order to achieve 100% curing. One can achieve 95% curing within 5 min by increasing the baking temperature up to 300°C. The quality of BCB bonded wafers is excellent. When bonded wafers with 10-μm-wide and 6 μm thick BCB are bonded and then forced apart, it has been observed that more then 90% of the patterned BCB was transferred from the host wafer onto the companion wafer. The original pattern remained on the host wafer as well. This indicates that the BCB bond strength is sufficient to maintain structural integrity. The 10-μm-wide, 6-μm-thick BCB lines and 5-μm-wide channels formed by the surrounding BCB still retain their shapes after bonding. BCB can serve as the air gap spacer as well as the bonding medium between two wafers, as shown in **Figure 15** (Chou 2001). The final air gap is defined by the cured BCB thickness. In conventional MEMS fabrication, the air gap is usually formed by deposition of sacrificial layers, such as low-pressure chemical vapor deposition (LPCVD) polysilicon and silicon dioxide. These sacrificial layers are limited in thickness to a few microns and take a long time to deposit and etch. By using photosensitive BCB as a spacer, microstructures with air gap >10 μm can be fabricated using a standard photolithographic process. The air gap variation was <5% across the 4″ silicon wafer in this experiment. BCB bonding technique has several advantages, including: (a) low-temperature (<250°C) bonding and low-stress (<30 MPa) film, (b) uniform air gap formation (<5% variation), (c) easy processing for different thicknesses with high throughput, and (d) high electrical and chemical resistance. It can be used not only as the localized bonding medium at low temperatures, but also as the air gap spacer

Table 4 Properties of BCB

Parameter	*Value*
Dielectric constant (1 kHz–20 GHz)	2.65
Breakdown voltage (V cm^{-1})	3×10^6
Resistivity (Ω cm)	1×10^{19}
Thermal conductivity (W $m^{-1}°K^{-1}$)	0.29
Thermal expansion coefficient (ppm per degree Celsius)	52
T_g (°C)	>350
Young's modulus (GPa)	2.9 ± 0.2
Tensile strength (MPa)	87 ± 9
Poisson ratio	0.34
Elongation at break point (%)	8 ± 2.5

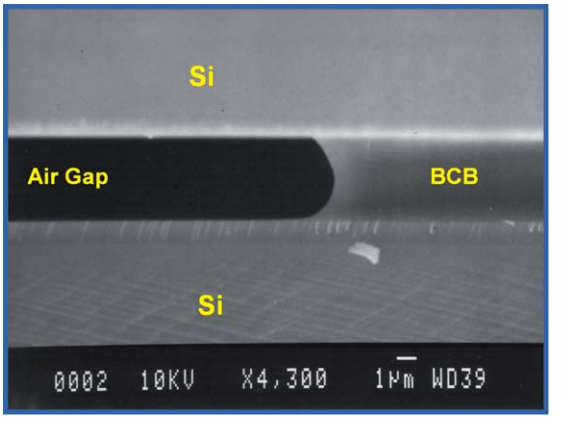


Figure 15 Cross section of locally bonded wafer with benzocyclobutene (BCB). (Source: Chou T K A 2001 All-silicon micromachined acoustic ejector array for micro propulsion and flow control. Doctoral dissertation, University of Michigan.)

between two structures with sufficient strength and uniformity.

Parylene (poly-*para*-xylylene), unlike conventional polymers, does not go through a liquid-phase transition; instead it is directly vapor-deposited from a solid powder. Thus, it can avoid problems associated with spinning of liquid materials, such as step coverage and outgassing. Parylene has not been widely used as a bonding agent, in spite of its widespread use in MEMS (Huang *et al.* 2004, Meng *et al.* 2000, Xu *et al.* 2003, Yang *et al.* 1999), until very recently (Kim and Najafi 2003, Noh and Hesketh 2000, Noh *et al.* 2004) for two main reasons. First, parylene does not form a strong covalent bond because it does not go through a curing process; it is deposited as a thin film through vapor deposition. Second, parylene is an inert film after deposition. Parylene is a high molecular weight, crystalline polymer, and consists of carbon backbones without any oxygen, nitrogen, or sulfur atom links. This crystalline carbon backbone film is naturally stable and highly resistant to chemical attacks. Thus, it does not allow permanent bonding or attachment to many surfaces. However, parylene is a thermoplastic polymer that becomes soft when it is heated and hard when it cools. Generally when a thermoplastic polymer is soft, portions of polymer chains can re-entangle with neighboring chains, thus forming a macroscopic bond (Ageoroges and Ye 2002, Smith 1993). In thermoplastic polymers, this phenomenon is reversible. It is believed that parylene undergoes this phenomenon.

In a recent paper, Noh *et al.* (2003, 2004) investigated wafer bonding using microwave heating of parylene on small (75 mm) wafers. His investigation mainly focused on the usability of microwave heating, the resultant uniformity, and the strength of parylene bonding. Kim and coworkers have done extensive characterization and testing of parylene-assisted wafer bonding (Kim and Najafi 2003, 2005, Noh *et al.* 2003). The bonding process starts by first preparing the wafers. Each of the two silicon wafers is coated with a thin (~0.381 μm) parylene layer, which serves as the glue layer. Each wafer is then cleaned using acetone and IPA, dried, and then bonded in a standard commercial bonder. Kim and coworkers demonstrated that wafers can be bonded using parylene; the optimized bonding conditions are a temperature of ~230°C, a vacuum of ~0.153 mbar, and a force of 800 N on a 100-mm wafer. It is noted that wafer bonding at 80°C (below the glass transition temperature) is almost too weak to be used. However, wafer bonding at temperatures above 130°C produces reasonable bond strength (>1.5 MPa) for MEMS applications. The measured bond strength versus bonding temperature is shown in **Figure 16** (Kim 2006, Kim and Najafi 2003, 2005).

Parylene bonding provides excellent qualities needed for many MEMS applications, as listed in **Table 3**. It produces reasonable bond strength above 3.60 MPa that is comparable to other polymer bonding methods. It maintains its bond strength

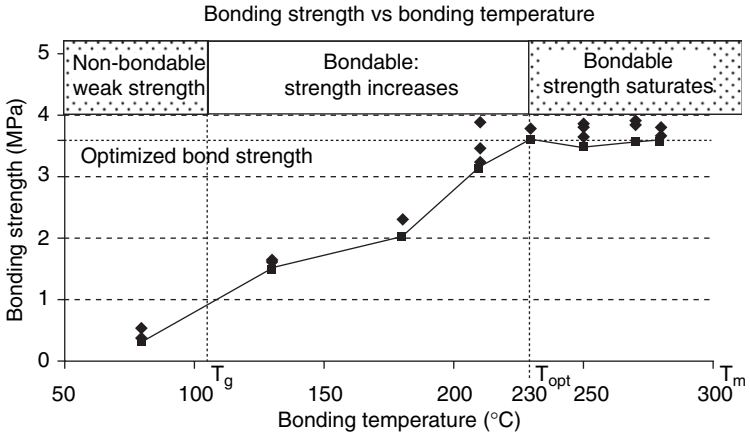


Figure 16 Measured bond strength versus bonding temperature for silicon wafers bonded using parylene. (Sources: Kim H, Najafi K 2003 Characterization of low-temperature wafer bonding using thin film parylene. *IEEE J. Microelectromech. Syst. (J. MEMS)* **14(6)**, 1347–55; Kim H, Najafi K 2005 Characterization of parylene-assisted wafer bonding: Long-term stability and influence of process chemicals. *Tech. Dig., 13th Int. Conf. Solid-State Sensors, Actuators, and Microsystems, Transducers '05*, Seoul, Korea, June, 2005, pp. 2015–18; Kim H 2006 An integrated electrostatic peristaltic micro gas pump with active microvalves. Doctoral dissertation, University of Michigan.)

under a typical room environment for at least 2 years. The bond is stable for long periods of time and parylene is resistant to many of the ordinary chemicals used for semiconductor fabrication. Parylene bonding is free of voids because it does not go through curing process. Parylene layers can be patterned to form an arbitrary shape for bonding a specific area.

1.09.4.3 Other Bonding Techniques

As mentioned before, the bonding techniques discussed above are differentiated by the different mechanisms used to form the bond. All these bonding mechanisms utilize heat to initiate and complete the bonding process. Bonding is usually performed in a bond chamber where the entire wafer is heated in a controlled ambient under an applied load. Therefore, if wafers contain either devices or materials that cannot withstand the required bonding temperature for any specific bonding technology, then that bonding technology cannot be used. This indeed limits the types of materials and bonding technologies that can be used for that specific application. In order to overcome some of these limitations, different techniques for heating wafers have been developed, as discussed below.

To achieve wafer bonding, one typically needs to heat only the regions between the two wafers that are in contact (i.e., the bond regions). There is usually no need to heat the entire wafer. It is sometimes required that the bonding temperature be raised above 400°C to allow various other materials and bonding techniques be used. One such example involves glass frit bonding, which needs bonding temperatures in the 500–700°C range. This temperature is too high for many materials used in MEMS and ICs. In some applications, the temperature of the MEMS devices cannot be raised much higher than 50–100°C. This is often the case when the MEMS device incorporates materials such as polymers or biological coatings, which cannot withstand high temperatures. Therefore, one needs techniques where the temperature of the MEMS parts can be kept below 100°C, or even lower, and the temperature where the bond between the two wafers is formed can be raised above 500°C. To achieve this goal, several techniques have been developed and proposed, including bonding based on localized resistive heating, microwave heating, laser heating, localized chemical vapor deposition (CVD) and heating, ultrasonic bonding, inductive heating, and RF heating (Chang-Chien and Wise 2001, Cheng *et al.* 1998, 1999, 2000b, c, 2001a, b, 2002, He *et al.* 1999, Lin *et al.* 1998, Mitchell and Najafi 2006).

1.09.4.3.1 Localized resistive heating

A new technique for heating the bond regions between two wafers using on-chip resistors was reported by Cheng *et al.* (1998, 1999, 2000a, b, 2001a, b, 2002), He *et al.* (1999), and Lin *et al.* (1998). In this approach, a resistive heater, using standard materials like aluminum or polysilicon, is formed only in those areas where the wafers are to be bonded. Because silicon is a very good conductor, the heat generated by the resistive heater can be confined and localized to the bond area. In the sensitive device regions, the temperature is maintained at room temperature or slightly above by directing most of the heat away to a heat sink attached to the bottom of the device substrate. Simulations and experiments have demonstrated that the temperature gradient falls rapidly from a temperature of about 700°C to around 50°C over a distance of <100 μm from the location of the resistive heater. A vacuum package based on localized aluminum/silicon-to-glass solder bonding technique was reported earlier (Cheng *et al.* 2002). The heater is formed using a layer of polysilicon and is isolated from the wafer using deposited dielectrics. A layer of aluminum is deposited on top of the polysilicon resistor. A glass wafer with an etched cavity is brought into contact with the wafer containing the polysilicon heater, and the aluminum bonding layer and the resistor are heated to raise the temperature locally in the bond region. At a temperature of about 800°C, the glass and the aluminum form a solder that flows over the nonplanar surface. An excellent bond is achieved in the bond region. This approach is quite useful because it allows to maintain a low temperature in the region where the sensitive devices are and to maintain a high temperature where the bonding region is located. The high temperature not only improves the quality of the bond, but also allows various materials to be used in the bonding region and to have these materials flow at the bond temperature and conform to the topology on the substrate, which is often nonplanar.

A different approach to localized heating has been recently reported by Mitchell and coworkers (Mitchell and Najafi 2006). In this approach instead of forming on-chip resistors in between the two wafers, heat is supplied to the backside of one of the two wafers to be bonded and removed from the backside of the second wafer as illustrated in **Figure 17**. Because most of the heat is drawn towards the heat sink, the noncontacting

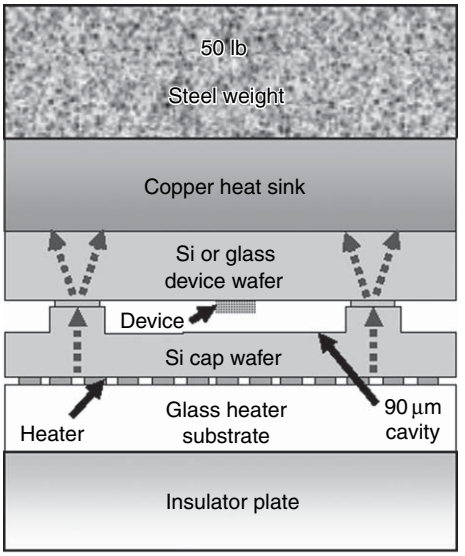


Figure 17 Cross-sectional view of the backside localized heating technique developed by Mitchell and coworkers One of the wafers is heated from the backside (in this figure it is called the silicon cap wafer), and the other wafer is cooled from its backside (this is called the silicon or glass device wafer). Heat is conducted through the bond regions (where the two wafers are in contact), thus heating these bond regions to a much higher temperature than the other noncontacting regions. (Source: Mitchell J, Najafi K 2006 Backside resistive localized heating for low temperature wafer-level bonding and packaging. *Tech. Dig., Solid-State Sensor, Actuator, and Microsystems*, Hilton Head Island, SC, USA, June 2006, pp. 352–5.)

regions are heated less than the bond regions. This method is attractive because it can be applied on a wide range of materials and at the wafer level without a need for on-chip resistors formed in between the two wafers. Using this technique, gold–silicon eutectic bonds were formed where the temperature of the bond regions were raised to about 400°C, and the sensitive device regions were heated to a temperature of only about 200°C.

1.09.4.3.2 Microwave, RF, and electromagnetic heating

Microwaves have long been used for heating various materials. Conductive layers absorb microwave energy and generate heat. This feature was first used by Budraa *et al.* (1999) to heat the bond region between two wafers to very high temperatures using microwave radiation. Microwave radiation selectively heats the metallic portion of the substrates. In addition microwave radiation (200 kHz–20 GHz) is transparent to a dielectric material that has a low dielectric constant, however energy is deposited in a skin depth (~1 pm) of a metal. This feature results in selective heating within the substrate. Heating of the silicon wafers is limited by the wafer material and is generally small for pure silicon. Budraa bonded silicon wafers with gold coatings for a MEMS packaging application by using microwave radiation in a single-mode cavity excited at 2.45 GHz. The experiment was performed in a high vacuum (~25 ptorr) to avoid the formation of plasma. The wafers were placed on top of each other without applying pressure inside a microwave transparent test tube. The quality of the bond was characterized by testing the hermeticity of a vacuum package formed using this bonding technology.

A variation of this microwave heating was used by Thompson *et al.* (2001, 2002) to heat and bond silicon wafers together. This work relies upon electromagnetic induction heating (EMIH), which uses electromagnetic radiation, ranging in frequency from a few megahertz to tens of gigahertz, to volumetrically heat silicon above 1000°C within a few seconds. Typical power requirements of silicon wafers of 75–100 mm in diameter fall between 900 and 1300 W. Note that in this work an intermediate layer need not be used. In order to get around the fact that silicon is transparent to electromagnetic radiation because of its small imaginary dielectric relaxation response to fields whose frequencies are below 100 GHz, Thompson took advantage of the ohmic response of silicon to an oscillating electromagnetic field; EMIH directly heats the silicon to temperatures at which the bonding can be achieved. This technique has successfully produced direct silicon wafer-to-wafer bonds without the use of an intermediate glue layer. Infrared images indicate void-free bonds that cannot be delaminated with knife-edge tests. In addition, four pairs of stacked wafers were bonded simultaneously in 5 min demonstrating the potential of multi-wafer bonds and high-throughput batch processing.

In yet another variation of microwave/RF heating, Bayrashev and Ziaie (2002) use RF radiation to heat a dielectric to achieve wafer bonding. RF dielectric heating uses a high-frequency electric field to send energy to an insulator placed between two metallic electrodes in a parallel plate capacitive configuration. Considerable energy can be imparted to dielectric molecules by agitating them as the field alternates each cycle. This energy appears as heat and since it is developed directly in the material, excellent uniformity and fast heating are possible. A

500-W, 14-MHz source was used to deliver RF power to the substrates. Two-inch diameter silicon wafers with 2- to 20-μm-thick polyimide or photoresist intermediate layers were successfully bonded (>95% bond area) in less than 7 min. The silicon substrate temperature remained below 280°C throughout the bonding process. The results of the pull tests indicate a bond strength of >1.5 MPa for fully cured substrates, which is greater than the strength of other low-temperature adhesive bonds. This method has several features. It is fast (<7 min bonding time), has a simple RF field configuration, is void-free, and the bond is stronger than that reported for adhesive bonds.

Finally, in yet another variation of the same basic concept, heating of a bonding layer, such as solder, using inductive heating was also reported by researchers (Cao *et al.* 2001, Yang *et al.* 2004). In this approach, a magnetic material is exposed to an electromagnetic field and is heated using the absorbed energy. The technique has been used for packaging of MEMS. In this approach an intermediate layer of a film that can absorb the electromagnetic field is needed. Yang *et al.* (2004) used this technique to bond a glass wafer to a silicon substrate using a nickel–cobalt layer and demonstrated that bond strength >18 MPa was measured. Both silicon–silicon and glass to silicon bonds were demonstrated.

1.09.4.3.3 Ultrasonic bonding

Ultrasonic wire bonding has been used extensively in the IC industry for bonding of conductive wires to pads on IC chips. The same basic technique has been used by Kim *et al.* (2002) and has been applied for attachment of a package die having a metal surface to a silicon substrate coated with another suitable metal layer. Ultrasonic bonding is a solid-phase welding process in which heat is generated by scrubbing two contacting surfaces during the bonding process, causing the materials to interdiffuse and bond together. The heat is insufficient to melt most metals and therefore melting temperatures and thermal conductivity of the materials are not decisive process factors. Ultrasonic vibration is usually generated at a frequency of about 60–80 kHz, and the force and pressure applied can be varied depending on the size of the bond ring. Bonds between aluminum–aluminum and indium–gold surfaces have been demonstrated (Yang *et al.* 2002). This technique is not easy to apply at the wafer level, and it has been mostly demonstrated for individual die or small area wafers.

1.09.4.3.4 Rapid thermal processing

A technology that is extensively used in the microelectronics industry is RTP. It is used for annealing and forming diffused junctions in ICs, among many other uses. In this approach, the wafer is rapidly, in the matter of a few seconds, heated to a high temperature and then cooled down. The limited exposure to these high temperatures is sufficient to achieve the desired objective of annealing or diffusion but not too long to damage the sensitive parts of the wafer. The same technique has been applied for wafer bonding and selective bonding in MEMS (Chiao and Lin, 2001, 2002). Surface-micromachined microresonators were sealed inside microcavities by an RTP aluminum-to-silicon nitride bonding. A glass wafer supporting 4-μm-thick aluminum rings was bonded to silicon wafer supporting a 0.4-μm silicon nitride layer using this technique. The assembled pair is loaded into an RTP chamber flushed with nitrogen gas, and then bonded within 10 s at 750°C. Due to the low thermal mass of the RTP system, the processing temperature rises from room temperature to 750°C and cools down to 350°C within 40 s. During the high-temperature bonding process, the molten aluminum reacts with silicon nitride, and a mechanically stable bond is formed at the aluminum–nitride interface. The measured aluminum-to-nitride bonding strength was found to be greater than the glass fracture toughness estimated at around 10 MPa. This technique has also been used for bonding other materials together (Chiao and Lin, 2001, 2002).

1.09.4.3.5 Laser heating

Lasers have long been used for welding, and have more recently been used for selective wafer bonding and welding of materials on wafers (Brown *et al.* 1999, Luo and Lin 2002, Luo *et al.* 2001, Mescheder *et al.* 2001, Sari *et al.* 2005, Stevenson 1988, Tao *et al.* 2004, Wild *et al.* 2001, Witte *et al.* 2002). The process is relatively straightforward. Two wafers with or without an intermediate layer are placed on top of one another and laser radiation is used to heat a selected region between the two wafers. One of the wafers or the material that is placed between the two wafers should be able to absorb the laser energy, and then be heated to a temperature where it melts and bonds with the materials surrounding it. Obviously, because of the high energy density of lasers, it is possible to quickly heat a selected region to very high temperatures and the literature cited demonstrates a number of different ways that this can be done. Laser welding and bonding are very attractive since the bond region

can be easily defined, and fast and localized heating can be easily achieved. The main drawback is that it is a serial process, although many emerging systems are capable of scanning large wafers for bonding. The display manufacturing industry uses this technology quite extensively for sealing and packaging of flat panel displays.

1.09.5 Summary and Conclusions

Wafer bonding has played a critical and extremely useful role in the development and successful commercialization of micromachined sensors and actuators. It is certain that it will continue to have an even more critical role in the development and the commercialization of future microsystems, ICs, nanoelectronics, and nano- and micromechanical systems. Its most important feature is that it allows the formation of complex microstructures and ICs through the precise and permanent attachment of multiple wafers, each of which can have other devices fabricated into it. Various wafer bonding technologies have been developed with each satisfying a different set of requirements and suitable for different applications. Because of this broad spectrum of applications and their specific requirements, it is difficult to compare these wafer bonding techniques. Nevertheless, Table 5 provides a summarized review and a comparison of wafer bonding technologies, and how they compare with respect to the set of desirable features that were presented in Section 1.09.1. Although many techniques have been developed and have matured, new technologies and improved understanding of the bonding mechanism of existing techniques are needed. Many of the existing technologies are constantly being pushed to reduce the bonding temperature or to improve the bonding quality and reduce cost. These efforts will continue and it is clear that we will see an expanding role for wafer bonding in many application areas.

Acknowledgments

The authors acknowledge the support of many organization and individuals who financially supported or participated in the development of some of the technologies reported here. The research results obtained at the University of Michigan were supported under various research grants funded by the US National Science Foundation (NSF), The National Institutes of Health (NIH), and the Defense Advanced Research Projects Agency (DARPA). The support of the faculty and the staff of the Solid-State Electronics Laboratory and the Michigan Nanofabrication Facility (MNF) at the University of Michigan is appreciated. Many former students and colleagues conducted the research that produced the results reported here. We especially thank Dr. T.K. Chou for his work on BCB bonding, Mr. Tzeno Galchev for his work on DBD plasma-activated bonding, Dr. Mehmet Dokmeci for his work on silicon–silicon bonding using evaporated

Table 5 Comparison of different wafer bonding technologies

	Bonding technique							
	Direct wafer bonding (DWB)			*Mediated wafer bonding (MWB)*				
	Anodic	*Fusion*						
Parameter	*Glass–Si*	*Wet activated*	*Plasma activated*	*Thermocompression*	*Eutectic*	*Solders*	*Glass seals*	*Polymers*
Temperature (°C)	< 400	< 1000	< 400	< 500	363	200–350	< 500	< 300
Voltage	1000 V	No	No	No	No	No	No	No
Surface	Very flat	Very flat			Flat	Flat		Flat
Bond strength	30–40 MPa	High	Good	High	17 MPa	14.6 MPa		4.3–16 MPa
Circuit compatibility	+	−−	++	++	++	+++	+++	+++
Hermeticity	+++	+++	+		+++	+	+++	−
Cost	Low	Medium	Medium	Low	Low	Low	Low	Low

glass, and Drs. Y. Mei and H. Lahiji for their work on silicon–gold eutectic bonding. This work made use of Engineering Research Centers Shared Facilities supported by the National Science Foundation under Award Number EEC-0096866.

References

Adachi T 2002 Eutectic reaction of gold thin-films deposited on silicon surface. *Surf. Sci.* **506**, 305–12

Adams D, Julies B A, Mayer J W, Alford T L 2003 The effect of metals (Au, Ag) and oxidizing ambient on interfacial reactions in Au/Si and Ag/Si systems. *Appl. Surf. Sci.* **216**, 163–8

Ageoroges C, Ye L 2002 *Fusion Bonding of Polymer Composites*. Springer, New York

Akatsu T, Hosoda N, Suga T, Rühle M 1999 Atomic Structure of Al/Al interface via the surface activated bonding method. *J. Mater. Sci.* **34**, 4133–9

Albaugh K B 1991 Electrode phenomena during anodic bonding of silicon to sodium borosilicate glass. *J. Electrochem. Soc.* **138(10)**, 3089–94

Alexe M and Gösele U (eds.) (2004) *Wafer Bonding: Applications and Technology* Springer, New York

Amirfeiz P, Bengtsson S, Bergh M, Zanghellini E, Borjesson L 2000 Formation of silicon structures by plasma-activated wafer bonding. *J. Electrochem. Soc.* **147**, 2693–8

Ando D, Oishi K, Nakamura T, Umeda S 1997 Glass direct bonding technology for hermetic seal package. *Proc. IEEE Micro Electro Mechanical Systems,* Nagoya, Japan, pp. 186–90

Anthony T R 1983 Anodic bonding of imperfect surfaces. *J. Appl. Phys.* **54(5)**, 2419–28

Barth P W 1990 Silicon fusion bonding for fabrication of sensors, actuators and microstructures. *Sens. Actuators* **23**, 919–26

Bayrashev A, Ziaie B 2002 Silicon wafer bonding with an insulator interlayer using RF dielectric heating. *15th IEEE Int. Conf. Micro Electro Mechanical Systems,* Las Vegas, NV, USA, pp. 419–22

Benoit O, Xuanxiong Z, Yannick B, Thomas P, Jean-Pierre R 2006 Effect of interfacial SiO_2 thickness for low temperature O_2 plasma activated wafer bonding. *Microsyst. Tech.* **V12**, 383

Berthold A, Nicola L, Sarro P M, Vellekoop M J 2000 Glass-to-glass anodic bonding with standard IC technology thin film as intermediate layers. *Sens. Actuators.* **A82**, 224–8

Bilenberg B, Nielsen T, Clausen B, Kristensen A 2004 PMMA to SU-8 bonding for polymer based lab-on-a-chip systems with integrated optics. *J. Micromech. Microeng.* **14**, 814–18

Blasquez G, Favaro P 2001 Silicon pyrex electrostatic bonding: Applicability to industrial microdevices production. *Technical Digest, The 11th IEEE Int. Conf. On Solid-State Sensors and Actuators (Transducers '01),* Munich, Germany, June 2001, pp. 194–7

Blasquez G, Favaro P 2002 Silicon glass anodic bonding under partial vacuum conditions: Problems and solutions. *Sens. Actuators A Phys.* **101(1)**, 156–9

Bosco N S, Zok F W 2004 Critical interlayer thickness for transient liquid phase bonding in the Cu–Sn system. *Acta Mater.* **52**, 2965–72

Bower R W, Chin F Y-J 1997 Low temperature direct silicon wafer bonding using argon activation. *Jpn. J. Appl. Phys. Part 2 (Lett.)* **36(5A)**, L527–8

Bower R W, Ismail M S, Roberds B E 1993 Low temperature Si_3N_4 direct bonding. *Appl. Phys. Lett.* **62(26)**, 3485–7

Briand D, Weber P, De Rooij N F 2004 Bonding properties of metals anodically bonded to glass. *Sens. Actuators A Phys.* **114(2–3)**, 543–9

Brooks A D, Donovan R P, Hardesty C A 1972 Low-temperature electrostatic silicon-to-silicon seals using sputtered borosilicate glass. *J. Electrochem. Soc. Solid-State Sci. Technol.* **119**, 545–6

Brown J, Maier N, Lee K Y, Ziegltrum L, St. Leger J 1999 Laser welding: Providing alignment precision and accuracy to substrate level packaging. *Proc. SPIE Int. Soc. Opt. Eng.* **3874**, 158–64

Bruson A, Gerl M 1982 Diffusion coefficient of Au in the liquid $Au_{0.81}Si_{0.19}$ alloy. *J. Appl. Phys.* **53**, 3616–19

Budraa N K, Jackson H W, Barmatz M, Pike W T, Mai J D 1999 Low pressure and low temperature hermetic wafer bonding using microwave heating. *IEEE Int. Conf. on Microelectromechanical Systems,* Orlando, FL, USA January 1999, pp. 490–2

Cao A, Chiao M, Lin L 2001 Selective and localized bonding using induction heating. *Tech. Dig., The 11th IEEE Int. Conf. Solid-State Sensors and Actuators, Transducers '01,* Munich, Germany, June 2001

Chang P-H, Berman G, Shen C C 1988 Transmission electron microscopy of gold–silicon interactions on the backside of silicon wafers. *J. Appl. Phys.* **63**, 1473–7

Chang-Chien P, Wise K D 2001 Wafer-level packaging using localized mass deposition. *Tech. Dig., 11th Int. Conf. Solid-State Sensors and Actuators, Transducers '01,* Munich, Germany, June 2001, pp. 182–5

Chavan A, Wise K 2001 Batch-processed vacuum-sealed capacitive pressure sensors. *J. Microelectromech. Syst.* **10(4)**, 580–8

Cheng Y T, Lin L, Najafi K 1998 Localized silicon fusion and eutectic bonding for MEMS fabrication and packaging. *Tech. Dig. Solid-State Sensor and Actuator Workshop,* Hilton Head, SC, USA, pp. 233–6

Cheng Y T, Lin L, Najafi K 1999 Localized bonding with PSG or indium solder as intermediate layer. *Digest, IEEE Microelectromechanical Systems (MEMS) Conference,* Orlando, FL, USA, January 1999, pp. 285–9

Cheng Y T, Lin L, Najafi K 2000a Localized silicon fusion and eutectic bonding for MEMS fabrication and packaging. *J. Microelectromech. Syst.* **9(1)**, 3–8

Cheng Y T, Lin L, Najafi K 2000b Fabrication and hermeticity testing of a glass–silicon package formed using localized aluminum/silicon-to-glass bonding. *Tech. Dig., 13th IEEE Int. Conf. Microelectromechanical Systems (MEMS),* Miyazaki, Japan January 2000, pp. 757–62

Cheng Y T, Lin L, Najafi K 2000c Localized silicon fusion and eutectic bonding for MEMS fabrication and packaging. *IEEE/ASME J. Microelectromech. Syst.* **9**, 3–8

Cheng Y T, Hsu W T, Lin L, Nguyen C T, Najafi K 2001a Vacuum packaging technology using localized aluminum/silicon-to-glass bonding. *Proc. 14th IEEE Microelectromechanical Systems Conference (MEMS 2001),* Interlaken, Switzerland, January 2001, pp. 18–21

Cheng Y T, Lin L, Najafi K 2001b A hermetic glass–silicon package formed using localized aluminum/silicon–glass bonding. *IEEE/ASME J. Microelectromech. Syst.* **10(3)**, 392–99

Cheng Y T, Hsu W T, Lin L, Nguyen C T, Najafi K 2002 Vacuum packaging using localized aluminum/silicon-to-glass bonding using localized aluminum/silicon-to-glass bonding. *IEEE/ASME J. Microelectromech. Syst.* **11**, 556–65

Chiao M, Lin L 2001 Hermetic wafer bonding based on rapid thermal processing. *Sens. Actuators Phys.* **91(3)**, 398–402

Chiao M, Lin L 2002 A wafer-level vacuum packaging process by RTP aluminum-to-nitride bonding. *Tech. Dig., Solid-State*

Sensor, Actuator and Microsystems Workshop, Hilton Head, SC, USA, pp. 81–4
Choi W B, Ju W B, Lee Y H, Haskard Y H, Sung M Y, Oh M H 1996 Anodic bonding technique under low-temperature and low-voltage using evaporated glass. *Proc. IVMC '96,* St. Petersburg, USA, pp. 427–30
Choi W, Ju B, Lee Y, Jeong J, Haskard M R, Lee N, Sung M, Oh M 1997 Experimental analysis on the anodic bonding with an evaporated glass layer. *J Micromech. Microeng.* **7**, 316–22
Chou T K A (2001) All-silicon micromachined acoustic ejector array for micro propulsion and flow control. Doctoral dissertation, University of Michigan
Chou T K A, Najafi K 2001 3D MEMS fabrication using low-temperature wafer bonding with benzocyclobutene (BCB). *Transducers '01, Eurosensors XV, 11th Int. Conf. Solid-State Sensors and Actuators,* Munich Germany, June 10–14, 2001, pp. 1570–2573
Christensen C, Bouwstra S 1996 Eutectic bonds on wafer scale by thin film multilayers. *Proc. SPIE Int. Soc. Opt. Eng.* **2879**, Austin, TX, pp. 288–90
Chuang R W, Lee C C 2002a High-temperature non-eutectic indium–tin joints fabricated by a fluxless process. *Thin Sol. Films* **414**, 175–9
Chuang R W, Lee C C 2002b Silver–indium joints produced at low temperature for high temperature devices. *IEEE Trans. Components Packaging Technol.* **25**, 453–8
Chung T R, Hosoda N, Takagi H, Suga T 1997 1.3-μm InGaAs/InP lasers on GaAs substrate fabricated by the surface activated wafer bonding method of room temperature. *Jpn. J. Appl. Phys.* **72**, 1565–6
Condra L W, Svitak J J, Pense A W 1975 The high temperature deformation properties of gold and thermocompression bonding. *IEEE Trans. Parts Hybrids Packaging* **PHP-11**, 290–6
Cook R F, Liniger E G 1999 Stress corrosion cracking of low dielectric constant spin-on-glass thin films. *J. Electrochem. Soc.* **146**, 4439–48
Cozma A, Puers B 1995 Characterization of the electrostatic bonding of silicon and pyrex glass. *J. Micromech. Microeng.* **5**, 98–102
Dalton R H 1956 Solder glass sealing. *J. Am. Ceramic Soc.* **39(3)**, 109–12
Deligeorgis G, Gallis S, Androulidaki M, Cengher D, Hatzopoulos Z, Alexe M, Dragoi V, Kyriakis-Bitzaros E D, Halkias G, Peiro F, Georgakilas A 2002 Properties of GaAs/Si heterostructure material fabricated by low temperature wafer bonding using a spin-on-glass intermediate layer. *2002 12th Int. Conf. Semiconducting and Insulating Materials (SIMC-XII-2002),* Smolenice, Castle, Slovakia June 30–July 5, 2002, pp. 125–8
Dokmeci M (1999) Hermetic glass–silicon micropackages and feedthroughs for neural prostheses. Doctoral dissertation, University of Michigan
Dokmeci M, Von Arx J, Najafi K 1997 Accelerated testing of anodically bonded glass–silicon packages in salt water. *Proc. 9th. Int. Conf. Solid-State Sensors and Actuators,* Chicago, IL, USA, June 1997, pp. 283–6
Dragoi V, Alexe M, Reiche M, Goesele U 1999 Low temperature direct wafer bonding of silicon using a glass intermediate layer. *Proc. Int. Semicond. Conf., CAS, **2**,* Sinaia, Romania, pp. 443–6
Dragoi V, Glinsner T, Mittendorfer G, Wieder B, Lindner P 2003 Adhesive wafer bonding for MEMS applications. *Smart Sensors, Actuators, and MEMS,* Maspalomas, Spain, May 19–21, 2003, pp. 160–7
Drost A, Klink G, Scherbaum S, Feil M 1998 Simultaneous fabrication of dielectric and electrical joints by wafer bonding. *Proc. 1998 Conf. Micromachined Devices and Components IV,* Santa Clara, CA, USA, September 21–22, 1998, pp. 62–71
Dunn M, Cunningham S J, Labossiere P E W 2000 Initiation toughness of silicon/glass anodic bonds. *Acta Mater.* **48(3)**, 735–44
Eaton W P, Risbud S H, Smith R L 1994 Silicon wafer-to-wafer bonding at $T<200^{\circ}$C with polymethylmethacrylate. *Appl. Phys. Lett.* **65**, 439–41
Ebert M, Bagdahn J 2004 Determination of residual stress in glass frit bonded MEMS by finite element analysis. *Proc. 5th Int. Conf. Thermal and Mechanical Simulation and Experiments in Microelectronics and Microsystems, EuroSimE 2004,* Brussels, Belgium, May 10–12, 2004, pp. 407–12
Eichler M, Michel B, Thomas M, Klages C P, Ruddy C, Reinecke H, Reiche M, Radu I, Gabriel M 2006 Ambient pressure plasma activation for low-temperature bonding. *2nd Int. Workshop Wafer Bonding for MEMS Technologies,* Halle, Germany, March 2006, pp. 35–6
Esashi M 1995 Packaged sensors, microactuators and three-dimensional microfabrication, *J. Robot. Mechatron.* **7**, 200–3
Esashi M, Nakano A, Shoji S, Hebiguchi H 1990 Low-temperature silicon-to-silicon anodic bonding with intermediate low melting point glass. *Sens. Actuators A Phys.* **23(1–3)**, 931–4
Esashi M, Ura N, Matsumoto Y 1992 Anodic bonding for integrated capacitive sensors. *Proc. of Micro Electro Mechanical Systems,* Travemünde, Germany, February 4–7, 1992, pp. 43–8
Farrens S N 2005 Packaging methods and techniques for MOEMS and MEMS. *Proc. SPIE* **5716**, 9–18
Farrens S N, Dekker J R, Smith J K, Roberds B E 1995 Chemical free room temperature wafer to wafer direct bonding. *J. Electrochem. Soc.* **142**, 3949–55
Fehlner F P, Dambaugh W H, Miller R A 1986 Refractory glass substrates for thin film transistors. *Proc. 6th Annu. Int. Display Conf.,* Tokyo, Japan, pp. 200–3
Folsom J, Haller J L, and Chilcott D W (1999) Method of protecting silicon wafers during wet chemical etching US Pat. 5 879 572
Frye R C, Griffith J E, Wong Y H 1986 A field-assisted bonding process for silicon dielectric isolation. *J. Electrochem. Soc.* **13(8)**, 1673
Furman B K, Mita S G 1992 Gold–Gold (Au-Au) thermocompression (TC) bonding of very large arrays. *Proc. 42nd Electronic Components and Technology Conference,* San Diego, CA, USA, May 18–20, 1992, pp. 883–9
Galchev T, Welch W C III, Najafi K 2007 Low-temperature MEMS process using plasma activated silicon-on-silicon (SOS) bonding. *Proc. IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS Conference),* Kobe, Japan, January 2007
Gale W F 2003 Transient liquid phase bonding of intermetallic compounds. *Thermec 2003 Processing and Manufacturing of Advanced Materials,* Madrid, Spain, July 7–11, 2003, pp. 1891–6
Go J S, Cho Y 1998 Experimental evaluation of anodic bonding process using Taguchi method for maximum interfacial fracture toughness. *Proc. 1998 IEEE 11th Annu. Int. Workshop Micro Electro Mechanical Systems,* Heidelberg, Germany, January 25–29, 1998, pp. 318–21
Goesele U, Tong Q, Schumacher A, Kraeuter G, Reiche M, Ploessl A, Kopperschmidt P, Lee T, Kim W 1999 Wafer bonding for microsystems technologies. *Sens. Actuators A Phys.* **74**, 161–8
Goyal A, Tadigadapa S, Islam R 2003 Solder bonding for microelectromechanical systems (MEMS) applications. *Reliability, Testing, and Characterization of MEMS/MOEMS II,* San Jose, CA, USA, January 27–29, 2003, pp. 281–8

Goyal A, Cheong J, Tadigadapa S 2004 Tin-based solder bonding for MEMS fabrication and packaging applications. *J. Micromech. Microeng.* **14**, 819–25

Han A, Oh K W, Bhansali S, Henderson T H, Ahn C H 2000 A low temperature biochemically compatible bonding technique using fluoropolymers for biochemical microfluidic systems. *13th Annu. Int. Conf. Micro Electro Mechanical Systems,* Miyazaki, Japan, pp. 414–28

Hansford D, Desai T A, Tu J, Ferrari M 1998 Biocompatible silicon wafer bonding for biomedical microdevices. *Proc. SPIE, BiOS '98: Micro and Nano-Fabricated Structures and Devices for Biomedical and Environmental Applications,* [3258-23], San Jose, CA, USA

Hara T, Kobayashi S, Ohwada K 1999 A new fabrication method for low-pressure package with glass–silicon–glass structure and its stability. *Int. Conf. Solid-State Sensors and Actuators, Transducers '99,* Sendai, Japan

Harendt C, Graf H G, Penteker E, Höfflinger B 1990 Wafer bonding: Investigation and *in situ* observation of the bond process. *Sens. Actuators A Phys.* **23(1–3)**, 927–30

Harendt C, Höfflinger B, Graf H G, Penteker E 1991 Silicon direct bonding for sensor applications: Characterization of the bond quality. *Sens. Actuators A Phys.* **25–27**, 87–92

Harendt C, Höfflinger B, Graf H G, Penteker E 1992a Wafer fusion bonding: Characterization of the bond quality. *Sens. Mater.* **3**, 173–87

Harendt, Graf H G, Höfflinger B, Penteker J E 1992b Silicon fusion bonding and its characterization. *J. Micromech. Microeng.* **2**, 113–16

Harpster T J 2005 Hermetic packaging and bonding technologies. Doctoral dissertation, University of Michigan

Harpster T J, Najafi K 2002 Long-term testing of hermetic anodically bonded glass–silicon packages. *15th IEEE Int. Conf. Micro Electro Mechanical Systems MEMS 2002,* January 20–24, 2002, Las Vegas, NV, USA, pp. 423–6

Harpster T J, Najafi K 2003 Field-assisted bonding of glass to Si–Au eutectic solder for packaging applications. *Proceedings, International IEEE Microelectromechanical Systems Conference (MEMS Conference),* Kyoto, Japan, January 2003

He G H, Lin L, Cheng Y T 1999 Localized CVD bonding for MEMS packaging. *10th Int. Conf. Solid-State Sensors and Actuators, Transducers '99, Tech. Dig.,* Sendai, Japan, June 1999, pp. 1312–15

Henmi H, Shoji S, Shoji Y, Yosimi K, Esashi M 1993 Vacuum package for microresonators by glass–silicon anodic bonding. *Int. Conf. Solid-State Sensors and Actuators, Transducers '93,* Yokahoma, Japan, pp. 584–7

Henmi H, Shoji S, Shoji Y, Yoshimi K, Esashi M 1994 Vacuum packaging for microsensors by glass–silicon anodic bonding. *Sens. Actuators* **A43**, 243–8

Hosoda N, Kyogoku Y, Suga T 1998 Effect of the surface treatment on the room-temperature bonding of Al to Si and SiO_2. *J. Mater. Sci.* **33**, 253–8

Huang A, Lew J, Xu Y, Tai Y-C, Ho C-M 2004 Microsensors and actuators for macrofluidic control. *IEEE Sensor. J.* **4**, 494–502

Humpston G, Jacobson D M 2004 *Principles of Soldering*. ASM International, Materials Park, OH

Humpston G, Jacobson D M, Sangha S P S 1994 Diffusion soldering for electronics manufacturing. *Endeavour* **18**, 55–60

Itoh, Okada T, Takagi H, Maeda R, Suga T 2003 Room temperature vacuum sealing using surface activated bonding method. *12th Int. Conf. Solid-State Sensors, Actuators and Microsystems, Transducers '03*, Boston, MA, USA, Vol. 2, pp. 1828–31

Iyer S S, Auberton-Herve A J 2002 *Silicon Wafer Bonding Technology for VLSI and MEMS Applications (EMIS Processing Series 1)*. INSPEC, Inc., London

Jeon N L, Chiu D T, Wargo C J, Wu H, Choi I S, Anderson J R, Whitesides G M 2002 Design and fabrication of integrated passive valves and pumps for flexible polymer 3-dimensional microfluidic systems. *Biomed. Microdevices* **4**, 117–21

Johnson A A, Johnson D N 1987 Metallurgy of gold–silicon alloys, welding, failure analysis, and metallography. *Microstructural Science,* vol. 14, edited by: Louthan M R Jr, LeMay I, Vander Voort G F, published by the American Society for Metals

Johnson D N, Biagtan E C, Johnson A A 1987 Determination of the solubility of silicon in gold and the gold–silicon eutectic composition using a new quantitative metallographic technique. *Scr. Metall.* **21**, 1689–92

Kanda Y, Matsuda K, Murayama C, Sugaya J 1990 The mechanism of field-assisted silicon glass bonding. *Sens Actuators* **A21–A23**, 939–43

Kato H 1987 A simple method for characterizing the bondability of metallization surfaces. *IEEE Trans. Component Packaging Manufact. Tech.* **CHMT-10**, 232–5

Kato H 1989 Eutectic reactions and textures of Au–Si alloy films on single-crystal silicon. *Jpn. J. Appl. Phys. Part 1 Regular Pap. Short Notes* **28**, 953–6

Katz A, Nakahara S, Geva M 1991 *In-situ* stress measurements of gold films on glass substrates during thermal cycling. *J. Appl. Phys.* **70**, 7342–8

Kikuchi H, Torigoe H, Shoji S 1997 Anodic bonding below 180°C for packaging and assembling of MEMS using lithium aluminosilicate–beta quartz glass–ceramic. *IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS),* Nagoya, Japan, January 26, 1997, pp. 482–7

Kim H 2006 An integrated electrostatic peristaltic micro gas pump with active microvalves. Doctoral dissertation, University of Michigan

Kim H, Najafi K 2005 Characterization of low-temperature wafer bonding using thin film Parylene. *IEEE J. Microelectromech. Syst.* **14(6)**, 1347–55

Kim H, Najafi K 2005 Characterization of parylene-assisted wafer bonding: Long-term stability and influence of process chemicals. *Tech. Dig., 13th Int. Conf. Solid-State Sensors, Actuators, and Microsystems, Transducers '05,* Seoul, Korea, June 2005, pp. 2015–18

Kim Y G, Pavuluri J K, White J R, Busch-Vishniac I J, Masada G Y 1995 Thermocompression bonding effects on bump-pad adhesion. *IEEE Trans. Component Packaging Manufact. Technol.* **18**, 192–9

Kim J, Chiao M, Lin L 2002 Ultrasonic bonding of In/Au and Al/Al for hermetic sealing of MEMS packaging. *Tech. Dig., IEEE 2002 Int. Conf. Micro Electro Mechanical Systems (MEMS 2002)* Las Vegas, NV, USA, January 2002, pp. 415–18

Kim S-A, Seo Y H, Cho Y H, Kim G, Jong U B 2006 Fabrication and characterization of a low-temperature hermetic MEMS package bonded by a closed-loop AuSn Solder Line. *Sens. Mater.* **18(4)**, 199–213

Ko W H, Suminto J T, Yeh G J 1985 Bonding techniques for microsensors. *Micromachining and Micro-packaging for Transducers*. Elsevier Science Publishers, Amsterdam, B.V., Vol. 20, pp. 41–61

Krause P, Sporys M, Obermeier E, Lange K, Grigull S 1995 Silicon to silicon anodic bonding using evaporated glass. *Int. Conf. Solid-State Sensors and Actuators. Transducers '95,* Stockholm, Sweden, pp. 228–31

Lan J, Chou T-K, Kanicki J 1997 Electrical and optical properties of low dielectric constant planarization polymer for high-aperture-ratio a-Si:H TFT-LCDs. *Proc. MRS Meet.* **471**, 27–33

Lasky J B 1986 Wafer bonding for silicon-on-insulator. *Appl. Phys. Lett.* **48** 78–80

Lee J-B, English J, Ahn C-H, Allen M G 1997 Planarization techniques for vertically integrated metallic MEMS on silicon foundry circuits. *Micromech. J. Microeng.* **7**, 44–54

Lee T M H, Hsing I, Liaw C Y N 2000 Improved anodic bonding process using pulsed voltage technique. *J. Microelectromech. Syst.* **9**, 469–73

Lee H S, Kim D S, Kwon T H 2003 A novel low temperature bonding technique for plastic substrates using X-ray irradiation. *Transducers '03,* Boston, MA, USA, pp. 1331–2334

Liang S L 1979 Argon concentration of RF sputtered silicon films. *Chin. J. Phys.* **17**, 102–6

Lin L, Cheng Y T, Najafi K 1998 Formation of silicon-gold eutectic bond using localized heating method. *Jpn. J. Appl. Phys. Part II* **11B**, 1412–14

Liu Y, Cui T, Sunkam R K, Coane P J, Vasile M J, Geoettert J 2003 Novel approach to form and pattern sol-gel polymethylsilsesquioxane-based spin-on glass thin and thick films. *Sens. Actuators B Chem.* **88**, 75–9

Lugscheider E, Bobzin K, Erdle A 2003 Solder deposition for transient liquid phase (TLP)-bonding by MSIP-PVD-process. *Surf. Coating Tech.* **174–175**, 704–7

Luo C, Lin L 2002 The application of nanosecond-pulsed laser welding technology in MEMS packaging with a shadow mask. *Transducers '01 Eurosensors XV,* Munich, Germany June 10–24, 2001, pp. 398–404

Luo C, Lin L, Chiao M 2001 Nanosecond-pulsed laser bonding with a built-in mask for MEMS packaging applications. *Tech. Dig., 11th IEEE Int. Conf. Solid-State Sensors and Actuators (Transducers '01),* Munich, Germany June 2001

Mao P, Han J 2005 Fabrication and characterization of 20 nm nanofluidic channels by glass-glass and glass–silicon bonding. *Lab Chip* **5**, 837–44

Mehra A, Ayon A A, Waitz I A, Schmidt M A 1999 Microfabrication of high-temperature silicon devices using wafer bonding and deep reactive ion etching. *J. Microelectromech. Syst.* **8(2)**, 152–60

Mei Y, Lahiji G R, Najafi K A 2002 Robust gold-silicon eutectic wafer bonding technology for vacuum packaging. *Solid-State Sensor, Actuator and Microsystem Workshop,* Hilton Head, SC, USA, pp. 772–4

Meindl J 1980 Biomedical implantable systems. *Science* **210**, 263–7

Meng E, Wang X-Q, Mak H, Tai Y-C 2000 A check-valved silicone diaphragm pump. *13th Annu. Int. Conf. Micro Electro Mechanical Systems (MEMS '00),* Miyazaki, Japan, pp. 62–7

Mescheder U M, Alavi M, Hiltmann K, Lizeau C H, Nachtigall C H, Sandmaier H 2001 Local laser bonding for low temperature budget. *Tech. Dig., 11th IEEE Int. Conf. Solid-State Sensors and Actuators (Transducers '01),* Munich, Germany, June 2001

Mitani K, Lehmann V, Stengl R, Feijoo D, Gösele U, Massoud H Z 1991 Causes and prevention of temperature dependent bubbles in silicon wafer bonding. *Jpn. J. Appl. Phys.* **30(4)**, 615–22

Mitchell J, Najafi K 2006 Backside resistive localized heating for low temperature wafer-level bonding and packaging. *Tech. Dig., Solid-State Sensor, Actuator, and Microsystems,* Hilton Head Island, SC, USA, June 2006, pp. 352–5

Mitchell J S, Lahiji G R, Najafi K 2005a Encapsulation of vacuum sensors on a wafer level package using a gold-silicon eutectic. *Tech. Dig., IEEE Int. Conf. Solid-State Sensors, Actuators, and Microsystems, Transducers '05,* Seoul, Korea, June 2005

Mitchell J S, Lahiji G R, Najafi K 2005b Reliability and characterization of micro-packages in a wafer level Au–Si eutectic vacuum bonding process. *Proc., ASME/Pacific Rim Tech. Conf. Exhibit. Integration and Packaging of MEMS, NEMS, and Electronic Systems, InterPack05,* San Francisco, CA, USA July 2005

Mizuno J, Ishida H, Farrens S, Dragoi V, Shinohara H, Suzuki T, Ishizuka M, Glinsner T, Lindner F P, Shoji S 2005 Cyclo-olefin polymer direct bonding using low temperature plasma activation bonding. *Int. Conf. Solid-State Sensors and Actuators, Transducers '05,* Seoul, Korea, June 5, 2005, pp. 1346–9

Morra M, Occhiello E, Marola R, Garbassi F, Humphrey P, Johnson D 1990 On the aging of oxygen plasma-treated polydimethylsiloxane surfaces. *J. Colloid Interf. Sci.* **137(1)**, 11–24

Najafi K 2001–2006 Wafer bonding. Lecture notes for Introduction to MEMS, A course taught at the University of Michigan

Najafi K 2003 Micropackaging technologies for integrated microsystems: Applications to MEMS and MOEMS. *Invited Plenary Talk, Proc. SPIE's Micromachining and Microfabrication Symposium,* San Jose, CA, USA, January 27–29, 2003, pp. 1–19

Nakanishi H, Nishimoto T, Nakamura N, Nagamashi S, Arai A, Iwata Y, Mito Y 1997 Fabrication of electrophoresis devices on quartz and glass substrates using a bonding with HF solution. *MEMS '97, IEEE,* Nagoya, Japan, January 26–30, 1997, pp. 299–304

Nakanishi H, Nishimoto T, Nakamura R, Yotsumoto A L, Shoji S 1998 Studies on SiO_2-SiO_2 bonding with hydrofluoric acid. *Proc. MEMS-98,* Heidelberg, Germany, pp. 609–14

Nese M, Hanneborg A 1993 Anodic bonding of silicon to silicon wafers coated with aluminium, silicon oxide, polysilicon or silicon nitride. *Sens. Actuators A,* **37–38**, 61–7

Niklaus F, Enoksson P, Kalvesten E, Stemme G 2000 Void-free full wafer adhesive bonding. *Proc. MEMS 2000,* Miyazaki, Japan, pp. 247–52

Niklaus F, Enoksson P, Kalvesten E, Stemme G 2001 Low-temperature full wafer adhesive bonding. *J. Micromech. Microeng.* **11(2)**, 100–207

Noh H-S, Hesketh P J 2000 Fabrication of parylene column for micro gas chromatograph. *The 7th Mechatronics Forum International Conference,* Atlanta, GA, USA

Noh H-S, Choi Y, Wu C-F, Hesketh P J, Allen M G 2003 Rapid, low-cost fabrication of Parylene microchannels for microfluidic applications. *12th Int. Conf. Solid-State Sensors, Actuators, and Microsystems,* Boston, MA, USA, pp. 798–801

Noh H-S, Moon K-S, Cannon A, Hesketh P J, Wong C P 2004 Wafer bonding using microwave of heating of parylene intermediate layers. *J. Micromech. Microeng.* **14**, 625–31

Oberhammer J, Niklaus F, Stemme G 2003 Selective wafer-level adhesive bonding with benzocyclobutene for fabrication of cavities. *Sens. Actuators A Phys.* **105(3)**, 297–304

Oh K W, Han A, Bhansali-S, Ahn C H 2002 A low-temperature bonding technique using spin-on fluorocarbon polymers to assemble microsystems. *J. Micromech. Microeng.* **12**, 187–291

Owen M J, Smith P J 1995 Plasma treatment of polydimethylsiloxane. In: Mittal K L (ed.), *Polymer Surface Modification: Relevance to Adhesion.* VSP, Utrecht, The Netherlands, pp. 3–15

Pantelidis D, Lee H J, Bravman J C 1999 Effect of microstructure and chemical bonding on the adhesion strength of a silicon/polymer interface for microelectronic packaging applications. *Mater. Res. Soc. Symp.–Proc.* **535**, 165–70

Park C B, Hong S M, Jung J B, Kang C S, Shin Y E 2001 A study on the fluxless soldering of Si-wafer/glass substrate using Sn-3.5 mass% Ag and Sn-37 mass% Pb solder. *Mater. Trans.* **42**, 820–4

Philips J. Res. Spec. Issue Direct Bonding **49**, 1

Plaza J A, Esteve J, Lora-Tamayo E 1998 Effect of silicon dioxide, silicon nitride and polysilicon layers on the electrostatic pressure during anodic bonding, *Sens. Actuators A* **67**, 181–4

Pomerantz D I 1968 Anodic bonding. *US Pat*. 3 397 278

Pomerantz D I 1969 Anodic bonding to liquid metals. *US Pat*. 3 470 348

Qingfeng X, Sasaki G, Fukunaga H 2002 Interfacial microstructure of anodic-bonded Al/glass. *J. Mater. Sci. Mater. Electron.* **13**, 83–8

Quenzer H J, Dell C, Wagner B 1996 Silicon–silicon anodic-bonding with intermediate glass layers using spin-on glasses. *Proc. 1995 9th Annu. Int. Workshop MEMS,* San Diego, CA, USA, February 11–15, 1996, pp. 272–6

de Reus R, Lindahl M 1997 Si-to-Si wafer bonding using evaporated glass. *Int. Conf. Solid-State Sensors, Actuators, and Microsystems, Transducers '97,* Chicago, IL, USA, June 1997, pp. 661–4

Rogers T 1992 Considerations of anodic bonding for capacitive type silicon/glass sensor fabrication. *J. Micromech. Microeng.* **2**, 164–6

Rogers T, Kowal J 1995 Selection of glass, anodic bonding conditions and material compatibility for silicon–glass capacitive sensors. *Sens. Actuators A Phys.,* **46–47**, 113–20

Rogge B, Moser D, Oppermann H, Paul O, Baltes H 1998 Solder-bonded micromachined capacitive pressure sensors. *Proc. 1998 Conf. Micromachined Devices and Components IV,* Santa Clara, CA, USA, September 21–22, 1998, pp. 307–15

Sari F, Rupf M, Gillner A, Poprawe R 2005 Advances in selective laser radiation bonding of silicon and glass for microsystems. *Proc. 3rd Int. WLT-Conf. Lasers in Manufacturing 2005,* Munich, Germany, June 2005

Sassen S, Kupke W, Bauer K 2000 Anodic bonding of evaporated glass structured with lift-off technology for hermetical sealing. *Sens. Actuators A Phys.* **83(1)**, 150–5

SchjØlberg-Henriksen K, Jensen G U, Hanneborg A, Jakobsen H 2003 Sodium contamination of SiO_2 caused by anodic bonding. *J. Micromech. Microeng.* **13(6)**, 845–52

Schmidt M A 1998 Wafer-to-wafer bonding for microstructure formation. *Proc. IEEE* **86**, 1575–85

Singh A, Horsley D A, Cohn M B, Pisano A P, Howe R T 1999 Batch transfer of microstructures using flip-chip solder bonding. *IEEE J. Microelectromech. Syst.* **8**, 27–33

Smith W F 1993 *Foundations of Materials Science and Engineering*. McGraw-Hill, New York

Sommadossi S, Litynska L, Zieba P, Gust W, Mittemeijer E J 2003 Transmission electron microscopy investigation of the microstructure and chemistry of Si/Cu/In/Cu/Si interconnections. *Mater. Chem. Phys.* **81**, 566–8

Spangler L J, Wise K D 1987 A technology for high-performance single-crystal silicon-on-insulator transistors. *IEEE Electron Device Lett.* **EDL-8(4)**, 137–139

Spangler L S, Wise K D 1990 A bulk silicon SOI process for active integrated sensor. *Sens. Actuators A Phys.* **24(2)**, 117–22

Sparks D, Queen G, Weston R, Woodward G, Putty M, Jordan L, Zarabadi S, Jayakar K 2001 Wafer-to-wafer bonding of nonplanarized MEMS surfaces using solder. *J. Micromech. Microeng.* **11(6)**, 630–4

Sparks D R, Najafi N, Massoud-Ansari S 2003 Chip-level vacuum packaging of micromachines using nanogetters. *IEEE Trans. Adv. Packaging* **26**, 277–82

Sparks D, Massoud-Ansari S, Najafi N 2004 Reliable vacuum packaging using Nanogetters^{TM} and glass frit bonding. In: Tanner D and Ramesham R (eds.) *Reliability,Testing and Characterization of MEMS/MOEMS III, Proc. SPIE, Vol. 5343.* SPIE, Bellingham WA

Stevenson P 1988 Sealing pressurized capsules by laser welding. *Welding Metal Fabricat.* **56(2)**, 80–2

Su Y-C, Lin L 2001 Localized plastic bonding for micro assembly, packaging and liquid encapsulation. *14th IEEE Int. Conf. Micro Electro Mechanical Systems,* Interlaken, Switzerland, pp. 50–53

Suga T, Otsuka K 2001 Bump-less interconnect for next generation system packaging. *IEEE, 51st Electronic Components and Technology Conf. (ECTC),* Orlando, FL, USA, May 29–June 1, 2001, pp. 1003–8

Suga T, Miyazawa K, Yamagata Y 1989 Direct bonding of ceramics and metals by means of a surface activation method in ultrahigh vacuum. *MRS Int. Meet. Adv. Mater. Mater. Res. Soc.,* **8**, 257–62

Suga T, Fujiwaka T, Sasaki G 1993 Surface activated bonding and its application on microbonding at room temperature. *9th Eur. Hybrid Microelectron. Conf.,* Nice, France, June 1993, pp. 314–21

Suni T, Henttinen K, Suni I, Mäkinen J 2002 Effects of plasma activation on hydrophilic bonding of Si and SiO_2. *J. Electrochem. Soc.* **149**, G348–51

Takagi H, Kikuchi K, Maeda R, Chung T R, Suga T 1996 Surface activated bonding of silicon wafers at room temperature. *Appl. Phys. Lett.* **68**, 2222–4

Takagi H, Maeda R, Hosoda N, Suga T 1999 Room-temperature bonding of lithium niobate and silicon wafers by argon-beam surface activation. *Appl. Phys. Lett.* **74**, 2387–9

Takahashi K, Ogura H, Sagawa M 2000 Miniaturized millimeter-wave hybrid IC technology using non-photosensitive multi-layered BCB thin films and stud bump bonding. *IEICE Trans. Electron.* **E83-C(11)**, 2029–37

Taklo M M V, Storas P, Schj⌀lberg-Henriksen K, Hasting H K, Jakobsen H 2004 Strong, high-yield and low-temperature thermocompression silicon wafer-level bonding with gold. *J Micromech. Microeng.* **14**, 884–90

Tao Y, Malshe A P, Brown W D 2004 Selective bonding and encapsulation for wafer-level vacuum packaging of MEMS and related micro systems. *Microelectron. Reliab.* **44**, 251–2

Tatic-Lucic S, Ames J, Boardman B, McIntyre D, Jaramillo P, Starr L, Lim M 1997 Bond-quality characterization of silicon-glass anodic bonding. *Sens. Actuators A Phys.* **60**, 223–7

Thompson K, Gianchandani Y B, Booske J, Cooper R 2001 Si–Si bonding using RF and microwave radiation. *11th Int. Conf. Solid-State Sensors and Actuators,* Munich, Germany, June 2001, pp. 226–9

Thompson K, Gianchandani Y B, Booske J, Cooper R F 2002 Direct silicon–silicon bonding by electromagnetic induction heating. *J. Microelectromech. Syst.* **11(4)**, 285–92

Tiensuu L, Bexwell M, Schweitz J, Smith L, Johansson S 1994 Assembling three-dimensional microstructures using gold–silicon eutectic bonding. *Sens. Actuators* **A45(3)**, 227–36

Tong Q Y, Gösele U 1999 *Semiconductor Wafer Bonding: Science and Technology*. John Wiley & Sons Inc., New York , ISBN 0-471-57481-3

Tsau H 2003 *Fabrication and Characterization of Wafer-level Gold Thermocompression Bonding*. MIT, Cambridge, MA

Tsau H, Spearing S, Schmidt M A 2004 Characterization of wafer-level thermocompression bonds. *IEEE J. Microelectromech. Syst.* **13(6)**, 963–71

Valero L 1984 The fundamentals of eutectic die attach. *Semicond. Int.* **7**, 236–41

Wallis G 1970 Direct-current polarization during field-assisted glass-metal sealing. *J. Am. Ceram. Soc.* **53**, 563–7

Wallis G, Pomerantz D I 1969 Field-assisted glass–metal sealing. *J. Appl. Phys.* **40**, 3946–9

Wang C Y, Lee C C 1992 A eutectic bonding technology at a temperature below the eutectic point. *Proc. 42nd Electron. Components and Technology Conference,* San Diego, CA, USA, May 18–20, 1992, pp. 502–7

Weichel S, de Reus R 1998 Silicon-silicon wafer bonding using evaporated glass. *Sens. Actuators* **A70(1–2)**, 179–84

Welch W C III, Najafi K 2005 Transfer of metal MEMS packages using a wafer-level solder sacrificial layer. *18th IEEE Int. Conf. Micro Electro Mechanical Systems,* Miami Beach, FL, USA, January 30–February 3, 2005, pp. 584–7

Welch W C III, Dokmeci M, Yazdi N, Najafi K 2005a A flux-free Pb-Sn solder bonding technology for wafer-level chip-scale vacuum packaging. *Proc. ASME/Pacific Rim Tech. Conf. Exhibit. Integration and Packaging of MEMS, NEMS, and Electronic Systems, InterPack05,* San Francisco, CA, USA, July 2005

Welch W C III, Chae J, Najafi K 2005b Transfer of metal MEMS packages using a wafer-level solder transfer technique. *IEEE Trans. Adv. Packaging (IEEE Trans. Components Packaging Manufact. Technol. Part B Adv. Packaging)* **28(4)**, 643–9

Welch W C III, Chae J, Lee S, Najafi K 2005c Transient liquid phase (TLP) bonding for MEMS packaging applications. *Tech. Dig., IEEE Int. Conf. Solid-State Sensors, Actuators, and Microsystems, Transducers '05,* Seoul, Korea, June 2005

Wiegand M, Reiche M, Gosele U 2000 Time-dependent surface properties and wafer bonding of O_2-plasma-treated silicon (100) surfaces. *J. Electrochem. Soc.* **147**, 2734–40

Wild M J, Gillner A, Poprawe R 2001 Advances in silicon to glass bonding with laser. *Proc. SPIE–Int. Soc. Opt. Eng.* **4407**, 135–41

Witte R, Herfurth H, Heinemann S 2002 Laser joining of glass with silicon. *Proc. SPIE–Int. Soc. Opt. Eng.* **4637**, 487–95

Wolffenbuttel R F 1997 Low-temperature intermediate Au–Si wafer bonding: eutectic or silicide bond. *Sens. Actuators A Phys.* **62**, 680–686

Wolffenbuttel R F, Wise K D 1994 Low-temperature silicon wafer-to-wafer bonding using gold at eutectic temperature. *Sens. Actuators* **A43(1–3)**, 223–9

Xiao Z-X, Wu G-Y, Zhang D, Zhang G, Li Z-H, Hao Y-L, Wang Y-Y 1998 Silicon/glass wafer-to-wafer bonding with Ti/Ni intermediate bonding. *Sens. Actuators A* **71**, 123–6

Xu Y, Tai Y-C, Huang A, Ho C-M 2003 IC-integrated flexible shear-stress sensor skin. *J. Microelectromech. Syst.* **12**, 740–7

Yang L, Hosoda N, Suga T 1997 Investigation on the interface microstructure of stainless steel/aluminum joints created by the surface activated bonding method. *Interf. Sci.* **5**, 279–86

Yang X, Grosjean C, Tai Y-C 1999 Design, fabrication, and testing of micromachined silicone rubber membrane valves. *J. Microelectromech. Syst.* **8**, 393–402

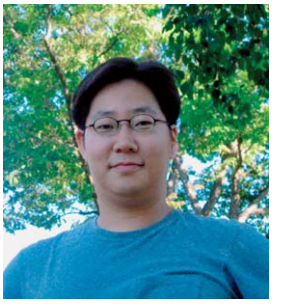
Yang H A, Wu M, Fang W 2004 Localized induction heating solder bonding for wafer level MEMS packaging. *17th IEEE Int. Conf. Microelectromechanical Systems (MEMS),* Masstricht, Netherlands, January 2004, pp. 729–32

Ziaie B, Arx J A V, Dokmeci M R, Najafi K A 1996 Hermetic glass–silicon micropackage with high-density on-chip feedthroughs for sensors and actuators. *IEEE/ASME J. Microelectromech. Syst.* **5(3)**, 166–79

Biographies

Khalil Najafi received his BS, MS, and the Ph.D. degrees in 1980, 1981, and 1986, respectively, all in Electrical Engineering from the University of Michigan, Ann Arbor, MI. From 1986 to 1988 he was employed as a research fellow, from 1988 to 1990 as an assistant research scientist, from 1990 to 1993 as an assistant professor, from 1993 to 1998 as an associate professor, and since September 1998 as Professor in the Solid-State Electronics Laboratory, Department of Electrical Engineering and Computer Science, University of Michigan. His research interests include micromachining technologies, micromachined sensors, actuators, and MEMS; analog integrated circuits; implantable biomedical microsystems; micropackaging; and low-power wireless sensing/actuating systems. Dr. Najafi was awarded a National Science Foundation Young Investigator Award for the year 1992–1997, and has been active in the field of solid-state sensors and actuators for more than 20 years. He has been involved in several conferences and workshops dealing with microsensors, actuators, and microsystems, including the *International Conference on Solid-State Sensors and Actuators, The Hilton-Head Solid-State Sensors and Actuators Workshop* and *The IEEE/ASME Micro Electromechanical Systems (MEMS) Conference*. Dr. Najafi has served as the editor for several journals, including the *IEEE Transaction on Electron Devices*, *Journal of Solid-State Circuits*, *Transactions on Biomedical Engineering*, and is currently an associate editor for the *IEEE Journal of Micro Electromechanical Systems*, the *Journal of Micromechanics and Microengineering*, Institute of Physics Publishing, and an editor for the *Journal of Sensors and Materials*. He is a fellow of the IEEE and the AIBME.

Timothy J. Harpster received his BS degree in Electrical Engineering with a minor in Physics from the University of Dayton Ohio in 1998. He was the recipient of the University of Dayton Presidential scholarship and NAECON (National Aerospace and Electronics Conference) scholarship. During his undergraduate career, he interned at Intek Inc. where he participated in software development for the RheoVac® Diagnostic System, a real-time air in-leak measurement system for power plant condensers. He entered the University of Michigan in September 1998 to pursue his Ph.D. in Electrical Engineering with a major in Solid State Electronics and a minor in Circuits and Microsystems. He received his MS degree in May 2000 and his Ph.D. degree in December 2005 for his work on the development of passive wireless sensor technologies for packaging applications and glass silicon packaging technologies for implantable devices. Dr. Harpster is currently Director of Engineering at Intek Inc. in Westerville, OH. Intek is a manufacturing and research company that has established itself as a leader in low flow rate measurement instrumentation, a preferred supplier of unique aerospace payloads, the manufacturer of RheoVac air in-leak and condenser monitoring systems, and expert in steam surface condenser theory and design.



Hanseup Kim (IEEE S '00) received his BS degree in Electrical Engineering from the Seoul National University, Korea, in 1997, and the MS and Ph.D. degree in Electrical Engineering from the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MA, in 2003 and 2006, respectively, where he is currently working as a postdoctoral research fellow. Since September 1999, he has been with the Center for Wireless Integrated MicroSystems (WIMS), University of Michigan, where he has been developing a high-performance polymer-based micropump with integrated microvalves for microfluidics applications including a high-speed micro gas chromatography. His research interests include microelectromechanical systems (MEMS), biomicrosystems, lab-on-a-chip, microfluidics, microsensors and actuators, micro/nanofabrication, and polymer technology development. He, with four other coauthors, was the recipient of the 2001 Best Paper and the 1st place award in the 38th Student Design Contest of Design Automation Conference (DAC 2001). Till date, he has authored and coauthored over 20 papers in refereed journals and

conferences, held one patent, and served as a technical reviewer for the *Journal of Microelectro-mechanical Systems*, *IEEE Transactions on Electron Devices (TED)*, and *Sensors and Actuators Journal (SNA)*.

Jay S. Mitchell was born in Madison, WI, in 1976. Jay received his BS and MS from Case Western Reserve University, Cleveland, OH, in 1999 and 2000, respectively, where he investigated the mechanical properties of silicon carbide for MEMS applications. In 2000 and 2001, he worked for Movaz Networks in the testing and design of micromirrors for telecommunications applications. In the fall of 2002, he began the Ph.D. program at the University of Michigan in mechanical engineering and is currently working on Au–Si eutectic vacuum packaging and localized heating methods for micropackaging. He also interned for Sandia National Laboratories in the summer of 2005 working in micropackaging.

Neil Welch was born in Jacksonville Beach, FL, in 1979. He received his BS in Electrical Engineering from Lehigh University in June 2001. During his time at Lehigh, he spent a summer at the Department of Energy's Princeton Plasma Physics Lab developing code to facilitate the visualization of scientific data. The following summer, he was part of the National Science Foundation (NSF) Research Experience for Undergraduates program at Lehigh's Sherman Fairchild Research Lab. At the lab, his work focused on reducing the number of oxide traps at the surface of a silicon carbide MOS capacitor. He enrolled at the University of Michigan in the fall of 2001 and received his master's degree in May 2003. In 2004, he was awarded the Outstanding Student Leadership Award for his efforts as part of the NSF Engineering Research Center for Wireless Integrated MicroSystems. He is currently pursuing his Ph.D. research into applications of solder for vacuum and hermetic packaging of MEMS. His other interests include miniature power sources and bioMEMS.

1.10 Electrodeposition

Matthew W. Losey[1] **and James J. Kelly**[2], [1]Touchdown Technologies, Baldwin Park, CA, USA, [2]IBM/TJ Watson Research Center, Yorktown Heights, NY, USA

Published by Elsevier B.V.

Glossary

Damascene Process for forming copper interconnects whereby a lithographically patterned dielectric is overfilled with metal by electrodeposition and then planarized to create metal patterns (lines and vias) within the dielectric

LIGA Lithographie, galvonoformung, abformtechnik, German acronym associated with the process of lithographically defining molds for electroforming

MEMS Microelectromechanical Systems

PMMA Polymethylmethacrylate, resist material used in X-ray lithography or LIGA

Wagner number, ***Wa*** Ratio of resistances in an electrochemical process that correlates to uniformity. Resistances may be due to the seed layer, charge transfer process, diffusion, or electrolyte conduction

1.10.1 General Principles

The electrodeposition of metals, alloys, and semiconductors has found broad use in the fabrication of microsystems and microelectromechanical systems (MEMS). While a large number of metals found in microsystems can be electroplated from aqueous electrolytes (Ni, Cu, Au, Pt, Fe, Pb, to name a few), a few metals cannot, notably Al, Ti, and pure W. Nickel and its high-strength alloys constitute some of the more useful materials available to microsystem fabrication by electrodeposition, owing to their use in forming mechanical and magnetic elements such as precision gears, latches, motors, and flexure spring arms (Malek and Saile 2004). Electroplating is a versatile technique spanning the range of coatings and thin films less than a micron to thick electroformed mechanical elements millimeters tall. Materials are deposited near ambient conditions without the need for expensive vacuum equipment. Deposition rates can be much faster than vapor deposition methods and film thicknesses can be as high as a millimeter or two. The essential elements of the process are the cathode, or the workpiece to be plated, an anode or counter electrode, and an electrolyte that possesses the metal ion in reducible form. Current is supplied to the workpiece, which for MEMS is often a flat substrate metallized by vapor deposition. This thin metal coating forms a conductive seed layer to which electrical contact is made. The seed layer can cover micropatterned films of insulator, as in the copper Damascene process for integrated circuit interconnects. Or, an insulating, patterned layer for through-resist plating can mask the seed layer. The latter approach is more often utilized for making freestanding parts, exemplified by the LIGA (lithographie, galvonoformung, abformtechnik, German acronym associated with the process of lithographically defining molds for electroforming) and X-ray lithographic methods.

In this chapter, some of the fundamentals of electrodeposition are reviewed first; there are numerous references that provide far more detail (Bard and Faulkner 2001, Dini 1993, Lowenheim 1995, Schlesinger and Paunovic 2000). In addition to kinetics, some general methods (including common issues) are detailed to give the MEMS practitioner a starting point for conducting an electrodeposition process. Some of the ramifications of plating through lithographically generated patterns are discussed. Although there is a rich literature and much recent progress has been made on the modeling and simulation of patterned electrodeposition, that topic is beyond the scope here. Instead, we cover some of the general principles and focus on the properties and processes of electrodeposits most relevant for MEMS. Finally, an assortment of materials, recipes, and applications are provided as a starting point for electrodeposition for MEMS, with emphasis on nickel and its alloys for application to micromechanical elements.

1.10.1.1 Kinetics

The rate at which a metal is deposited is related to the current density by Faraday's law, which states that the extent of an electrochemical reaction that occurs is proportional to the electrical charge passed. Each metal has an electrochemical equivalent Z, or the proportionality constant relating deposition rate to current density. For example, the growth rate can be expressed as follows:

$$\text{Rate}\left(\mu\text{m h}^{-1}\right) = 3.6 \times 10^4 \times \frac{Z \times \text{CE} \times \mathcal{J}}{\rho} \quad [1]$$

where CE is the current efficiency, $\mathcal{J}$ is the current density (in mA cm^{-2}), ρ is the density of the metal (in g cm^{-3}), and

$$Z = \frac{M}{nF} \quad [2]$$

where M is the molecular weight (in g mol^{-1}), n is the number of electrons involved, and F is Faraday's constant, 96 485 C mol^{-1}.

For the case of nickel electrodeposition Ni^{2+}, $n = 2$, $M = 58.71\,\text{g mol}^{-1}$, $Z = 3 \times 10^{-4}\,\text{g C}^{-1}$ and the linear growth rate ($\rho = 8.9\,\text{g cm}^{-3}$) at a current density of 10 mA cm^{-2} is 11.5 μm h^{-1}, assuming the CE is 95%, which is typical for Ni. The CE is the ratio of charge consumed toward the deposition reaction to the total charge consumed. Side reactions, such as the reduction of hydrogen, detract from the CE. Copper electrodeposition usually has nearly 100% CE with current densities up to 50 mA cm^{-2}. Nickel electrodeposition generally has 95–100% CE, but unlike copper, CE will drop at current densities lower than 3 mA cm^{-2} CE for gold electrodeposition can be as low as 50%, depending on the bath formulation.

The upper limit to the rate of metal deposition is determined by the mass transport, or diffusion, of the metal ion species to the cathode surface. In the case of

a planar surface with no convection or mixing of the electrolyte, the current decreases with time as metal ions are consumed near the vicinity of the electrode. For this case, the time-varying current density is described by the Cotrell equation,

$$j(t) = \frac{nFD^{1/2}C}{\pi^{1/2}t^{1/2}} \quad [3]$$

where D is the diffusion coefficient for the electroactive metal species and C is the concentration of the metal in the bulk solution.

Eventually, as the metal ions are consumed at the cathode surface, the deposition rate, or current density, becomes limited by the rate of diffusive flux of ions to the surface. When the length scale for diffusion, δ, is known, such as the height of a resist feature or for a rotating disk electrode, the limiting current can be described by

$$j_L = \frac{zFDC}{\delta} \quad [4]$$

For extreme cases of resist height (>100 μm) and plating rates, it has been shown that natural convection can be a significant component of the mass transport (Nilson and Griffiths 2003). Under these conditions, the limiting current can be larger than that expected by diffusion alone.

For a rotating disk electrode (the preferred tool of the electrochemist and the arrangement in many commercial electrodeposition tools), the diffusion layer thickness, δ, can be precisely described and the limiting current is described by the Levich equation,

$$j_{L,\,RDE} = \frac{0.62nFD^{2/3}\omega^{1/2}C}{v^{1/6}} \quad [5]$$

where ω is the angular frequency of rotation and v is the kinematic viscosity.

In some cases, it is preferable to control an electroplating process by the potential, rather than current. This can be beneficial if undesired side reactions are to be avoided or the plated area is quickly changing. The rates of electrochemical processes are strongly dependent on the potential, in fact exponential. The conventional approach to describing electrode kinetics in the absence of mass transfer limitations is the Butler–Volmer model:

$$j = j_0\left[\exp\left(\frac{-F\alpha\eta}{RT}\right) - \exp\left(\frac{(1-\alpha)F\eta}{RT}\right)\right] \quad [6]$$

where j_0 is the exchange current density, α is known as the transfer coefficient, and η is the overpotential or the difference between the applied potential and the equilibrium potential. The first term refers to the cathodic process and the second term refers to the anodic process. For irreversible electrochemical processes where the cathodic overpotential is large, the anodic component is negligible and the Butler–Volmer equation reduces to the Tafel equation:

$$\begin{aligned} \eta &= a + b\log j \\ a &= \frac{2.3RT}{\alpha F}\log j_0 \\ b &= -\frac{2.3RT}{\alpha F} \end{aligned} \quad [7]$$

A Tafel plot (log i versus η) is used to extract these kinetic parameters and glean insight toward the nature of the kinetic steps involved.

In describing the overpotential, the equilibrium potential E_{eq} is related to the concentration of the electroactive species by the Nernst equation, which for the reduction of a metal to its solid in an ideal solution is:

$$E_{eq} = E_0 + \frac{RT}{nF}\ln C \quad [8]$$

where E_0 is the standard potential.

1.10.1.2 Principles of Alloy Deposition

The performance (mechanical properties, corrosion resistance) of many electrodeposited materials can be greatly enhanced as a result of alloying. High-strength nickel can be made thermally stable with small amounts of manganese, a range of useful magnetic materials are possible with alloys of nickel and iron, and alloys of bismuth and tellurium possess superior thermoelectric properties. The principles of electrodepositing alloys and the broad range of alloys available are covered in several comprehensive references (Brenner 1963, Landolt 1994). A starting basis for understanding the relationship between electrolyte composition and resulting alloy content of the deposit is to consider the equilibrium potentials and relative concentrations of the metal ions. At first glance, metals whose equilibrium potentials fall close together can be codeposited to form an alloy. Alloy electrodeposition can be distinguished as normal codeposition when the resulting ratio of the metals in the alloy is that expected according to the equilibrium potentials and concentrations. Tin–lead alloys used in printed circuit boards as solderable coatings fall in this category (Jordan 2000). The equilibrium potentials are close (Pb = −0.126 V, Sn = −0.136 V) and the full range of

alloy composition is possible. Lead is a little more noble and will deposit slightly faster than tin, but for the most part the concentration of the electrolyte dictates the alloy content.

Metals whose equilibrium potentials differ significantly can still be made to codeposit by reducing the activity of the more noble metal. The classic example is Cu–Zn, where the equilibrium potentials are considerably different (Cu = +0.342, Zn = −0.762). In electrolytes of the simple salts, alloy deposition is impossible, as copper will deposit to the virtual exclusion of zinc. However, Cu–Zn alloys are readily plated using electrolytes where the copper species is complexed, typically as a cyanide.

Unfortunately, one of the most useful alloys for MEMS, Ni–Fe (and Ni–Co), falls in the category of anomalous codeposition. In this category of abnormal alloy electrodeposition, the less noble metal, iron, preferentially deposits and the alloy content does not prescribe to the usual equilibrium potential relationships (Ni = −0.257, Fe = −0.447). For this reason, the iron concentration in the bath is much lower than nickel and the process is highly sensitive to mass transport and mixing effects.

Another category of abnormal alloy deposition is induced codeposition, where one element is codeposited as a result of the catalytic activity or reduced overpotential of the alloy surface. Tungsten and molybdenum are examples of metals that can be electrodeposited as part of an alloy, but not as a pure metal, Ni–W being the example relevant to MEMS. Ni–W has many desirable properties including high hardness, improved corrosion and wear resistance, and thermal stability owing to the high melting point of tungsten. Historically, Ni–W has been difficult to process owing to the high temperatures (80–90°C) and high pH (8–11) required of the electrolyte. Recent work involving a sulfamate electrolyte has better enabled integration into a MEMS process flow (Slavcheva *et al.* 2005).

For the case of Ni–Mn (Ni = −0.257, Mn = −1.18), the manganese content is very low as expected, usually less than 1 wt.%. In fact, given the large difference in equilibrium potentials and the lower manganese concentrations in the electrolyte used (see **Table 2**), it is surprising that any manganese is codeposited. It has been argued that manganese is incorporated as a result of adsorption onto the nickel surface rather than a codeposition reaction (Atanassov and Schils 1999). This is consistent with the observed grain-refining effect that the manganese has on the resulting film.

1.10.1.3 Methods

Electroplating can be a simple laboratory exercise involving a beaker and a power supply or an elaborate operation with means for agitation, filtration, electric field shaping, temperature control, and multistep or pulsed current programs. As a matter of practicality, most industrial electroplating processes are galvanostatic, rather than potentiostatic. This obviates the need for a reference electrode and the variation in resistance that may occur as a result in subtle differences in cell configuration. The current may either be direct or pulsed. It is common practice for substrates to be immersed hot, that is with potential applied and current initiating immediately upon entry. This prevents any dissolution or deleterious oxidation of the seed by the electrolyte media.

1.10.1.3.1 Thickness uniformity

The most common issue confronted once the desired chemistry and metal properties are established is uniformity of the deposit across the plated area of the substrate. Although subsequent removal steps, such as polishing, can provide planarization, it is generally undesirable to have extreme variations in plating rate. The desired properties of the deposit, be they mechanical, electrical, or just cosmetic, vary with the current density and hence the deposition rate. An example of an extreme case, illustrated in **Figure 1**, demonstrates what can happen as deposition nonuniformity builds upon itself. If the plating surface is allowed to surpass the height of the mold or

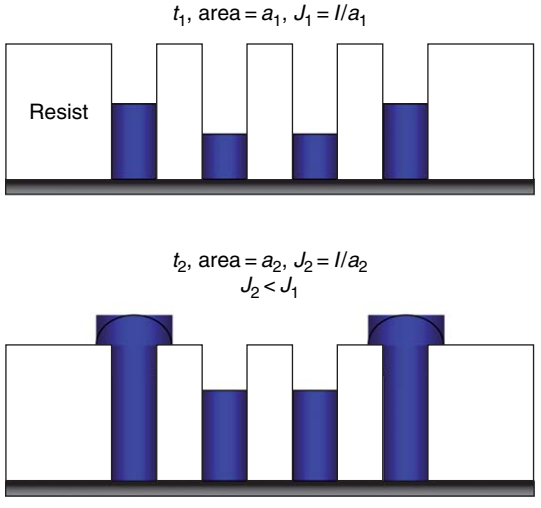


Figure 1 Overplating features increases surface area and reduces the overall current density, changing the properties of the deposit.

resist at any point, the plating area begins to increase and the current density drops, resulting in changing deposit properties for those features still filling. If the polarization resistance is low, the protruding overplated areas may attract a disproportionate amount of current and make it extremely difficult to fill the remaining underplated features.

There are several origins of nonuniformity, and there are usually several clever approaches that can be taken to address them. The first, described as the terminal effect results when the resistance of the metallization or seed layer is significant compared to other resistances such as conduction through the electrolyte or the charge-transfer process itself. Deposition rates are greater near the edges or points of the substrate where electrical contact is made. This can become an issue for large substrates and thin seed layers (such as copper Damascene plating on 300-mm-diameter substrates with sputtered copper seed layers <1000 Å thick). For most MEMS applications, a conductive metallization layer of a micron or so is sufficient to eliminate the problem. If not, the electrolyte can be made more resistive, for example, by reducing the sulfuric acid concentration in the case of copper plating. To describe the relative influence of these various resistances, a dimensionless Wagner number (Wa) is often employed

$$Wa = \frac{\mathrm{d}\eta/\mathrm{d}\mathcal{J}}{\rho_e L}$$

which in this case describes the relative influence of the polarization resistance, or the inherent resistance of the charge transfer reaction, to the electrical resistance of the electrolyte and ρ_e is the electrolyte resistivity, and L is a characteristic length for conduction. For $Wa \gg 1$, indicating that polarization resistance dominates, the current distribution is uniform and the secondary current distribution is said to predominate. For $Wa < 1$, the primary current distribution predominates, that is the current density is a result of the potential distribution of the cell.

In the cases where polarization resistance is relatively insignificant and the cell geometry dictates the potential field, numerous approaches have been used to improve the current density distribution. One way is to insert an insulating barrier at the cathode, or a cathode shield, to help shape the field lines and direct the current toward or away the center of the substrate. Another common approach is to add auxiliary features (known as a robber or current thief) at the edge of the substrate to absorb the nonuniformity intrinsic to any edge. Furthermore, sacrificial auxiliary features can be used to improve the area density, by provided isolated features of a pattern with neighboring areas. The size and placement of the auxiliary electrodes may vary as well: larger areas may be required at the substrate edge to direct current. Yet another approach is to provide multiple anodes, each with individual current or potential current to provide real time shaping of the current distribution. With even more complexity, the cathode itself can be divided into multiple current paths. Each approach has its own merits unique to the pattern and application, as well as level of complexity and as a result all are found in industrial practice.

1.10.1.3.2 Throwing power

Many of the early applications for patterned electrodeposition, particularly for through-hole plating of printed circuit board vias, were concerned with the ability to uniformly deposit within deep recesses. Electrolytes and conditions that could achieve this uniformity were deemed to have high throwing power. Hence the throwing power could be described as the ratio of the deposition rate at the furthermost point to the closest or top surface. In these cases, the conduction layer coating the walls of the vias provide a significant resistance. Organic additives that adsorb to the metal surface and increase the polarization increase the throwing power by reducing the impact of the electrolyte resistance, according to the relationships described above for the Wa. Other uniformity issues relevant to MEMS processing, including through thickness uniformity and variations associated with feature size, are covered in a following section.

1.10.1.3.3 Bubbles

With patterned substrates, particularly with thick resists (10–100 μm) and high aspect ratio structures, surface tension and wetting effects predominate. Generally, lithographically patterned substrates are dried in between processing steps and require wetting of the features to provide electrolyte to the conducting metal surface. Bubbles that become trapped within resist features during this initial wetting result in voids in the final metal, as the growing electroplating surface will envelope the bubble. Successful wetting is highly dependent on the geometry and aspect ratio of the features, in addition to the state of the metal seed layer and the hydrophobicity of the resist. If typical approaches such as rinsing or agitation are unsuccessful, extreme measures such as

evacuating the resist and immersing under vacuum can work for most cases. Bubbles can be formed as a result of incomplete wetting of the initially dry surface, or can be generated during plating as a result of side reactions like hydrogen reduction. Nickel electrodeposition, where the CE is usually less than 100%, can generate sufficient hydrogen to form bubbles. For this reason, the nickel electrolytes are either operated at elevated temperatures or contain a surfactant (sodium lauryl sulfate) to reduce the surface tension. Depending on the nature of the substrate, the initial stages of electroplating can be particularly prone to hydrogen generation. Initiating metal deposition onto the substrate may require an additional overpotential due to the fact that it is a different metal surface or that it possesses a thin oxide layer; the hydrogen overpotential itself may be reduced for the substrate metal, such as the case for nucleating nickel on copper.

1.10.1.3.4 Pulse plating

Certain situations in electroplating benefit from intermittently turning off the current (pulse plating) or intermittently reversing the current (pulse reverse plating) (Puippe and Leaman 1986). The net effect is to reduce the average current density. This allows temporarily depositing metal at high current densities that would become mass transport limited under DC conditions. The zero current portion of the pulse allows recovery of the metal concentration by diffusion. The pulsing frequency can vary from a few hundred megahertz (microsecond on/off times) to a several hertz (on/off times on the order of seconds). In general, the upper limit to the frequency is set by the capacitance of the electrical double layer. Pulse timing should be longer than the time required to charge or discharge this intrinsic electrode capacitance. As an estimate, for a typical double layer capacitance of $50\,\mu F\,cm^{-2}$, the charging time (in microseconds) is $17\,J^{-1}\,(A\,cm^{-2})$, and the discharge time is $120\,J^{-1}\,(A\,cm^{-2})$ (Puippe and Leaman 1986). The principal benefit most often associated with pulse plating is grain refinement. Electrodeposition is a crystallization process involving the dynamic competition between the nucleation of new crystals and the growth of existing ones. Higher overpotentials favor nucleation over growth. However, at sufficiently high overpotentials, or current densities, the metal ion concentration can be depleted in the vicinity of the cathode. Copper with fine grains and remarkably high strength has been made in this way with extreme values of the pulsing parameters (Lu *et al.* 2004). For other applications, pulsing is simply a means to overcome diffusion limitations, such as the filling of high aspect ratio alumina pores to create nanowires (Nielsch *et al.* 2000).

1.10.1.3.5 Multilevel microfabrication with electrodeposition

Electrodeposition can also be used to fabricate multiple level structures by combining a sacrificial metal with an electroplated pattern. **Figure 2** illustrates the concept, first demonstrated by Guckel and coworkers (Guckel 1993, Massoud-Ansari *et al.* 1996). A first

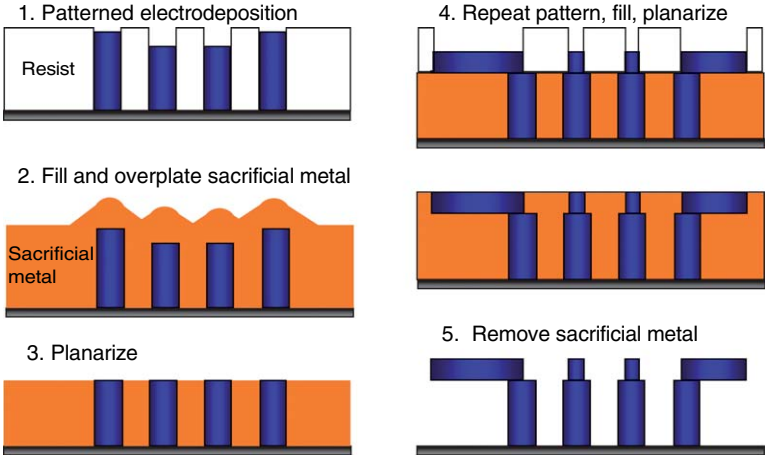


Figure 2 Multilevel structure fabrication by electrodeposition. Step 1: lithographically pattern resist on conductive substrate and electroplate structural metal. Step 2: remove resist and electroplate sacrificial metal such as copper to fill and cover structural elements. Step 3: planarize to desired thickness of the structural metal. Step 4: repeat steps 1–3, aligning to the previous level. Step 5: selectively etch the sacrificial metal to provide multilayer elements attached to the substrate.

metal, usually nickel, is electroplated through resist and the resist removed. A second metal covers the patterned metal deposit, by electrodeposition. The second metal such as silver or copper is one that can be etched selectively relative to the first. The sacrificial metal performs two functions. The first is to provide rigid support to the structures during planarization. The second function is to provide a conductive base to initiate electrodeposition of the subsequent level. This eliminates the restriction of plating on the structure exclusively and enables the creation more complicated by nearly 3D components.

1.10.1.4 Patterned Electrodeposition

The science and process sensitivities of electrodeposition through thin resist used to produce metallic conducting interconnects, packaging and contacts for the microelectronics industry have been much discussed (Dukovic 1994). In contrast, electrodeposition through thick resist for deep, high aspect ratio LIGA features has received relatively little attention (although we note that one study involving design rules for the LIGA electrodeposition step has recently become available (Drese 2004). In surveying the literature where electrodeposition through patterned photoresist is discussed, it is evident that the microelectronics industry has devoted a significant amount of experimental and modeling resources to understanding through-mask electrodeposition (more detailed reviews of these developments are available (Datta and Landolt 2000, Dukovic 1994). A central theme of the prior literature is the classification of three length scales over which electrochemical processes occur; the first, the workpiece scale, involves the distribution of current over the scale of the substrate or wafer; the second, pattern scale, involves the distribution of current between features defined in the insulating photoresist; the third, the feature scale, involves the distribution of current to and within an individual opening or feature in the resist. These three scales are depicted in **Figure 3** after DeBecker and West (West 1996). As will be discussed later, the principles governing the deposition of a metal structure or film having uniform thickness and properties over all these scales (usually desired for all electrodeposition processes) are largely the same, regardless of the final application. In terms of attaining uniformity across a substrate, the prior literature regarding the workpiece and pattern scales for electrodeposition

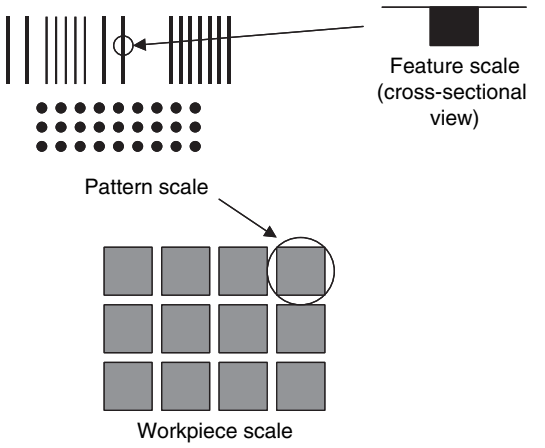


Figure 3 Schematic illustration of the workpiece, pattern, and feature scales, redrawn from a similar graphic from DeBecker and West. Groups of features are distributed across the substrate; these groups consist of a patterned array of features. The feature scale involves a single feature and its internal uniformity.

through thin resist is directly applicable to electrodeposition through thick resist.

The important difference regarding metal deposition through thin and thick resist comes when through-thickness uniformity is considered; as the electrodeposited metal films for microelectronics are usually only a few to several microns in thickness, appreciable through thickness variations in film structure do not usually occur as the thickness is so small. Moreover, any small microstructural deviation that may be present through the film thickness does not usually critically impact the function of the electrodeposited metal film in microelectronics. But in contrast to the thin films employed in microelectronics, large variations in deposit microstructure may occur over the hundreds of microns that comprise the through-mask electrodeposited thick films in LIGA. Because most of the issues in attaining uniformity across the substrate are similar to those in microelectronics, we will rather focus on a discussion on attaining through-thickness uniformity in structures electrodeposited through thick resist, a challenging aspect unique to LIGA. Besides differences in resist and metal thicknesses, another major difference between films electrodeposited for microelectronics and microsystem applications lies in the choice of materials. In microelectronics, the metal films are serving as conductors or contacts; hence, materials such as copper, gold, and lead–tin alloys (more recently lead-free alternatives are under development) have attractive physical properties

and are widely employed. In microsystems, the thick metal structures are typically serving as components in some mechanical device or as a replication mold; the mechanical properties are thus of most interest in these situations. Electrodeposited nickel has a long history of use as a structural material as its mechanical properties may be tailored by the use of alloying elements and electrolyte bath additives (both organic and inorganic) (Marti and Lanza 1969). Depending on deposition conditions and chemistries, Ni may be electrodeposited with low stress, making thick films possible. The generally good corrosion resistance of electrodeposited Ni also makes it attractive for mechanical applications in a wide variety of environments. The material properties of Ni and Ni alloys will be discussed in more detail later in the chapter. In comparing electrodeposition processes for thin and thick resists, we will keep in mind that the materials that are deposited into these lithographic molds may have very different deposition characteristics as well.

1.10.1.4.1 Workpiece and pattern scale effects

In prior treatments of through-mask electrodeposition, the first length scale often identified is the workpiece scale; this is usually the characteristic length of the substrate (or wafer) itself onto which resist defining the desired structures has been applied. Variations in the local current density typically occur over the workpiece scale. The cell geometry, the size of the substrate, and other deposition parameters dictate the uniformity across the substrate; the degree of this uniformity may be described by the Wagner number (*Wa*), a dimensionless parameter that has been discussed in more detail previously. Generally speaking, if no measures are taken to improve the current distribution across the wafer, regions close to the edge will have a locally higher current density than areas near the center of the substrate.

A few different approaches may be used to render the current distribution across the wafer more uniform. The appropriate placement of insulating shields with respect to the wafer may result in a more uniform current distribution. The use of auxiliary electrodes (also known as thief electrodes) in minimizing variations across the workpiece has been discussed previously (Mehdizadeh *et al.* 1990). For example, another cathode placed around the workpiece (as a type of surrounding frame) may improve the workpiece current distribution significantly. In industrial plating tools, all these aspects are carefully considered in optimizing the wafer scale uniformity. Obviously, as a large number of wafers are typically run under fixed conditions in the microelectronics industry, a fair amount of time and effort is spent in finding the optimal cell geometry and deposition conditions. These approaches for homogenizing the wafer scale uniformity are directly applicable to electrodeposition through thick resist in LIGA, and have in some cases been adopted.

The second length scale is referred to as the pattern scale. The relative amounts of exposed active metal and insulating photoresist may vary across the substrate due to the geometry and layout of the pattern defining the structures to be electrodeposited. Thus, the active area density varies depending on essentially the distribution of the insulating resist material. This leads to variations in the local current density (the deposition rate) depending on the local pattern geometry in the resist. Mehdizadeh *et al.* (1992) studied this problem for electrodeposition through photoresist having various pattern densities. West and coworkers considered the effect of patterning on multiple electrodes of disks and lines. These and other studies may be drawn upon in designing layouts where pattern scale effects are minimized. In our experience, LIGA pattern layouts that consist of an array of uniformly spaced features having similar dimensions are not commonly encountered; instead, patterns tend to consist of fine features (dense resist areas) located in close proximity to large, open areas of exposed seed metal (defining a wide, low aspect ratio feature). Thus, pattern scale effects must be managed on a fairly routine basis. Although it may not be possible to alter the particular part geometry and size (necessary for the device function), some discretion may be exercised in distributing the features in the resist defining the part geometry so as to mitigate pattern scale effects. Moreover, sacrificial features (openings in the resist) may be added to the pattern to break up areas that have long runs of uninterrupted insulating resist, resulting in a more uniform active area density distribution. This practice of inserting sacrificial moats or frames around features where dimensional fidelity is critical has been studied with regard to minimizing dimensional errors due to polymethylmethacrylate (PMMA) swelling and thermal expansion after immersion in the electroplating bath (Griffiths *et al.* 2004). It is somewhat fortuitous the same measures that aid in minimizing these dimensional errors in LIGA also may help in rendering the pattern scale current

distribution more uniform. Mehdizadeh *et al.* have discussed the possibility of carefully considering the wafer and pattern scale current distribution and arranging features in the resist across the workpiece so as to attain a more balanced current distribution. Even when detailed current distribution models are unavailable, past experience can be a useful guide in judiciously arranging the part layout to avoid uneven distributions of current at the pattern scale.

1.10.1.4.2 Feature scale effects

As mentioned previously, the similarity between microelectronics processing and LIGA through-mask electrodeposition diminishes as one moves from the workpiece and pattern to the feature scale. The principal reason for this is the large resist thicknesses typically employed for X-ray lithography. As pointed out by Dukovic, at the feature scale, the dimensions of the lithographic features are usually small enough (microns) that electrical field effects become less important, and the influence of the concentration field of the reacting species grows in determining the feature scale current distribution. Because mixing is limited inside the feature due to the presence of the surrounding resist material, the concentration of a reacting species may vary with position within the feature. Another aspect that must be considered for both thin and thick resists is that as deposition proceeds, the feature geometry changes as the deposited metal fills the feature. Generally speaking, obtaining good mixing inside lithographic features becomes more difficult with increasing resist thickness. When the electrodeposition reaction involves just a single metal cation, variations in its concentration inside the feature typically result in variations in the local rate of deposition and therefore deposit thickness (Mehdizadeh *et al.* 1993). The deposition characteristics of copper, gold, and tin alloys through relatively thin resists (less than 20 μm) for bump structures have been considered with attention to the thickness distribution of the deposit within a single feature (Hayashi *et al.* 2001, Kim and Ritzdorf 2003, Kondo *et al.* 1996, Watanabe *et al.* 1999). It is desirable that these bump structures have good as-plated uniformity for subsequent packaging processing. Another example of an important feature scale electrodeposition problem is the case of copper Damascene electrodeposition for copper interconnects. For this process, proprietary electrolyte bath additives and active feature sidewalls lead to more complicated feature filling behavior, despite the fact that only a single elemental species is being deposited. In the case of electrodeposition through thick resist for LIGA, the importance of the feature scale is not its influence on the local deposit thickness uniformity (the wafer is planarized after plating, removing thickness variations across the wafer at all scales). Rather, it is the more serious issue of poor mixing in the mold features themselves due to their small widths and large heights. This has broad implications with respect to the potential for variations in the concentrations of electrolyte species that may, in turn, impact the local deposit composition, morphology, and properties. For example, if additives or other metal species are present as alloying elements, profound changes may occur through the thickness owing to the constantly changing feature geometry that accompanies the metal growth. If the concentration of one of these species is close to zero at the feature bottom (e.g., due to its rapid incorporation into the deposit), large changes in the flux of this species will occur as the deposit grows and the diffusion length from the bulk electrolyte shortens. The resulting nonuniform deposit morphology or alloy composition may be unacceptable for some applications. The electrodeposition of the NiCo and NiMn systems and the resulting feature scale uniformity will be contrasted to demonstrate the importance of transport into LIGA features later in the chapter.

Interesting experiments involving the feature scale were carried out by Leyendecker *et al.* (1994). These authors studied the effect of thick photoresist on mass transport and on the pH within a LIGA feature using special electrodes. In the first series of experiments, a rotating disk electrode masked by an insulating patterned polymer layer of various thicknesses and feature sizes is employed to study a model redox reaction occurring under transport control. The aqueous electrolyte consists of 2 mM potassium ferro- and ferricyanide each in 1 M potassium chloride at 25°C. Results of their experiments are reproduced in **Figure 4**; although various geometries were investigated, they present results for a constant hole diameter of 200 μm and resist heights of 100, 200, and 500 μm (in addition to an unmasked electrode).

Another series of experiments relevant to the feature scale in LIGA involving transport limitations is reported (Griffiths *et al.* 1998). They used a commercial Ni sulfamate electrolyte to deposit Ni into features drilled into PMMA. A single piece of Cu foil attached to one side of the PMMA acted as the substrate, from which the deposited Ni structures

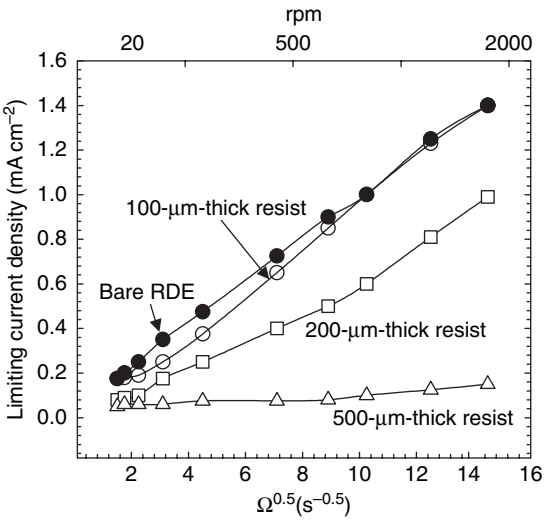


Figure 4 Limiting current densities for ferri–ferro cyanide redox couple on bare and patterned rotating disk electrodes adapted from Leyendecker *et al.* (1994). Feature hole diameter is 200 μm. The presence of the resist has a large effect on the transport of species to the feature bottom, even for an aspect ratio of one. Solid points are the bare rotating disk electrode, while the hollow points are electrodes patterned with resists of various thicknesses as indicated. (Source: Reprinted from *Electrochim. Acta*, **39**, Leyendecker K *et al*., New microelectrodes for the investigation of the electroforming of LIGA microstructures, 1139, Copyright (1994), with permission from Elsevier.)

could be removed for weighing in the determination of the CE. Current densities ranged from 1.1 to 108 mA cm^{-2} The drilled holes had diameters of 1.7, 3.2, and 6.4 mm and depths from 17 to 42 mm. They calculated the Sherwood number (the ratio between the measured current to the diffusion-limited current) for each case. The experimental results are surprising in that the Sherwood numbers are all between 10 and 100, indicating a much larger deposition current than expected based upon the diffusion-limited case.

The authors explain the unexpectedly high currents as resulting from enhanced transport of the reacting metal ion due to natural convection in the feature driven by its depletion at the substrate. This hypothesis is further supported by the fact that when the experimental cell was inverted, the measured currents were an order of magnitude less than with the deposition surface facing up. These authors go on to demonstrate by numerical modeling that flow across the top of the features cannot support the transport of the reacting metal ions at the observed rates, and that instead buoyancy-driven flows are probably important in allowing such high deposition rates. Nilson and Griffiths (2003) study natural convection in small features with respect to LIGA electrodeposition. They find that the relative importance of natural convection depends on the feature geometry and the inclination of the substrate. The authors conclude that some mixing from natural convection is expected to occur in features having aspect ratios less than 1.2, 6.6, and 37 for resist thicknesses of 0.1, 1.0, and 10 mm, respectively. Thus, natural convection becomes important for a wider range of feature aspect ratios as the resist thickness increases. Because resist thicknesses in microelectronics are usually only tens of microns or less, the importance of convective flow had not been widely discussed until electrodeposition into deep LIGA features was considered. It is interesting to contrast the findings of Leyendecker *et al.*, discussed at the beginning of this section, with those of Griffiths *et al.* In the former case, the insulating PMMA mask was shown to have strongly diminished mixing, resulting in lower transport rates of the reacting ions, while in the latter case, buoyancy-driven transport rates into deep high aspect ratio features higher than those expected from pure diffusion control were observed. Because Leyendecker *et al.* used a redox couple to measure transport rates through microstructured electrodes, they were not able to benefit from potential buoyancy-driven flows induced by the consumption of metal ions at the feature bottom.

Another series of studies focusing on feature scale uniformity in LIGA was carried out by Schwartz and coworkers (Leith and Schwartz 1999a). Leith *et al* considered the electrodeposition of NiFe at 23°C from sulfamate–chloride chemistries having various ratios of Ni to Fe electrolyte loadings. They pointed out that the plating characteristics of an electrolyte must be carefully considered before employing it for through-mask electrodeposition using thick resist. They found that electrolytes having less dissolved iron (e.g., 20:1 Ni^{2+}:Fe^{2+} as opposed to 10:1 and 5:1 also considered in their study) are less sensitive to mixing variations at relatively high current densities, suggesting that they would be more suitable for the electrodeposition of LIGA structures. Leith and Schwartz (1999b) also describe a cell designed to maximize workpiece and pattern scale uniformity, using the NiFe system (as it is a challenging system for uniformity) to demonstrate good uniformity across the wafer. The through thickness uniformity was not investigated in this study. In another article, Wang *et al.* (2002) used an electrolyte with a

10:l Ni^{2+}:Fe^{2+} loading (where the deposited NiFe alloy composition is more sensitive to mixing) to study convective–diffusion conditions during the through-mask deposition of LIGA structures. They term the approach process archaeology as the plated structure was planarized to half its deposited thickness for compositional analysis via energy-dispersive X-ray analysis. By observing in plan view the planarized structures, regions in the feature where locally high degrees of mixing occur could be easily identified as the Fe content as the alloy increases with mixing. In this case, the authors were using the transport-sensitive deposition kinetics of the Fe as a type of tracer tool to study the mixing within small cavities such as LIGA resist features. Another group of authors also considered the NiFe system to focus on the feature scale uniformity of through-mask deposited LIGA structures. Thommes *et al.* used a sulfate-based NiFe electrolyte at 50°C with a Ni^{2+}:Fe^{2+} ratio of about 12:1 to deposit NiFe structures at 5 mA cm^{-2} with good through-thickness uniformity. **Figure 5**, reproduced from their data, shows the compositional behavior of the deposition system using a rotating disk electrode. It is clear from **Figure 5** that for low current densities the electrode rotation rate does not influence the deposited alloy composition, suggesting that under these conditions the alloy composition is not greatly influenced by the amount of mixing. This relative insensitivity to mixing would appear promising for the feature scale uniformity of LIGA NiFe structures.

The suitability of the authors' deposition conditions for the electrodeposition of uniform high aspect ratio NiFe microstructures is evidenced by the through-thickness compositional data shown in **Figure 6**. For a resist thickness of 180 μm and a feature linewidth of 8 μm (aspect ratio 22.5), the authors obtain very good NiFe alloy uniformity throughout the entire structure thickness using a deposition current density of 5 mA cm^{-2}. Through-thickness uniformity suffers at higher deposition current densities. **Figure 6** also shows experimental results for deposition at 14 mA cm^{-2} using the same feature geometry. The Fe content increases through the deposit as the top PMMA surface is approached,

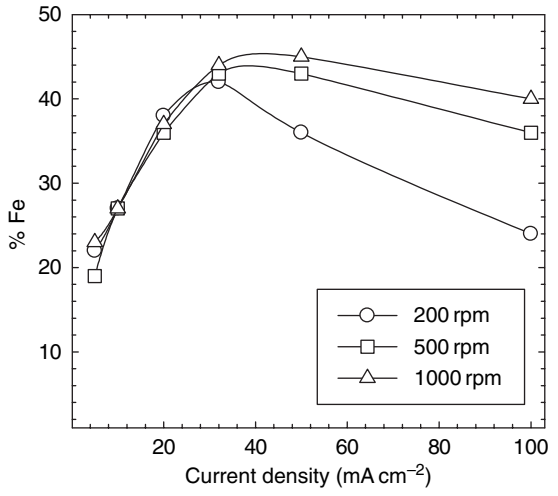


Figure 5 Electrodeposited NiFE alloy composition dependence on deposition conditions, for conditions given by and replotted from Thommes *et al.* (1994). The relative compositional insensitivity at low current densities suggests that the system may be a appropriate for the electrodeposition of uniform, high aspect ratio structures. (Source: Thommes A, Stark W, Leyendecker K, Bacher W, Liebscher H, Ilmenau C 1994 LIGA microstructures from a NiFe alloy: Preparation by electroforming and their magnetic properties. In: Romankiw L T, Herman D A Jr. (eds.) *Proc. 3rd Int. Symp. Magnetic Materials, Processes, and Devices PV 94–6*; reproduced with permission from the Electrochemical Society, Inc.)

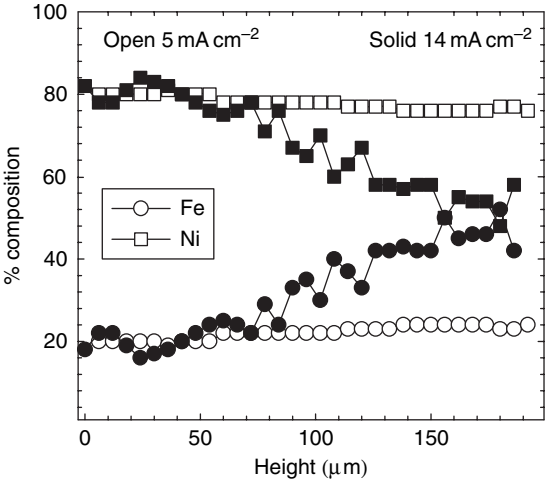


Figure 6 Electrodeposited NiFE alloy composition through-thickness uniformity replotted from Thommes *et al.* (1994). The structure width is 8 μm and its height is 180 μm. Open points are for deposition at 5 mA cm^{-2}, while the solid points are at 14 mA cm^{-2}. Deposition at the lower current density (where composition was shown to be less sensitive to mixing) leads to uniform alloy content through the structure thickness, as compared to 14 mA cm^{-2}. (Source: Thommes A, Stark W, Leyendecker K, Bacher W, Liebscher H, Ilmenau C 1994 LIGA microstructures from a NiFe alloy: Preparation by electroforming and their magnetic properties. In: Romankiw L T, Herman D A Jr. (eds.) *Proc. 3rd Int. Symp. Magnetic Materials, Processes, and Devices PV 94–6*; reproduced with permission from the Electrochemical Society, Inc.)

varying from about 20–50% across the thickness. Though not shown here, the authors filled cylindrical features with a diameter of 5 μm and a height of 90 μm depositing NiFe at 5 mA cm^{-2}. They show similarly good results as for the structure in **Figure 6** deposited at the same current density. The electrodeposition process outlined by Thommes *et al.* (1994) has been shown to be capable of producing high aspect ratio structures with good alloy through-thickness compositional uniformity. To the authors' knowledge, few other studies exist of alloy deposition through thick resist where the feature scale uniformity is considered in such detail.

An aspect that is likely important in explaining the good results obtained by these authors is the bath temperature; Brenner has discussed the fact that NiFe electrodeposition becomes less anomalous (and hence less transport sensitive) with increasing bath temperature. Andricacos and Romankiw (1993) also discuss the importance of bath temperature as well as other process variables on electrodeposited magnetic alloy composition. Given the limitations in obtaining good mixing in small LIGA features, it is not surprising that achieving good feature scale uniformity for anomalous deposition systems (such as NiFe) is challenging for geometries encountered in LIGA electrodeposition. New techniques are being employed for studying the electrodeposition reaction inside of small features like those fabricated by the LIGA process; for example, special microelectrodes and chronoamperometric methods have been developed to measure the local metal cation concentrations during alloy electrodeposition inside small cavities (Kuepper and Schultze 1997). In the study mentioned previously, concentration variations inside of a cavity where a NiCu alloy is being electrodeposited are investigated; they also observe the formation of a Ni hydroxide and nickel oxide at the deposition surface during electrodeposition of the NiCu alloy from a sulfate electrolyte with potassium nitrate as supporting electrolyte.

1.10.1.4.3 Summary: Uniformity at the workpiece, pattern, and feature scales in LIGA

The detailed study of Thommes *et al.* is a good example of the considerable effort required to develop an electrodeposition process capable of producing uniform structures at the feature scale for geometries encountered in LIGA electrodeposition. When mixing effects are important (as often the case in Ni alloy deposition), this is especially difficult. After some confidence in the feature scale uniformity is attained for a given chemistry and set of deposition conditions, the uniformity at the pattern and workpiece scales must still be addressed as well. This makes LIGA through-mask electrodeposition in some ways more challenging than electrodeposition through thin resist, as the feature scale adds literally another dimension to the problem of obtaining uniformity across multiple length scales. Considerable time and effort may be saved by first characterizing the electrodeposition kinetics and the importance of mass transport for an electrodeposition process before attempting to apply it to LIGA fabrication (especially in the case of an alloy system). If a good understanding of kinetics, transport, and alloy composition is in hand, electrolytes and deposition parameters suitable for deposition through thick resist may be chosen more prudently.

1.10.2 Properties of Electrodeposits

1.10.2.1 Internal Stress

For many electrodeposited metals, as-deposited internal stresses can be severe enough to deform the substrate or cause catastrophic detachment of the metal film. The severity of the effect depends on the metal, the electrodeposition process, the thickness, and the geometry of the plated structure. This stress is not to be confused with the stresses that originate from the substrate thermal expansion mismatch during subsequent elevated temperature processing (Keller *et al.* 1999). If the stress is tensile, the metal film will peel upward at corners and edges where discontinuities create a localized increase in force. If the stress is compressive, the film can buckle or blister. The degree of internal stress will limit the ultimate thickness of the plated film before it fails by either delamination from the substrate or cracks internally. For this reason, hard, stressed electrodeposits of Cr, Pt, and Rh are usually limited to a few microns. Copper and soft gold are relatively low stress as-deposited and can be plated hundreds of microns thick.

Several methods are available to measure the stress including the measurement of wafer or substrate bow by interferometry or capacitance, as well as *in situ* instruments like the spiral contractometer. In the bent strip method, the simplest technique for measuring stress, thin copper strips are coated with an insulating film on one side. As the thickness of the film builds on the copper surface, the strip deflects

according to the nature and magnitude of the internal stress. The amount of curvature depends on the properties of the deposit, the thickness, and the mechanical properties of the substrate. If the thickness of the substrate, t_s, is much thicker than the deposit, t_d, then the Stoney equation can be used to relate the radius of curvature, R, to the stress, σ:

$$\frac{Et_s^2}{6(1-v)t_d R}$$

where E and v are the Young's modulus and Poisson ratio of the substrate, respectively, and the substrate is initially flat (Stoney 1909).

Stress is a complicated function of all the plating parameters and is very much process dependent. For example, high-strength alloys of nickel with manganese iron or cobalt are highly stressed in tensile. High-strength nickel produced from organic grain refining agents is often stressed in compression. Varying the intrinsic stress through thickness has been used to intentionally induce curvature in gold electroplated beams and flexures (Chinthakindi *et al.* 2002).

1.10.2.2 Adhesion

For many multilayer processes encountered in MEMS and semiconductor processing, adhesion of the metal deposit to the substrate and pattern are critical to the success of the fabrication. Numerous forces can be applied to the metal interface that result in failure, including the internal stress of the deposit itself, stresses generated as a result of the coefficient of thermal expansion (CTE) mismatch during thermal processes, and stresses applied during the subsequent polishing of the metal such as that exerted by chemical–mechanical polishing (CMP) in the copper Damascene process or by grinding of excess deposit in the LIGA process.

Metals that readily form a native oxide, such as aluminum, or that form a chemically resistant oxide, such as nickel, prove to be difficult substrates for achieving good adhesion of electrodeposits. Standard methods are available and usually employ a series of acid treatments followed by a strike layer or a thin layer of metal generated in a specially adapted electrolyte. The Woods Nickel strike is such an example, consisting of 125 ml l^{-1} of HCl and 240 g l^{-1} of $NiCl_2$ operated at 50 mA cm^{-2} for a few minutes time (Lowenheim 1995).

While no universal method exists for determining an absolute value of adhesion, several methods are adequate for relative evaluations. Methods include the peel test, shear test, ring shear test (Dini 1993, Lowenheim 1995). The measurement of peel strength is readily adapted to thin films, but is highly dependent on strength and ductility of the metal film and requires large areas (Klingenmaier and Dobrash 1976). Other methods are purely qualitative, such as those to evaluate gold coatings: the coated components are subjected to elevated temperatures and the presence of blisters indicates poor adhesion. Instruments that measure the maximum shear force at failure provide a more quantitative measure of the adhesion. Thick-resist generated structures are plated onto a coating or substrate and provided the failure occurs at the intended interface, the adhesion can be described. The shortcoming of this approach is that the bulk strength of the substrate may be weaker than the interfacial strength, in which case only a minimum strength can be assigned.

1.10.2.2.1 Electroplating on aluminum

Aluminum is a notoriously difficult metal on which to electroplate, both for initiating metal film growth and for achieving sufficient adhesion. Despite this, successful methods have evolved and illustrate some of the principles of achieving adhesion of a electrodeposited film to its substrate. Aluminum is a common metal available to MEMS fabrication: it has low electrical resistivity, a self-passivating oxide, and it is easily patterned by wet etchants. Recently, aluminum thin films have been implemented as a substrate material in the X-ray lithographic LIGA process in order to provide better feature resolution and PMMA resist adhesion after exposure (Griffiths *et al.* 2005).

As with most electroplated metal adhesion issues, the critical factor is removal of a surface oxide layer. For aluminum, this native oxide can develop rather quickly. A established method is to use an intermediate layer of a displacement plated metal, such as zinc or tin. Employed in a caustic electrolyte to prevent oxide formation on the aluminum, the zinc or tin displaces the aluminum owing to their more noble electrochemical potential. Furthermore, optimal zinc films are obtained by a double immersion process whereby the first zinc layer is removed by a short nitric acid etch and the subsequent second zinc immersion has improved nucleation density and uniformity. A thin copper layer is often plated onto the zinc to provide a good base for subsequent metal deposition and to protect the zinc from more corrosive plating electrolytes.

Another common approach is to create a porous surface oxide by anodizing the aluminum. Recent use of the technique has focused on producing ordered arrays of directional nanopores. With more aggressive conditions such as stronger acids and higher temperatures, porosity is increased and it is this porosity that provides the physical interlocking for good adhesion. With respect to providing a base for electrodeposited metals, certain conditions for the anodization are preferred. Although sulfuric acid is common for producing hard anodic coatings, phosphoric acid anodized aluminum is optimal for electroplating due to the larger pores that are produced. Higher anodizing potentials also produce larger pores that correlate well with better adhesion. The alloy content of the aluminum also has an important role to play. It was found that aluminum with small amounts of copper and silicon (1% of each, a common alloy used for integrated circuits) greatly facilitated subsequent adhesion of electroplated copper onto the anodized surface (Griffiths *et al.* 2005).

1.10.2.3 Crystal Orientation and Microstructure

The polycrystalline nature of most electrodeposits render their properties highly sensitive to the nature of the microstructure, that is the size distribution of the individual crystals, or grains, their boundaries, and their lattice orientation. Many properties, including strength, wear resistance, corrosion resistance, etch rates, magnetic properties, are strongly dependent on the texture or crystal orientation of the electrodeposit, which in turn is dependent on all of the processing variables. For nickel electrodeposits, a fiber-oriented growth mode with a preferred orientation in the <110> direction is associated with high-strength deposits, and can be influenced by pH and current density (Amblard *et al.* 1979). The influence of the substrate is generally negligible once the deposit exceeds a thickness of a few microns (Dini 1993).

1.10.2.4 Mechanical Properties of Electrodeposits

As coatings, the mechanical properties of electrodeposits, particularly the strength and hardness, have been tailored to achieve high wear resistance. A wealth of historical mechanical properties and the respective electrodeposition processes are available (Safranek 1986). As the structural material in micromechanical elements, such as spring flexure arms, the electrodeposit needs to resist plastic or permanent deformation, and hence requires high strength. In addition, the mechanical properties pertinent to spring behavior, such as the stiffness or modulus, need to be precisely characterized. **Table 1** lists a range of strengths and modulii for various electrodeposited metals. These values are provided as an estimate: variation depending upon exact processing variables is expected. The most commonly used alloys for microsystem springs are those based on nickel, owing to the superior strength.

1.10.2.4.1 Tensile test

The traditional method for determining the strength and modulus of a metal is through a tensile test, in which a precisely shaped specimen is pulled in tension until failure. This produces the stress–strain curve for which the modulus, yield strength, ultimate tensile strength, and strain to failure can be determined. This approach can be applied to electrodeposits in one of two ways. The first is to electroplate a sheet of the desired electrodeposit to sufficient thickness, and then form the shape required through machining, stamping, or wire electrodischarge machining. The second approach is to lithographically define the tensile mold, provided

Table 1 Mechanical properties of common electrodeposited metals

	Modulus (GPa)	*Yield strength (MPa)*	*Vickers hardness*
Nickel, course grained	160	300–600	200–300
Nickel, organic additive refined	180	800–1200	400–600
Ni–1% Mn	190	600–1200	400–600
Ni–35% Co	200	600–1100	400–600
Gold	70	100	50–100
Gold – hard	70	200	150–200
Copper	11	300–400	50–200

the resist is thick enough (a few hundred micrometers is needed for a conventional tester). This latter approach captures the effects associated with feature scale transport, but suffers from long processing times and specialized lithography methods. Some specialized instruments have been developed specifically for thin-film tensile testing, but the approaches are not standard.

1.10.2.4.2 Hardness

Another gauge for the strength of an electrodeposited metal is its hardness, as determined by indenting a surface with a pyramidal diamond tip. For many metals, an excellent correlation between the yield strength and the hardness can be achieved. For a given load, the surface deforms plastically and the size of the remaining indentation is used to characterize the strength of the material. The resulting hardness number has been correlated to the flow stress, or yield strength, for numerous metals, although the exact relationship is unique for each material. A number of scales exist for the hardness, including Vickers, Knoop, Rockwell, and Brinell. The Vickers hardness is defined as

$$H = 1854\frac{P}{a^2}$$

where a is the length of the indent diagonal in microns and P is the load in grams-force. For thin films and coatings, a microhardness test is commonly used, where the applied load is small enough to reduce the indent size far below the thickness of the deposit. So for a 500 VHN_{25} nickel electrodeposit and a 25-g load, the indent diagonal is 9.6 μm, setting the minimum feature size for measurement around 40 μm. A requirement for the microhardness test is that the penetration depth is less than a third of the total depth of the material; otherwise, the substrate will influence the measurement. Furthermore, the surface being indented must be optically flat to provide an accurate measurement. The granularity of as-deposited films makes taking a reliable measurement without polishing difficult. As a result, samples must be carefully prepared to provide rigid backing, depth of material, and a smooth surface. Nevertheless, the technique is straightforward and commonly used to characterize electrodeposited metals.

1.10.2.4.3 Bulge test

Yet another technique for measuring these mechanical properties of thin films is the bulge test, in which the deflection of a membrane under pressure provides a measure of stress and strain, according to:

$$\sigma = \frac{2Pw^2}{bt}$$

where the stress is related to the pressure P, the width of the membrane w, the center deflection b, and the film thickness t (Xiang *et al.* 2002). And,

$$\varepsilon = \varepsilon_0 + \frac{b^2}{6w^2}$$

where ε is the strain and ε_0 the residual strain.

1.10.2.4.4 Origins of strength

It has been well documented that a principal source of strengthening in electrodeposits is the high concentration of grain boundaries, where the Hall–Petch relation correlates the strength to the grain size:

$$\sigma = \sigma_0 + Kd^{-1/2}$$

where σ can be the flow stress or microhardness and d is the grain diameter. Li and Ebrahimi (2003) have examined the Hall–Petch relation for a number of nanocrystalline nickel and nickel alloy electrodeposits.

1.10.2.5 Thermal Stability

For many microsystem applications, fabricated components will have an elevated temperature of operation or must retain their performance after being subjected to elevated temperatures. Unfortunately, for many electrolytic metals, the very nature of the deposit that renders it useful (fine grain size for high strength in the case of nickel, or fine grain size for bright films in the case of copper and gold) also produces a less stable microstructure. The higher density of grain boundaries and lattice defects provide a higher energy state of the metal, for which modest increases in temperature can activate modifications to the microstructure. In some cases, changes to the microstructure such as recrystallization and grain growth can cause the electrodeposit to lose its most desirable properties.

For Damascene copper interconnects however, recrystallization is desired in order to reduce the electrical resistivity and render the copper stable toward further high-temperature processing (Harper *et al.* 1999). Recrystallization manifests itself by an approximately 20% decrease in electrical

resistivity from the as-deposited state. Copper films prepared by electrodeposition are known to recrystallize easily at temperatures as low as room temperature. This self-annealing process is highly dependent on the deposition conditions (Perez-Prado and Vlassak 2002).

For nickel films, it is desired to preserve the fine-grained microstructure in order retain high-strength material. Pure nickel electrodeposits are thermally unstable, manifesting grain growth and texture evolution commencing near 300°C (Czerwinksi and Szpunar 2000). Thuvander *et al.* (2001) found the onset of grain growth for nanocrystalline electrodeposited Ni to be as low as 220°C; the onset for a Ni–Fe (50%) alloy was found to be near 400°C. For a Ni–Fe (20%) electrodeposit, Li and Ebrahimi (2003) found rapid grain growth to commence near 300°C, and the hardness falling by 50% with an annealing temperature of 400°C. Ni–W alloys have been shown to possess better thermal stability. Although the hardness was not characterized, Choi *et al.* (2003) demonstrated that a nanocrystalline electrodeposited Ni–W (20%) was microstructurally preserved up to 400°C, with good stability even at 700°C. Ni–Mn (1%) alloys have also demonstrated remarkable thermal stability, with high strengths and a fine-grained microstructure preserved with heat treatments up to 600°C (Talin *et al.* 2006).

1.10.3 Selected Metals and Applications

1.10.3.1 Nickel and Nickel Alloys for Mechanical Actuation

A number of electrodeposited nickel alloys have desirable mechanical properties with high strength and toughness and find use as structural materials in microsystem components like motors, gears, latches, nozzles, and flexure spring arms. While wrought nickel possess a yield strength around 300 MPa, the yield strength of electrodeposited nanocrystalline nickel and nickel alloys can exceed 1000 MPa. The small size of the as-deposited grains imparts strength according to the Hall–Petch relation. Nanocrystalline Ni, NiFe, NiCo, and NiMn can all have exceptionally high strengths. The deposition of alloys of NiFe (20% or 50% Fe) and NiCo (20% Co) are more susceptible to mass transport effects and can be difficult to control in high aspect ratio structures. Pure nanocrystalline nickel can have equally high strength if an organic grain-refining agent, such as saccharin, is used. While reducing the stress significantly, the use of these sulfur bearing agents results in sulfur inclusions in the metal that can lead to catastrophic embrittlement at temperatures near 300°C. Nickel alloys (NiFe, NiCo) can be as strong without the use of sulfur-bearing organics, but lose strength with temperature exposures larger than 300°C, as a result of grain growth. Electrodeposited NiMn alloys have been shown to have exceptional thermal stability, in addition to high strength.

Resist-patterned electrodeposited nickel has found use as a spring material in flexure elements for forming mechanically compliant electrical contacts. Individual spring components, millimeters thick, have been electroformed using a LIGA process (Goods *et al.* 2004). The electrical contact spring was electroformed with nickel manganese first and then coated with gold after release to provide electrical conduction and corrosion resistance.

Another recent application of resist-patterned electrodeposited nickel is in the fabrication of arrays of microminiature springs for making electrical contact to the bond pads of integrated circuits during testing. As the size of the chips decrease and the number of electrical connections increase, the requirements for the probe cards used during electrical test exceed conventional fabrication methods. MEMS fabrication is uniquely suited to provide the reduced size and pitch of electrical contact springs needed for advanced probe cards (Itoh *et al.* 2003, Kataoka *et al.* 2003, Kim *et al.* 2004). Higher strength nickel alloys afforded by electrodeposition, allow for the design of smaller, more tightly packed springs with the contact forces needed to make reliable connection. The electrical test can be at elevated temperatures for accelerated reliability prediction that demands a strong, thermally stable nickel material. Repeated and aggressive cycling of the springs requires an electrodeposit strong enough to resist deformation. **Figure 7**, taken from Kataoka *et al.*, illustrates one concept for the creation of cantilever-based nickel contact probes fabricated through lithography and electrodeposition.

In addition to microsystem components, electrodeposited nickel finds use as the tool material for hot embossing and injection molding of plastics for microfluidics, optical recording media (Czerwinksi and Szpunar 2000), replication of nanostructured surfaces (Lee and Kwon 2006), and nanoimprint lithography (Park *et al.* 2004). Electrolytes for the

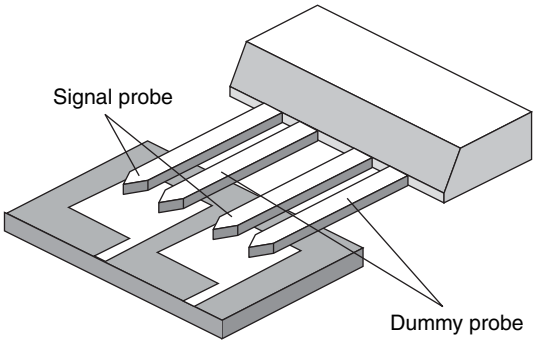


Figure 7 Example of nickel-plated microelectromechanical systems (MEMS) cantilever for use as an electrical test probe. (Source: Reprinted from *Sens. Actuators*, **103**, Kataoka K *et al*., Electroplating Ni micro-cantilevers for low contact-force IC probing, 116–21, Copyright (2003), with permission from Elsevier.)

electrodeposition of nickel are based on nickel sulfate (the Watts bath) or nickel sulfamate.

$$Ni^{2+} + 2e^- \rightarrow Ni_{(s)}$$

Some amount of chloride is present to assist dissolution of the nickel anode, but specially formulated nickel anodes possessing sulfur will sufficiently release nickel without the chloride. **Table 2** lists a common formulation for a sulfamate electrolyte, which is used for high-speed electroforming owing to the reduced stress and slightly higher concentrations of nickel available. Typically, a concentrate of nickel as nickel sulfamate (24 oz gal^{-1} or 180 g l^{-1}), available from a number of commercial vendors, is diluted to the desired concentration.

Table 3 lists a representative formulation for a Watts nickel electrolyte with saccharin as an organic brightner to reduce stress and refine the grain size. Sulfamate electrolytes can be used in conjunction with these additives as well, and the resulting deposits are lustrous, smooth, low stresses, and high strength. The downside to these addition agents, as mentioned previously, is the incorporation of sulfur into the nickel deposit.

A range of temperatures is feasible for nickel electroplating, but usually falls within 32–60°C.

Table 2 Sulfamate nickel electrolyte

Ni, as sulfamate (g l^{-1})	85
Boric acid (g l^{-1})	30
Sodium lauryl sulfate (g l^{-1})	0.2
pH	4
Temperature (°C)	50
J (mA cm^{-2})	15
Anode	Ni–S

Table 3 Watts Ni, bright

Ni, as Ni sulfate $NiSO_4 \cdot 6H_2O$ (g l^{-1})	90
Ni, as chloride, $NiCl_2 \cdot 6H_2O$ (g l^{-1})	10
Boric acid (g l^{-1})	30
Saccharin (g l^{-1})	1
Sodium lauryl sulfate (g l^{-1})	0.2
pH	4
Temperature (°C)	40
J (mA cm^{-2})	15
Anode	Ni–S

The low end of the temperature range is dictated by the boric acid, a critical component to both types of electrolytes. Boric acid, in the range 30–40 g l^{-1}, is believed to act as a buffer to mitigate the local rise in pH associated with hydrogen formation. The amount of boric acid is close to the solubility limit, and temperatures below 32°C should be employed with caution. A wetting agent or surfactant such as sodium lauryl sulfate is also desirable to inhibit pitting that can result when hydrogen clings to the growing surface. A low surface tension, which can also be accomplished with higher temperatures, is especially critical for plating in patterned resists where feature walls can pin bubbles within the feature.

While nickel sulfamate-derived electrodeposits are stronger than most other metals and easily integrated due to low stresses, some applications require a thermally stable alloy with high strength and resistance to embrittlement. The trade-off is that these alloys (Ni–Fe, Ni–Co, and Ni–Mn) are highly stressed as-deposited with more complicated processes. An extensive literature exists for the electrodeposition of magnetic Ni–Fe and Ni–Co alloys, so the processes are only briefly described here. **Table 4** provides the electrolyte used in Leith and Schwartz (1999a) as an example of Ni–Fe deposition in high aspect ratio structures. The high stresses encountered required the saccharin stress-reducing additive, rendering the deposit susceptible

Table 4 Ni–Fe (20%) permalloy

Ni, as sulfamate (g l^{-1})	11.7
Fe, as chloride, $FeCl_2 \cdot 4H_2O$ (g l^{-1})	0.45
Boric acid (g l^{-1})	25
Saccharin (g l^{-1})	1.5
Sodium lauryl sulfate (g l^{-1})	0.2
Ascorbic acid (g l^{-1})	1.0
pH	3
Temperature (°C)	25°C
J (mA cm^{-2})	60

Table 5 NiMn electrolyte

Ni, as sulfamate (g l^{-1})	85
Mn, as $MnCl_2 \cdot 4H_2O$ (g l^{-1})	1–5
Boric acid (g l^{-1})	30
Sodium lauryl sulfate (g l^{-1})	0.2
pH	3.5
Temperature (°C)	28
J (mA cm^{-2})	5[a]
Anode	Ni–S

Source: Kelly *et al.* (2003).
[a]Average, pulsed: 15 mA cm^{-2} for 0.66 s, 3 mA cm^{-2} for 4.4 s.

to embrittlement. Furthermore, the alloy composition was sensitive to the extent of mass transfer, a difficulty for Ni–Fe and Ni–Co alloys.

A high-strength nickel alloy developed to avoid these issues with mass transport and electrolyte sensitivity is Ni–Mn, where the Mn content is low, less than 1% (Kelly *et al.* 2003). Uniform alloy content through the thickness of high aspect ratio molds demonstrates the reduced sensitivity to mass transport with respect to the codepositing element. The electrolyte is based on sulfamate and pulse plating is used to alleviate some of the tensile stress. The low temperatures employed dictate caution with extended use of the electrolyte (Goods *et al.* 2006) (**Table 5**).

1.10.3.2 Copper for RF MEMS and Electrical Interconnects

Copper finds use primarily as an electrical conductor. Thin deposits on the order of a micron are used in the Damascene process for copper interconnects. For other microsystems, thick deposits have been used to construct elements for radio frequency (RF) MEMS, such as transmission lines (Wu *et al.* 2003), wave guides, filters (Park *et al.* 2000), and switches (Amaya *et al.* 2004). The thick deposits afforded by electrodeposition are particularly useful for high-power microwave applications (Choi and Yoon 2004).

The most common electrolytes are based on copper sulfate and sulfuric acid near room temperature. Bright coatings are obtained through the use of addition agents that refine the grain size and level the surface morphology. The electrolytes used in Damascene interconnects possess multiple organic additives to induce super-filling or bottom-up feature deposition, resulting in the void-free deposition in features with aspect ratios greater than one. These and other organic additives may not be suitable for the thick deposits desired in MEMS applications, owing to the unstable microstructure that results. **Table 6** lists a typical copper electrolyte formulation, where pulse plating can be used to refine grains and reduce stress without the use of additives.

Table 6 Copper electrolyte, sulfuric acid

Cu, as sulfate $CuSO_4 \cdot 5H_2O$ (g l^{-1})	40
Sulfuric acid (g l^{-1})	180
Chloride (g l^{-1})	50
Temperature (°C)	25
J (mA cm^{-2})	10
Anode	Cu or Cu–P

1.10.3.3 Gold

Gold finds use as both a corrosion-resistant coating and an electrical conductor. Through-resist plating of gold has been extensively used in microelectronics, particularly for packaging. Lithographically patterned gold also finds use in the fabrication of X-ray masks for the LIGA process.

Early electrolytes were based on cyanide complexes (Green *et al.* 2003, Morrissey 1993) of the gold, which were generally incompatible with photoresists. Formulations common to microelectronics and MEMS consist of sulfite-based chemistries (Horkans and Romankiw 1977). **Table 7** gives a representative formulation based on Green *et al.*

1.10.3.4 Thermoelectric Thin-Film Microdevices: Bi_2Te_3

In addition to materials with enhanced mechanical properties, semiconductors like CdTe for photovoltaics and Bi_2Te_3 alloys for thermoelectric energy conversion or Peltier cooling are readily electrodeposited and incorporated into microsystem fabrication schemes (Snyder *et al.* 2003). A number of MEMS-based thermoelectric device designs and fabrication schemes have emerged recently (Bottner 2005, Bottner *et al.* 2004,

Table 7 Gold, sulfite

Au, as $Na_3Au(SO3)_2$ (g l^{-1})	10
K_2SO_3 (g l^{-1})	38
pH	9.5
Temperature (°C)	45
J (mA cm^{-2})	3.5

Source: Green *et al.* (2003).

Da Silva and Kaviany 2005). These devices are targeted at various cooling applications, such as the site-specific cooling of hot spots on microprocessors or optoelectronic components, as well as energy conversion applications such as remote power sources for sensors. The efficiency for which a thermoelectric material can convert heat into electricity or pump heat from a source is related to the materials intrinsic material properties by a figure of merit, z. This figure of merit is a strong function of temperature and as a thermoelectric material, bismuth telluride and its doped alloys have some of the highest figures of merit near room temperature. Electrodeposition is an attractive route for synthesizing the material as the technique affords thicker films for taller elements. Taller thermoelectric elements are crucial to increasing the thermal resistance in order to increase the temperature gradient and improve conversion efficiencies. A number of studies have examined the process–property relationships for this material (Magri *et al.* 1996, Martin-Gonzalez *et al.* 2002, Tittes *et al.* 2003). Moreover, electrodeposition has been shown to create the preferred crystal orientation with the highest figure of merit. The properties of Bi_2Te_3 are highly anisotropic, with a figure of merit for energy conversion nearly two times larger in along the preferred crystal axis. However, much like the high-strength nickel alloys Ni–Fe and Ni–Co, the electrical performance is highly sensitive to the alloy composition and process control is tantamount. By varying deposition conditions, investigators were able to achieve both n- and p-type semiconductor material, depending on the relative excess of one of the elements (Takahashi *et al.* 1994). Exacerbating the issue is the fact that the elements are in low concentrations (of the order of 10 mM) due to their low solubility, even in the 1 M nitric acid electrolyte commonly used. The overall reaction in the acid electrolyte is:

$$3HTeO_2^+ + 2Bi_3^+ + 18e^- + 9H \rightarrow Bi_2Te_3(s) + 6H_2O$$

Electrodeposition is also the key to the synthesis of many of the nanostructured thermoelectric materials. By using the template synthesis method in nanoporous alumina, nanowires of thermoelectric materials are electrodeposited and the properties improved over bulk materials (Borca-Tasciuc *et al.* 2004). A clever integration of such a nanostructured thermoelectric material into a MEMS device has been demonstrated (Wang *et al.* 2005). The promise for improved materials and novel applications promise to make this an important contribution to MEMS by electrodeposition.

References

Amaya R E, Levenets V, Tarr N G, Plett C 2004 Copper coplanar waveguides in Si substrates for 10 GHz applications. *J. Vac. Sci. Technol.* A **22**, 847–50

Amblard J, Epelboin I, Froment M, Maurin G 1979 Inhibition and nickel electrocrystallization. *J. Appl. Electrochem.* **9**, 233–42

Andricacos P C, Romankiw L T 1993 In: Tobias H G A C W (ed.) Magnetically soft materials in data storage: their properties and electrochemistry. *Advances in Electrochemical Science and Engineering.* VCH Verlagsgesellschaft, Weinheim

Atanassov N, Schils H W 1999 Influence of structural parameters on the properties of electrolytic Ni–Mn–S deposits. *J. Appl. Electrochem.* **29**, 51–7

Bard A J, Faulkner L 2001 *Electrochemical Methods and Fundamentals and Applications*. John Wiley and Sons, New York

Borca-Tasciuc D A, Chen G, Prieto A, Martin-Gonzalez M S, Stacy A, Sands T, Ryan M A, Fleurial J P 2004 Thermal properties of electrodeposited bismuth telluride nanowires embedded in amorphous alumina. *Appl. Phys. Lett.* **85**, 6001–03

Bottner H Micropelt miniaturized thermoelectric devices: Small size, high cooling power densities, short response time. In: IEEE (ed.) *24th Int. Conf. Thermoelectric (ICT)*, Clemson, SC, USA, June 19–23, 2005. IEEE, Clemson, SC

Bottner H, Nurnus J, Gavrikov A, Kuhner G, Jagle M, Kunzel C, Eberhard D, Plescher G, Schubert A, Schlereth K H 2004 New thermoelectric components using microsystem technologies. *J. Microelectromech. Syst.* **13**, 414–20

Brenner A 1963 *Electrodeposition of Alloys*. Academic Press, New York

Chinthakindi A K, Bhusari D, Dusch B P, Musolf J, Willemsen B A, Prophet E, Roberson M, Kohl P A 2002 Electrostatic actuators with intrinsic stress gradient. I. Materials and structures. *J. Electrochem. Soc.* **149**, H139–45

Choi Y-S, Yoon J-B 2004 Experimental analysis of the effect of metal thickness on the quality factor in integrated spiral inductors for RF ICs. *IEEE Electron Devices Lett.* **25**, 76–9

Choi P, Al-Kassab T, Gartner F, Kreye H, Kirchheim R 2003 Thermal stability of nanocrystalline nickel-18 at. % tungsten alloy investigated with the tomographic atom probe. *Mater. Sci. Eng.* **A353**, 74–9

Czerwinksi F, Szpunar J A 2000 Texture instability in Ni electrodeposits applied in optical disk technology. *J. Mater. Sci.* **35**, 331–5

Da Silva L W, Kaviany M 2005 Fabrication and measured performance of a first-generation microthermoelectric cooler. *J. Microelectromech. Syst.* **14**, 1110–17

Datta M, Landolt D 2000 Fundamental aspects and applications of electrochemical microfabrication. *Electrochim. Acta* **45**, 2535–58

Dini J W 1993 *Electrodeposition, the Materials Science of Coatings and Substrates*. Noyes Publication, Norwich, NY

Drese K S 2004 Design rules for electroforming in the LIGA process. *J. Electrochem. Soc.* **151**, D39–D45

Dukovic J O 1994 Current distribution and shape change in electrodeposition of thin films for microelectronic fabrication. In: Gerischer H and Tobias C W (eds.) *Advances in Electrochemical Science and Engineering.* VCH, Weinheim

Goods S H, Kelly J J, Talin A A, Michael J R, Watson R M 2006 Electrodeposition of Ni from low-temperature sulfamate electrolytes. *J. Electrochem. Soc.* **153**, C318–C324

Goods S H, Kelly J J, Yang N Y C 2004 Electrodeposited nickel–manganese: An alloy for microsystem applications. *Microsyst. Technol.* **10**, 498–505

Green T A, Liew M-J, Roy S 2003 Electrodeposition of gold from a thiosulfate–sulfite bath for microelectronic applications. *J. Electrochem. Soc.* **150**, C104–10

Griffiths S K, Nilson R H, Bradshaw R W, Ting A, Bonivert W D, Hachman J T, Hruby J M 1998 Transport limitations in electrodeposition for LIGA microdevice fabrication. *Proc. SPIE* **3511**, 364–375

Griffiths S K, Crowell J A W, Kistler B L, Dryden A S 2004 Dimensional errors in LIGA-produced metal structures due to thermal expansion and swelling of PMMA. *J. Micromech. Microeng.* **14**, 1548

Griffiths S K, Losey M W, Hachman J T, Skala D M, Hunter L L, Yang N Y C, Boehme D R, Korellis J S, Aigeldinger G, Lu W Y, Kelly L J, Hekmaty M A, Mclean D E, Yang P C Y, Hauck C A, Friedmann T A 2005 Resist substrate studies for LIGA microfabrication with application to a new anodized aluminum substrate. *J. Micromech. Microeng.* **15**, 1700–12

Guckel H 1993 *Formation of Microstructures by Multiple Level Deep X-ray Lithography with Sacrificial Metal Layers*. Wisconsin Alumni Research Foundation, Madison, WI, USA

Harper J M, Cabral C Jr., Andricacos P C, Gignac L, Noyan I C, Rodbell K P, Hu C K 1999 Mechanisms for microstructure evolution in electroplated copper thin films near room temperature. *J. Appl. Phys.* **86**, 2516–25

Hayashi K, Fukui K, Tanaka Z, Kondo K 2001 Shape evolution of electrodeposited bumps into deep cavities. *J. Electrochem. Soc.* **148**, C145–C148

Horkans J, Romankiw L T 1977 Pulsed potentiostatic deposition of gold from solutions of the Au(I) sulfite complex. *J. Electrochem. Soc.* **124**, 1499–505

Itoh T, Kawamura S, Kataoka K, Suga T 2003 Contact properties of micromachined Ni probes. *49th IEEE Holm Conference on Electrical Contacts*, *Washington DC, USA*.

Jordan M 2000 Electrodeposition of tin–lead alloys. In Paunovic M (ed.) *Modern Electrodeposition*, 4th edn. John Wiley and Sons, New York

Kataoka K, Kawamura S, Itoh T, Ishikawa K, Honma H, Suga T 2003 Electroplating Ni micro-cantilevers for low contact-force IC probing. *Sens. Actuators A* **103**, 116–21

Keller R M, Baker S P, Arzt E 1999 Stress–temperature behavior of unpassivated thin copper films. *Acta Mater.* **47**, 415–26

Kelly J J, Goods S H, Yang N Y C 2003 High performance nanostructured Ni–Mn alloy for microsystem applications. *Electrochem. Solid State Lett.* **6**, C88–C91

Kim B, Ritzdorf T 2003 Electrodeposition of near eutectic SnAg solders for wafer level packaging. *J. Electrochem. Soc.* **150**, C577–C584

Kim B-H, Kim H-C, Chun K, Ki J, Tak Y 2004 Cantilever-type microelectromechanical systems probe card with through-wafer interconnects for fine pitch and high-speed testing. *Jpn. J. Appl. Phys.* **43**, 3877–81

Klingenmaier O J, Dobrash S M 1976 Peel test for determining the adhesion of electrodeposits on metallic substrates. In: Mittal K L (ed.) *Adhesion Measurement of Thin Films, Thick Films and Bulk Coatings.* ASTM, Philadelphia, PA

Kondo K, Fukui K, Uno K, Shinohara K 1996 Shape evolution of electrodeposited copper bumps. *J. Electrochem. Soc.* **143**, 1880–6

Kuepper M, Schultze J W 1997 Spatially resolved concentration measurements during cathodic alloy deposition in microstructures. *Electrochim. Acta* **42**, 3023–31

Landolt D 1994 Electrochemical and materials science aspects of alloy deposition. *Electrochim. Acta* **39**, 1075–90

Lee S, Kwon T H 2006 Mass-producible replication of highly hydrophobic surfaces from plant leaves. *Nanotechnology* **17**, 3189–96

Leith S D, Schwartz D 1999a High-rate through-mold electro-deposition of thick (200 m) NIFE MEMS components with uniform composition. *J. Microelectromech. Syst.* **8**, 384–92

Leith S D, Schwartz D T 1999b Through-mold electrodeposition using the uniform injection cell (UIC): Workpiece and pattern scale uniformity. *Electrochim. Acta* **44**, 4017–27

Leyendecker K, Bacher W, Stark W, Thommes A 1994 New microelectrodes for the investigation of the electroforming of LIGA microstructures. *Electrochim. Acta* **39**, 1139–43

Li H Q, Ebrahimi F 2003 An investigation of thermal stability and microhardness of electrodeposited nanocrystalline nickel–21% iron alloys. *Acta Mater.* **51**, 3905–13

Lowenheim F A 1995 *Electroplating*. Technical Reference Publications, Bristol

Lu L, Shen Y, Chen X, Qian L, Lu K 2004 Ultrahigh strength and high electrical conductivity in copper. *Science* **304**, 422–6

Magri P, Boulanger C, Lecuire J M 1996 Synthesis, properties and performances of electrodeposited bismuth telluride films. *J Mater. Chem.* **6**, 773–9

Malek C K, Saile V 2004 Applications of LIGA technology to precision manufacturing of high-aspect-ratio micro-components and -systems: A review. *Microelectron. J.* **35**, 131–43

Marti J L, Lanza G P 1969 Hardness of sulfamate nickel deposits. *Plating* **56**, 377–85

Martin-Gonzalez M S, Prieto A L, Gronsky R, Sands T, Stacy A M 2002 Insights into the electrodeposition of Bi_2Te_3. *J. Electrochem. Soc.* **149**, C546–54

Massoud-Ansari S, Mangat P S, Klein J, Guckel H 1996 A multi-level, LIGA-like process for three dimensional actuators. *Micro Electro Mechanical Systems, 1996, MEMS '96, Proceedings. 'An Investigation of Micro Structures, Sensors, Actuators, Machines and Systems'. IEEE, 9th Annu. Int. Workshop*, IEEE, San Diego, CA, USA

Mehdizadeh S, Dukovic J, Andricacos P C, Romankiw L T, Cheh H Y 1990 Optimization of electrodeposit uniformity by the use of auxiliary electrodes. *J. Electrochem. Soc.* **137**, 110–17

Mehdizadeh S, Dukovic J, Andricacos P C, Romankiw L T, Cheh H Y 1992 The influence of lithographic patterning on current distribution: A model for microfabrication by electrodeposition. *J. Electrochem. Soc.* **139**, 78–91

Mehdizadeh S, Dukovic J, Andricacos P C, Romankiw L T, Cheh H Y 1993 The influence of lithographic patterning on current distribution in electrodeposition: Experimental study and mass-transfer effects. *J. Electrochem. Soc.* **140**, 3497–505

Morrissey R 1993 A versatile non-cyanide gold plating system. *Plating Surface Finishing* **80(4)**, 75–9

Nielsch K, Muller F, Li A P, Gosele U 2000 Uniform nickel deposition into ordered alumina pores by pulsed electrodeposition. *Adv. Mater.* **12**, 582–6

Nilson R H, Griffiths S K 2003 Natural convection in trenches of high aspect ratio. *J. Electrochem. Soc.* **150**, C401–12

Park K Y, Park J Y, Lee J C, Park J Y, Kim G H, Kim D W, Bu J U, Chung K W 2000 A new micromachined bandpass filter on a quartz substrate. *Jpn. J. Appl. Phys.* **39(6B)**, 638–40

Park S, Schift H, Padeste C, Schnyder B, Kotz R, Gobrecht J 2004 Anti-adhesive layers on nickel stamps for nanoimprint lithography. *Microelectron. Eng.* 73–4, 196–01

Perez-Prado M T, Vlassak J J 2002 Microstructural evolution in electroplated Cu thin films. *Scr. Mater.* **47**, 817–23

Puippe J C and Leaman F (eds.) (1986) *Theory and Practice of Pulse Plating.* American Electroplaters and Surface Finishers Society, Orlando, FL

Safranek W H 1986 *The Properties of Electrodeposited Metals and Alloys*. American Electroplaters and Surface Finishers Society, Orlando, FL

Schlesinger M and Paunovic M (eds.) 2000 *Modern Electroplating.* John Wiley and Sons, New York

Slavcheva E, Mokwa W, Schnakenberg U 2005 Electrodeposition and properties of NiW films for MEMS application. *Electrochim. Acta* **50**, 5573–80

Snyder G J, Lim J R, Huang C K, Fleurial J P 2003 Thermoelectric microdevice fabricated by a MEMS-like electrochemical process. *Nat. Mater.* **2**, 528–31

Stoney G G 1909 The tension of metallic films deposited by electrolysis. *Proc. R. Soc. London* **82**, 172–5

Takahashi M, Katou Y, Nagata K, Furuta S 1994 The composition and conductivity of electrodeposited Bi–Te alloy films. *Thin Solid Films* **240**, 70–2

Talin A A, Marquis E A, Goods S H, Kelly J J, Miller M K 2006 Thermal stability of Ni–Mn electrodeposits. *Acta Mater.* **54**, 1935–47

Thommes A, Stark W, Leyendecker K, Bacher W, Liebscher H, Ilmenau C 1994 LIGA microstructures from a NiFe alloy: Preparation by electroforming and their magnetic properties. In: Romankiw L T, Herman D A Jr. (eds.) *Proc. 3rd Int. Symp. Magnetic Materials*, *Processes*, *and Devices PV* New Orleans, LA, 94-6

Thuvander M, Abraham M, Cerezo A, Smith G D W 2001 Thermal stability of electrodeposited nanocrystalline nickel and iron–nickel alloys. *Mater. Sci. Technol.* **17**, 961–70

Tittes K, Bund A, Plieth W, Bentien A, Paschen S, Plotner M, Grafe H, Fischer W J 2003 Electrochemical deposition of Bi_2Te_3 for thermoelectric microdevices. *J. Solid State Electrochem.* **7**, 714–23

Wang W, Leith S D, Schwartz D T 2002 Convective–diffusive mass transfer inside complex micro-molds during electrodeposition. *J. Microelectromech. Syst.* **11(2)**, 118–124

Wang W, Jia F, Huang Q, Zhang J 2005 A new type of low power thermoelectric micro-generator fabricated by nanowire array thermoelectric material. *Microelectron. Eng.* **77**, 223–9

Watanabe H, Hayashi S, Honma H 1999 Microbump Formation by noncyanide gold electroplating. *J. Electrochem. Soc.* **146**, 574–9

West B D A A C 1996 Workpiece, pattern, and feature scale current distributions. *J Electrochem. Soc.* **143**, 486–92

Wu Q, Fu J, Gu X, Shi H, Lee J 2003 MEMS-based transmission lines for microwave applications. In: Chiao J C, Varadan V K, and Cané C (eds.) *Smart Sensors*, *Actuators*, *and MEMS*, *Proceedings of the SPIE*, Maspalomas, Gran Canaria, Spain

Xiang Y, Chen X, Vlassak J J 2002 The mechanical properties of electroplated Cu thin films measured by means of the bulge test technique, in *Thin Films: Stresses and Mechanical Properties IX*, edited by Ozkan CS *et al*. (*Mater. Res. Soc. Sym. Proc.* **695**, Warrendale, PA, 2002), pp. 189–95.

Biographies

Matt Losey is Director of MEMS Process Integration at Touchdown Technologies in Baldwin Park, CA, a start-up specializing in MEMS-based electrical test probe cards. Prior to joining Touchdown Technologies, he worked for 3 years in the Microsystem Processing Department at Sandia National Laboratories in Livermore, CA, doing electrodeposition process research and development. Prior to Sandia, he worked for 3 years in the Silicon Technology Development group at Texas Instruments in Dallas, TX, doing research in copper electrodeposition for submicron interconnects.

James Kelly is a Research Staff Member in the Electrochemical Processes Group at the IBM TJ Watson Research Center in Yorktown Heights, NY. Prior to joining IBM, he worked for 4 years in the Microsystem Processing Department at Sandia National Laboratories in Livermore, CA, doing electrodeposition process and materials development.

1.11 LIGA

Ulrike Wallrabe and Jan Korvink, Department of Microsystems Engineering-IMTEK, University of Freiburg, Freiburg, Germany

Jürgen Mohr, Institute for Microstructure Technology, Karlsruhe, Germany

1.11.1 Introduction

The LIGA process was first developed at the Forschungszentrum Karlsruhe GmbH in Germany towards the end of the 1970s. The process was required as part of the so-called nozzle separation process required for uranium isotopes (Becker *et al.* 1982, 1986), which needed inexpensive yet very small slotted separation nozzles. Once in place, it was soon recognized that the process could also accurately produce other high aspect ratio microstructures, and nowadays it is employed worldwide by a number of other research institutes and, increasingly, by industries.

The strength of LIGA lies in its aspect ratio, as it combines lateral resolution at the micrometer range and lower level with heights up to the millimeter range. The sidewalls have unprecedented verticality, yet lateral shapes can be freely formed. Another attractive feature is the alternative it offers to silicon-based microstructures, and indeed the designer can choose among polymers, metals, and ceramics, but of course, structuring can also be performed on top of a silicon substrate.

In essence, the process follows the steps outlined by its acronym: lithography is performed with highly parallel synchrotron radiation (X-rays), and after a developing step, metal structures are grown in the cavities via galvanic forming (or electroplating) upon which the polymer mold is stripped away. In the last step, the metal structures can be replicated by using a replication technology such as embossing, known in German as Abformung.

Since its invention, many interesting variants and improvements to the process have emerged, but its advantage is certainly the highly collimated and energetic beam of X-rays of a synchrotron that penetrate, with only marginal scattering, through hundreds of micrometers of polymer resist and thereby pattern it with exceptionally sharp-edged, smooth vertical sidewalls.

With the advent of the ultraviolet (UV)-sensitive resist family SU-8, the LIGA concept has been extended to UV lithography, thereby avoiding the need for a synchrotron source. Although the process cannot achieve the same qualities as traditional LIGA, it is growing in popularity and is known under the names UV–LIGA and poor man's LIGA. The fabrication of microstructures by the LIGA process is schematically represented in **Figures 1–4**.

In the first step of the process, also known as deep X-ray lithography (DXL), synchrotron radiation with a wavelength spectrum of 0.2–0.6 nm is used to transfer the pattern of an X-ray mask into a layer of resist. The resist, several hundreds of micrometers thick, is usually made of polymethylmethacrylate (PMMA, better known as Plexiglass). Its high contrast to radiation is

Figure 1 X-ray lithography with a thick resist film using an X-ray mask. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

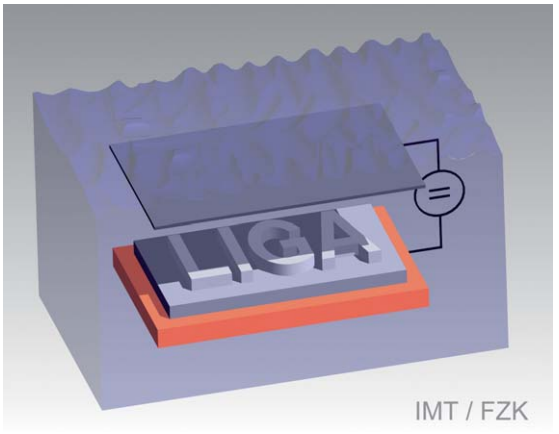


Figure 2 Electrodeposition of metal in a resist template. Typical metals are Ni, Au, Cu, or NiFe. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

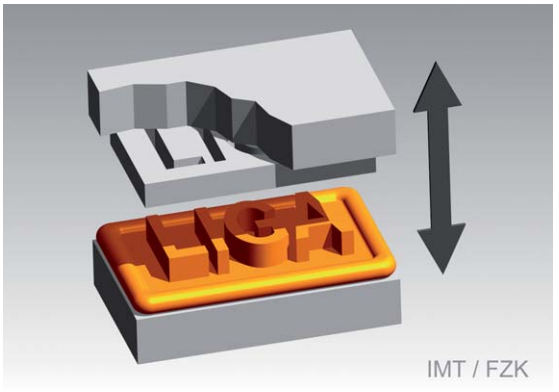


Figure 3 Molding (here hot embossing) of a polymer using a metallic molding tool. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

known from electron beam lithography. Nowadays, chemically strengthened X-ray-sensitive resists (Schenk *et al.* 1997) as well as the epoxy-based resist material, SU-8, are also used in the LIGA process.

The X-ray radiation that passes the mask is absorbed into the resist and causes a localized chemical or structural modification of the polymer chains. In PMMA, the localized molecular weight decreases due to scission of individual polymer chains. By using molecular-weight-dependent developer solvents, exposed regions of polymer can be selectively dissolved, leaving the desired cavities in the resist material. It is possible to create metal geometries complementary in structure to the cavities by

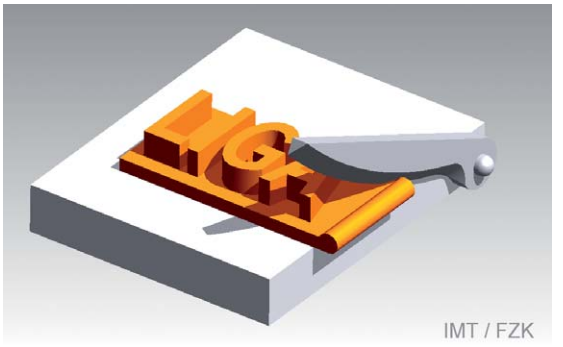


Figure 4 Postprocessing of molded component, here dicing. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

electroplating within the cavities. An electrically connected base plate has only those regions exposed that correspond to the dissolved resist, and hence the metal, either copper, nickel, or gold, will be selectively electrodeposited. Once the remaining resist has been removed, the remaining metal part can either end up as the final part or it can be used as a tool, a mold insert, for subsequent mass production by embossing or injection molding. In the latter case, the electrodeposition step is continued until the metal part has grown a thick metallic layer above the resist surface. This continuous and solid base plate strengthens the fragile microstructures underneath.

The mold insert allows the high-quality manufacture of numerous plastic copies of the microstructure by injection molding, reaction injection molding, or hot embossing (see **Figure 3**). Replication techniques extend the materials palette to numerous other thermoplasts, e.g., polyoxymethylene (POM), polyether ether ketone (PEEK), polyvinylidene fluoride (PVDF), polysulfone (PSU), and polycarbonate (PC). These molded parts can either be the end product or again be used for an intermediate step. As end products, they can be filled with metal through electrodeposition, and as intermediate forms, they may become lost templates during the fabrication of, for example, ceramic microstructures.

The key characteristics of the LIGA process are as follows:

- Structure heights of up to 3 mm, depending on design
- Freedom in lateral shape
- Smallest achievable lateral dimension of a structural detail down to 200 nm

- Highest achievable aspect ratio of freestanding structures is 50, and of supported structures is 500
- Surface roughness of sidewalls is less than 20 nm
- A broad range of materials for the final structures:
  Polymers: PMMA, POM, PSU, PEEK, PVDF, PC, liquid-crystal polymer (LCP), polyamide (PA), and polyethylene (PE)
  Metals: Ni, Cu, Au, NiFe, NiP
  Ceramics: lead zirconium titanate (PZT), $Pb(Mg_{1/3}Nb_{2/3})O_3$–$PbTiO_3$ (PMNT), Al_2O_3, ZrO_2

The long LIGA process chain is the reminiscent of semiconductor manufacturing. Thus, in order to make the LIGA process available as a competing manufacturing technology, the individual process steps have to be stable and achieve a high yield. A major effort in development was therefore directed toward establishing the so-called standard processes based on the better-characterized PMMA resist. Other professional equipments that were developed included X-ray scanners, megasonic development tanks, and electroplating facilities.

The ensuing and detailed description of the LIGA technology is presented in two sections. The first part is concerned with the process itself, whereas the second part provides an overview of typical applications that can be achieved with the technology. Because lithography is the most critical process step and X-ray lithography places strong constraints on the employed masks, it is explained in somewhat more detail than the other processing steps. The applications will focus on optical and mechanical devices that range from industrially commercialized products all the way to components that are still under investigation at either the research or the development level.

1.11.2 The LIGA Technology

1.11.2.1 Generation of X-rays

The resolution in lithography is directly dependent on the wavelength of light used. Optical lithography made an unexpected progress by the repeated development of new exposure equipment that already works in the far UV region. Even smaller wavelengths in the range from 0.2 to 2 nm are provided, for example, by a synchrotron. A synchrotron is a storage ring for relativistic electrons, which emit characteristic X-rays while accelerated on a circular path. In the case of synchrotron radiation, the X-rays are extremely parallel, which allows for the exposure of very thick resists behind the mask without loosing the required accuracy. In addition, they are highly energetic and very powerful, i.e., the quantum flux is high, which guaranties for sufficient penetration depth at still moderate exposure times.

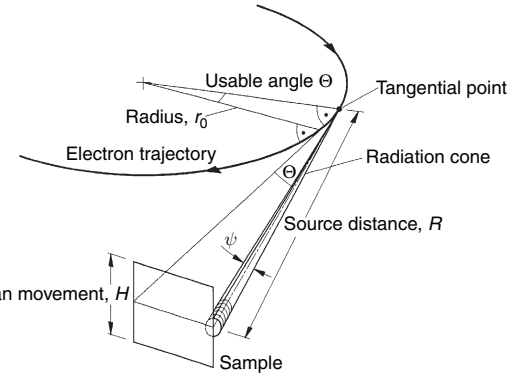


Figure 5 Principle of the generation of synchrotron radiation.

Figure 5 shows schematically the generation of synchrotron radiation. A large number of free electrons run in an electron storage ring at a highly relativistic energy, that is, almost at the speed of light, on a circular path with a fix annular frequency. Due to the strong centrifugal acceleration that the electrons suffer they emit an electromagnetic radiation along the whole circumference of the storage ring.

The radiation is extremely collimated in vertical direction, and it is emitted in tangential direction from the electron's path of flight. The opening angle ψ of the beam is dependent on the electron energy E, and is typically on the order of 1 mrad.

$$\Psi = \frac{m_e c^2}{E} \qquad [1]$$

The radiation power P and the characteristic wavelength λ_c are also strongly dependent on the energy, i.e., with the fourth and third power on the radius R of the ring and of the electron current I, respectively.

$$P = 88.5 \times \frac{E^4 I}{R} \qquad [2]$$

$$\lambda_c = 5.99 \times \frac{R}{E^3} \qquad [3]$$

The emitted power as a function of wavelength is plotted in Figure 6 for several electron energies. The electron current was chosen to be $I = 0.1$ A, and the radius of the ring $R = 10$ m. The characteristic wavelength is also given. It is defined in a way that the

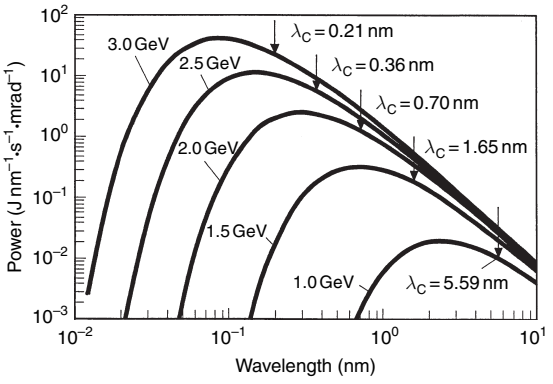


Figure 6 Emitted synchrotron radiation power as a function of wavelength for various energies of the stored electrons.

integrated radiation power above and below λ_c is equal. The wavelength with the maximum radiation power, however, is different from λ_c by a factor of 0.650.

Different from the basic sketch of a synchrotron featured above, a real synchrotron is not a circle but a polygonal-like shape. The bending radius of the electron flight path is defined by dipole magnets that are arranged regularly along the storage ring. In this case, the synchrotron radiation is generated only at the dipole magnets; hence, the steel pipes that guide the X-rays toward the exposure chamber, the so-called beamlines, are flanged to the ring in a tangential orientation only here. Between the dipole magnets, that is, in the straight ring sections where no centrifugal force acts on the electrons, crosswise arranged quadrupole magnets are placed for refocusing the electron beam, which otherwise tends to spread because of electrostatic repulsion. Additional sextupole magnets provide an energy refocusing, as the electrons suffer energy loss due to the emitted radiation on the one hand and due to scattering at residual gas molecules on the other hand. In order to keep the scattering losses to a minimum, the whole synchrotron is evacuated to vacuum, which is of the order of 2×10^{-9} mbar. **Figure 7** shows a schematic top view of the synchrotron ANKA in Karlsruhe, Germany. The polygonal shape, the star-shaped wall of concrete that surrounds the ring in order to protect the environment against the X-rays and the tangential beamlines, which are also shielded by concrete walls, can be seen. The electrons are generated and preaccelerated in the injector, which is positioned in the center of the ring. Here, the electrons are generated by a hot cathode and are brought to an energy of 500 MeV by a microtron or by a linear accelerator. From there, they are injected into the ring and accelerated to their final speed by a kicker magnet.

For using synchrotron radiation for lithographic purposes, an exposure chamber is flanged to the beamline. The mechanical separation of the vacuum in the ring and an air or helium atmosphere in the exposure chamber is done by a vacuum proof but X-ray transparent vacuum shielding. Typically this is a beryllium metal sheet that combines mechanical stability with minimum X-ray absorbance (see Section 1.11.2.2.2). **Figure 8** shows a basic X-ray lithography setup. The X-rays from the beamline enter the exposure chamber through the vacuum shielding. Due to the fact that the X-ray beam is so well collimated (it spans vertically 1 cm at a distance of 10 m to the ring), the substrates need to be scanned through the beam to ensure

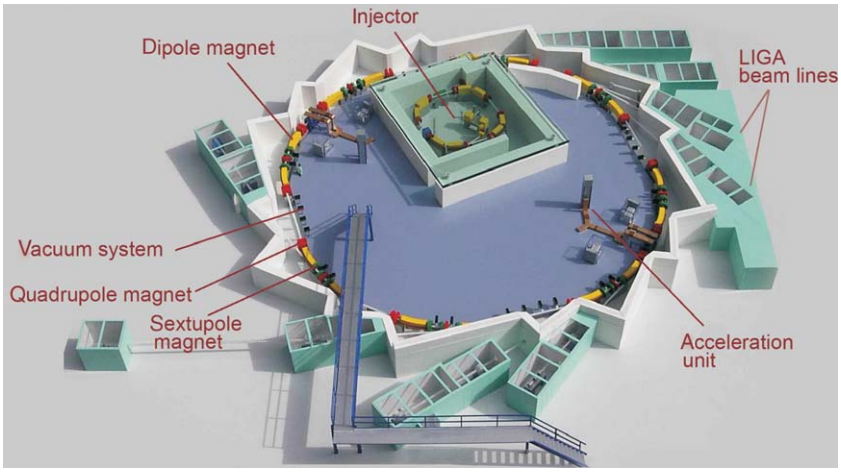


Figure 7 Top view of the synchrotron ANKA in Karlsruhe, Germany. On the left side there are three beamlines for LIGA applications.

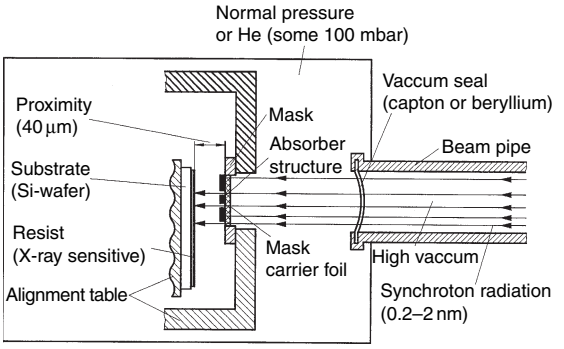


Figure 8 Schematic drawing of an X-ray lithography setup for shadow projection with beamline, mask holder, and scanner table for the substrate.

Figure 9 Professional scanner from Jenoptik Company, Germany: the mask is fixed to the scanner table in the rear side and the substrate on the front plate.

sufficiently large exposure areas. Hence, the exposure chamber features a substrate holder and a mask holder for the X-ray mask on a movable scanner table. Between the mask and the substrate remains a proximity gap of several tens of micrometers to protect the fragile membrane mask (see Section 1.11.2.2.2) from the damage. **Figure 9** shows a professional scanner from Jenoptik Company, Germany, from the user's view. The mask is mounted onto the back plate and the substrate onto the front plate.

A pure shadow projection of the X-rays through a suitable shadow mask is recommended as X-ray optics are very expensive and available only to a limited extent. In addition, as the synchrotron radiation is so highly parallel and highly energetic, it may penetrate into an X-ray-sensitive resist by hundreds of micrometers without the risk of dramatic structural losses and size deviations. This finally gives rise to the uniqueness of X-ray lithography, namely of DXL; it allows the patterning of very thick resists with extremely high resolution and very smooth sidewalls.

1.11.2.2 X-ray Masks

A mask suitable for the LIGA process consists of an X-ray absorber, a carrier foil or a substrate for the absorbers, and the frame of the mask (Bacher *et al.* 1991). In contrast to masks for optical lithography, the absorbers for high-energetic X-rays need to be quite thick, whereas the carrier is typically very thin to ensure sufficient transparency. These requirements call for a specific mask fabrication process.

1.11.2.2.1 X-ray opaque absorbers

The pattern that is to be transferred in a thick X-ray-sensitive resist is given by the pattern of the absorber. While in optical lithography using UV light, an approximately 0.1-µm-thick chromium layer on the mask is already sufficient, the absorber in X-ray lithography must consist of a material with a high absorbance for X-ray radiation in the wavelength region used. Materials, such as gold, tantalum, or tungsten, that have a high atomic weight and therefore a high absorption coefficient are considered. Mostly gold is used due to its ability of electroplating deposition. Tantalum and tungsten find few applications and are structured by reactive ion etch processes.

The retention capability of X-ray radiation is given not only by the absorption coefficient $\alpha_{Au}(\lambda)$ but also by the thickness d_{Au} of the layer. The transmission of a layer is given by the following equation:

$$T_{Au}(\lambda) = \exp(-\alpha_{Au}(\lambda) d_{Au}) \quad [4]$$

The minimum thickness of the gold absorber depends on the characteristic wavelength of the synchrotron radiation, and it increases with decreasing wavelength and increasing resist thickness. As a rule of thumb, 20 µm of gold can be considered as an appropriate average.

Because the gold absorbers are made by electroplating, they are deposited on a negative form of a resist template, which is somewhat higher than the absorber thickness, i.e., 30–40 µm. With the exception of structuring with synchrotron radiation itself, there is no existing process that guarantees the required precision on the one hand and that can grant for the structuring of resist layer thicknesses of several tens of micrometers and a precision in the submicrometer region, on the other hand. Therefore, such a mask has to be fabricated in two steps. First, an initial

X-ray mask with an absorber thickness of only 3 μm is produced, a so-called intermediate mask. The pattern of the intermediate mask is then transferred by X-ray lithography in an approximately 30-μm-thick resist layer, which after development serves as a template for 20-μm gold electrodeposition to form a second mask. The mask produced in this way is called the process mask. The synchrotron radiation is linked to almost no structure loss in this copy step because of the high parallelism and the small wavelengths.

In total, the mask production in the LIGA process comprises the following process steps:

- Manufacture of a carrier substrate for the intermediate mask
- Structuring the resist layer of the intermediate mask
- Electroplating of the absorber structure with gold
- Copying the intermediate mask to a process mask by the first LIGA-like process

1.11.2.2.2 X-ray transparent carriers for absorbers

The absorber structures have to be fabricated on a suitable carrier (Schomburg *et al.* 1991). In optical lithography approximately 2-mm-thick polished glass or quartz substrates are used. However, in X-ray lithography glass of this thickness is not transparent enough to the relevant X-rays. Hence, the carrier must not only have a low absorption coefficient, but also be thin enough to ensure sufficient transparency. Therefore, materials with low atomic weight, for example, beryllium, carbon (diamond), silicon and its compounds, plastics, or metals with low atomic number are considered as carrier materials. Additional requirements on the optimum material are mechanical rigidity, dimensional stability, and – very important – resistance against X-ray radiation. All these limit the use of plastics. Therefore, earlier, thin metal foils have been used as the carrier material in the LIGA process. Only in that way was it possible to realize a large area of a freely strained carrier (e.g., 25 mm × 65 mm) with low stress. Among the metals, beryllium shows an ideal transmission (**Figure 10** shows the absorption coefficient *vice versa*) even at a thickness of several hundreds of micrometers. The transparency is still high in the useful wavelength region. However, because of high toxicity (dust from beryllium and its oxides can cause lung cancer), the handling and mechanical treatment of beryllium is difficult. As a consequence, beryllium substrates are delivered only by very few companies, and they are quite expensive (>1 k€ per 4″ substrate). Nevertheless, in the past years beryllium gained more and more interest because it guarantees very low mask distortions generated by thermal load during the X-ray exposure, and they are mechanically stable due to their mentionable thickness.

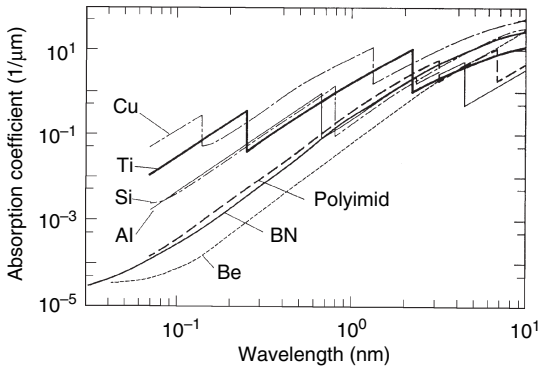


Figure 10 Absorption coefficient of diverse low-absorbance materials, which are, in principle, suitable carriers for absorber structures.

Titanium is an often-used alternative to beryllium as a carrier layer. Because of the high absorption coefficients of titanium compared to beryllium, the membrane thickness must be considerably smaller and should only be several micrometers; a typical value is 3μm. The thin carrier foils are stretched across a frame to give the necessary mechanical robustness to the mask and to ease the handling during fabrication or alignment.

1.11.2.2.3 Fabrication of an intermediate mask

1.11.2.2.3.(i) Fabrication of the carrier membrane For the fabrication of a 3-μm-thick titanium membrane, first a layer of carbon is sputtered onto a silicon wafer. During sputtering, a ring-shaped aperture covers the outer edge of the wafer so that no carbon is deposited at the periphery. Substrate temperature and other coating parameters are adjusted in such a way that the carbon adheres poorly to the wafer. Then, a titanium layer of 2- to 3-μm thickness is sputtered on top of the carbon layer. This layer exhibits good adhesion to the substrate in the outermost periphery of the wafer, while that on the carbon layer is applied only relatively loosely. After the patterning of the mask, a rigid frame from Invar (a metal alloy with very low thermal expansion coefficient) is glued onto the inner part of the titanium layer (**Figure 11**). After

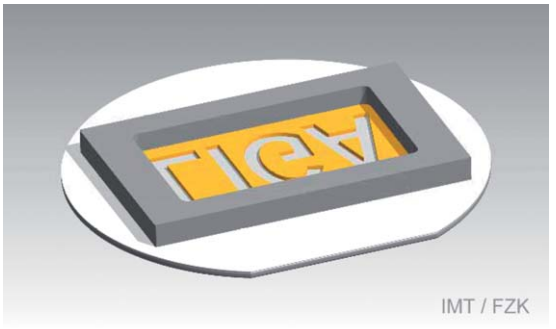


Figure 11 Electroplated mask pattern in gold with mask frame glued onto the substrate.

Figure 12 Release of the titanium membrane from substrate.

hardening the glue, the frame together with the titanium membrane can be lifted off the support by gently bending the silicon wafer resulting in a carrier membrane with excellent surface finishing, which is freely suspended onto a frame (**Figure 12**).

1.11.2.2.3.(ii) Patterning of the resist Depending on the requirements that the microstructures must fulfill, different processes can be utilized for the patterning of intermediate masks. For structures that do not necessarily call for the highest precision, the X-ray intermediate mask can be produced by an optical copy of a conventional chromium–quartz mask in an approximately 3-µm-thick photosensitive resist. The most perpendicular sidewalls can be attained by optimizing the baking process and the exposure and development conditions so that structures with a slope angle of about 88° are possible (Schulz *et al.* 1996). The minimum achievable lateral dimensions are about 2 µm because of the unavoidable effect of diffraction in optical lithography and

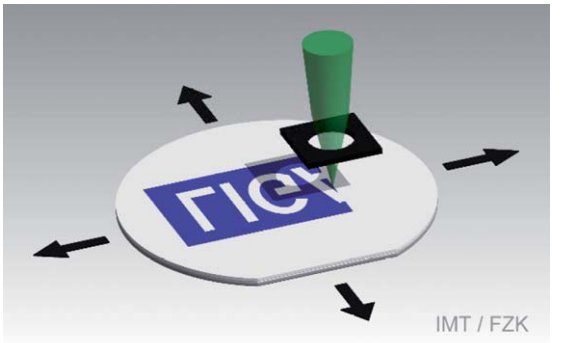


Figure 13 Writing of the master pattern with an electron beam writer. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

due to the resist thickness of 3 µm. Apart from that, rounding of edges is unavoidable. The structural loss, that is, a decrease in the structural width of about 0.5 µm, can be controlled by a corresponding size compensation on the chromium mask.

For high-precision requirements, electron beam lithography is recommended as indicated in **Figure 13**. For this purpose, an electron-sensitive resist of about 3 µm thick is spun onto the wafer with the titanium layer. An electron beam writer with high acceleration voltage (e.g., 100 keV) is suitable for direct structuring of an approximately 3-µm-thick resist layer (Hein *et al.* 1992). The 100-keV acceleration voltage ensures the required steepness of the sidewalls because the electron lobe that results from scattered electrons in the material is shifted into the substrate. PMMA is used as the resist because it provides a very high resolution in electron beam lithography, however at the prize of a low sensitivity, which results in long writing times. The 3-µm-thick PMMA layers are produced by spin coating twice with a very homogeneous thickness, whereby an annealing step is carried out after each coating in order to reduce internal cracking and avoid stress cracking. In case of less demands on the precision, negative diazo-based resist materials are also available representing a compromise between resolution and writing time.

In order to attain precision in the submicrometer range, the pattern is divided in a fine region, which contains the border, and in a coarse inner region. The fine region of a width of 1 µm, for example, is exposed to a small beam diameter (0.02–0.05 µm). For the coarse region, a considerably larger beam

diameter of up to 0.5 µm is sufficient. Thus, the border can be structured with high precision and at the same time the total writing time is considerably reduced, as the writing time per surface unit for the coarse region is about a factor of 50 smaller than it is for the fine region. This procedure results in a smallest line and space structure of about 1 µm, limited by the proximity effect. The dimensions of the structure details, however, are between 0.1 and 0.2 µm.

1.11.2.2.3.(iii) Electroplating of gold The electroplating of microstructures is as an essential process step of the LIGA process (see Section 1.11.2.7). As for the fabrication of both intermediate mask and process masks, electroplating of gold is required (Maner *et al.* 1988). Typically, in the macroscopic world, cyanide electrolytes are used. However, the cyanide electrolytes do have some disadvantages; above all the toxicity, but they also attack the resist and produce rough surfaces. By using suitable sulfite gold electrolytes instead, fine-grained depositions with smoother surfaces can be achieved.

The essential requirements of gold electroplating of the absorbers are as follows:

- The absorber structure must adhere well to the carrier material, as the cross-sectional area of the absorber structure often amounts to only a few square micrometers.
- The deposition should be carried out with the least amount of stress as possible in order to provide good adhesion of the absorber to the carrier after resist stripping and to avoid mask distortions.
- The thickness of the deposited layer must show a high uniformity, both in the microscopic as well as in the macroscopic range so that a uniform and the greatest possible X-ray contrast results.
- Also, pattern details of the resist structure in the submicrometer range must be copied by electroplating.

As titanium belong to the materials that are difficult for electrodeposition of other metals, either suitable intermediate layers or a chemical pretreatment is necessary. For example, a wet chemical oxidation of the surface has proven to be a suitable method (Mohr *et al.* 1988). It increases not only the adhesion of the resist, but also the adhesion of the gold absorbers. In order to ensure a uniform initiation of the electroplating, a gold starting layer of about 10 nm thick is sputtered onto this titanium oxide layer.

1.11.2.2.4 Fabrication of a process mask The process mask is produced by transferring the pattern of the intermediate mask into a relatively thick resist layer using synchrotron radiation (**Figure 14**). PMMA of 30–40 µm thickness is used as a resist material, which is applied to the mask carrier by direct polymerization (see Section 1.11.2.3.1). For the process mask, besides the titanium membranes, the above-mentioned beryllium substrates can also be used as a carrier. They are produced by fine mechanical processing of metal sheets with a currently limited thickness of 500 µm. This metal sheet is inactivated on both sides by a silicon nitride layer in order to avoid corrosion during utilization. A polyvinylidene (PVD)-deposited gold layer is used as the electroplating starting layer. This may cause problems in resist adhesion, which have to be overcome by initial adhesion promoters. If titanium is used for the process mask, the same fabrication sequence as for the intermediate mask can be used for the membrane.

As the X-ray lithography is a pure shadow projection of the absorber structure, the process mask features an exactly negative image of the intermediate mask after development and electroplating of gold. Even the lateral roughness of the absorber structures of the intermediate mask, which is only on the order of 20 nm, is completely transferred to the process mask. Because the X-ray contrast of the intermediate mask with an absorber thickness of only 3 µm is limited, the wavelength of the synchrotron radiation must be considerably larger compared to structuring of resists with a thickness of several hundreds of micrometers. Sufficient contrast between the illuminated and nonilluminated areas can be achieved only by using softer

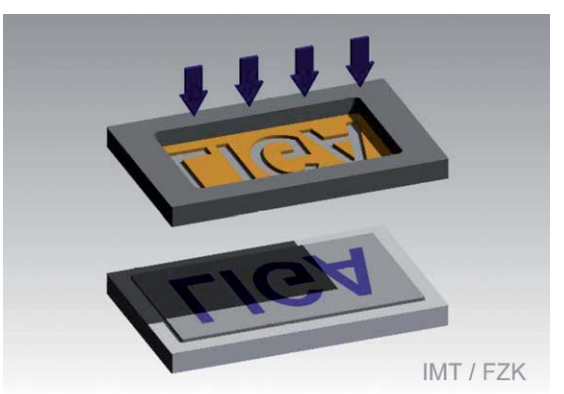


Figure 14 Copy of intermediate mask to thick resist layer for process mask using X-ray lithography.

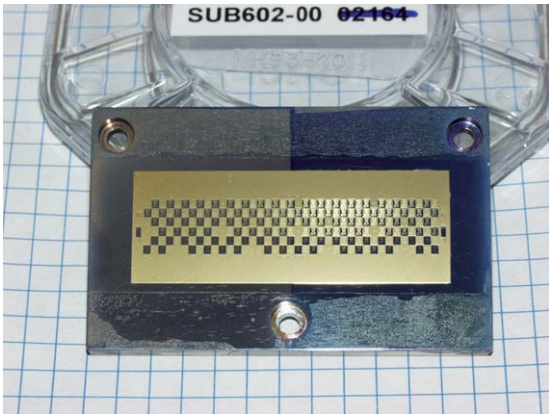


Figure 15 Photograph of a process mask on a titanium membrane: the holes in the frame are for fixation of the mask on the mask holder.

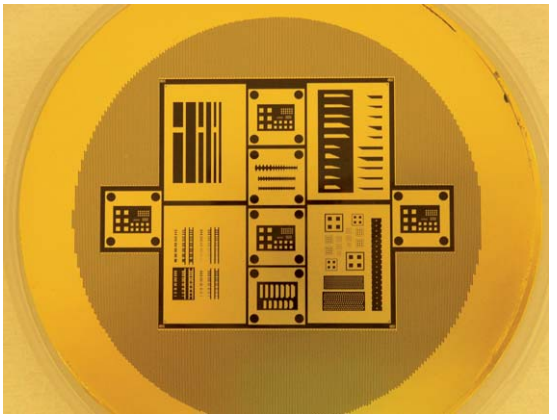


Figure 16 Photograph of a process mask on a Be substrate; this mask can be fixed by clamping the substrate due to its mechanical stability.

X-rays. A fully manufactured process mask is shown in Figure 15 for a titanium membrane and in Figure 16 for a beryllium substrate, respectively.

1.11.2.2.5 Alignment marks in X-ray process masks

If microstructures that lie accurately aligned on top of previously structured auxiliary layers are to be produced by X-ray lithography then the process mask must be adjusted relative to these auxiliary layers before the X-ray exposure. This can be done not only with the help of alignment marks that are part of the X-ray mask layout on the one hand, but also by the prepatterned auxiliary layer on the other hand. Because titanium or beryllium mask membranes are not optically transparent, perforations are locally

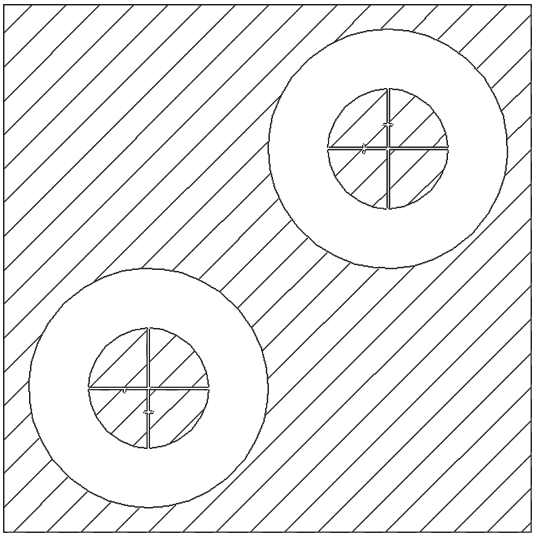


Figure 17 Alignment marks on X-ray process mask.

etched in the mask membrane through which the alignment marks can be observed. In order to etch little holes into the membrane, the entire mask is covered with a photoresist. It is opened at the positions of the alignment marks by optical lithography and wet etching. In the case of titanium, undercut etching of the photosensitive coating plays a subordinate role due to the thin membrane and hence the short etching times. In contrast, with thick beryllium masks, the entire membrane cannot be entirely etched through, as this would lead to a strong undercut due to isotropic etching. For these reasons, mechanical cavities must be introduced into the beryllium before etching.

The alignment marks, typically thin- and long-legged crosses, on the X-ray mask are produced simultaneously with the other absorber structures. Therefore, they are precisely positioned relative to the other structures. The crosses are surrounded by a broad ring, which is also electroplated in gold and which helps to prevent undercutting and thus defines the final dimensions of the etched hole. Figure 17 shows a typical layout of the alignment marks. The four quarter sections within the rings are opened by etching and are viewed through the mask membrane with a microscope. For the alignment marks on the substrate, a respective pattern shown in Figure 18 is recommended. The arrows help the user to move the mask in the correct direction in case the starting position is so bad that the cross on the substrate is out of sight when observed through the mask hole. The alignment accuracy obtained with this procedure is on the order of 3 μm.

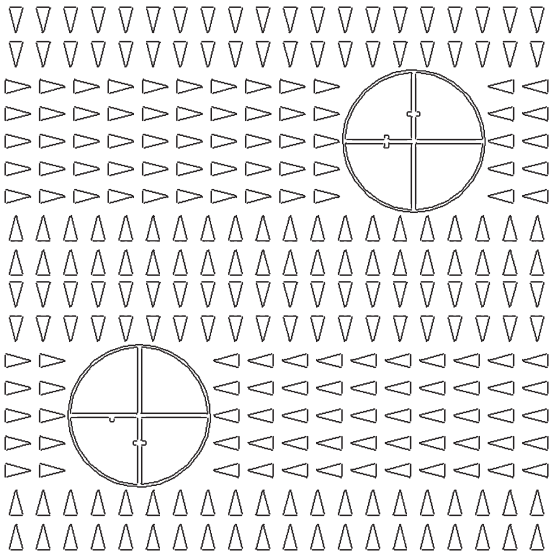


Figure 18 Alignment marks on the substrate; the arrows help orientation.

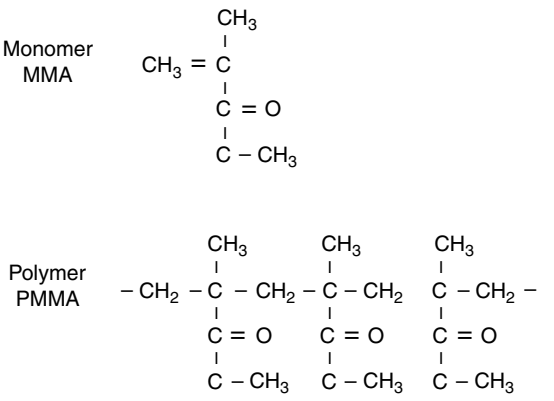


Figure 19 Structure of methacrylate (MMA) and polymethylmethacrylate (PMMA). (Source: Menz W, Mohr J, Paul O 2001 Microsystem Technology, Wiley-VCH.)

1.11.2.3 X-ray Resists

1.11.2.3.1 Production of thick resist layers

The resist layers used in microelectronics rarely exceed thicknesses of 1 μm and are usually polymers that can be dissolved to the desired viscosity with a solvent. By spin-coating the resist solution onto a substrate, the solvent is partially centrifuged off leaving a homogeneous resist layer on the substrate. The thickness of the layer can be precisely controlled by the speed of rotation.

Spin coating is, however, not suitable for thick resist layers on the order of a few hundred micrometers, as required for the LIGA process. Spin-coated layers produce only loosely intertwined polymer strands and produce large internal stresses for increasing layer thickness. During development, stresses cause nonuniform penetration of the developer solution and stress-driven cracking of the polymer material. Hence, thick resist layers are directly polymerized onto a base plate (Mohr *et al.* 1988) or alternatively, a polymer plate is glued or welded to a base plate. PMMA is almost exclusively used as the resist material.

For the direct polymerization process, the starting material is a viscous cast resin that consists of the low-viscosity monomer methylmethacrylate (MMA) and a solid fraction that is dissolved in it. After the addition of a polymerization starter, e.g., peroxide, the MMA polymerizes to PMMA, but only if the temperature is raised or through the addition of a room temperature initiator such as aniline. The molecular structures of MMA and PMMA are shown in Figure 19.

The solid additions to PMMA are not affected by the polymerization through secondary effects. The resulting molecular weight is strongly dependent on the concentration of the initiators and starters and the prevailing reaction conditions. The polymerization results in a typically bimodal molecular weight distribution. Internal stresses are reduced by subjecting the sample to a thermal tempering cycle.

In order to attach an existing polymer plate to a substrate, two variants are typically followed. In the first variant, a thin layer of the cast resin monomer is applied as a 10-μm-thick adhesion layer directly onto the base plate. The PMMA substrate is pressed directly onto this layer, and the assembly is hardened at room temperature as described before. In the second variant, a thin monomer layer is spin-coated onto the base plate. The monomer contains adhesion elements so that the application of pressure and temperature to the assembly results in the formation of a bond forming between the PMMA substrate and the base plate (Skrobis *et al.* 1995).

It is imperative that, subsequent to the lithography step that includes both irradiation and development, the finest high aspect ratio structures still sufficiently adhere to the base plate. In addition, as electroplating almost always follows lithography, the exposed adhesion layer on the substrate must guarantee uniform initiation of electroplating on the base plate. In addition, as X-ray lithography can result in the generation of

photoelectrons, the effective cross section must be very small, a requirement that is fulfilled by a material of low atomic number. In order to fulfill these requirements, a compromise is taken in the form of a sputtered titanium layer, which is subsequently oxidized in a wet chemical bath. Upon oxidization, a microporous layer is achieved, which in turn results in a microtoothing of the interface between the base plate and the substrate. To further strengthen the bond, an adhesive is added to the resin. The adhesives are typically siloxanes, which form an oxygen-bridging bond with the oxidized surface (Boerio and Williams 1981). It is also possible to use copper, nickel, or gold as adhesives and electroplating initiation layers, but none have the high adhesive strength associated with titanium. For gold, the adhesion strength can be improved by the addition of thiophenols. If the resulting microstructure is the polymer itself, then carbon is also well suited as a adhesion layer (El-Kholi *et al.* 2000). Though its adhesion strength exceeds that of titanium, it is not suitable as an electroplating initiator.

1.11.2.3.2 Beam-induced reactions and development of resists

Upon exposing a resist to radiation, its polymer chains are cracked resulting in a lower molecular weight distribution (Schnabel 1978). From **Figure 20** it can be deduced for PMMA that by increasing the dosage of radiation to high levels, it is possible to transform the molecular weight from an initial $650\,000\,\mathrm{g\,mol^{-1}}$ to a minimum that lies between 2500 and $3000\,\mathrm{g\,mol^{-1}}$.

The scission of the main chain of the polymer is initiated by the electronic excitement of the molecule's bonds. The X-ray photons reach the polymer with energies in the kiloelectron volt range and are absorbed by individual atoms via the photoeffect along the polymer chain. In turn, they release high-energy photoelectrons and Auger electrons. In this way, energy is progressively transferred to other molecules or molecular building blocks, and secondary electrons are generated. This excitation chain continues until only thermal energy remains in the form of ionized and excited molecules and thermal electrons.

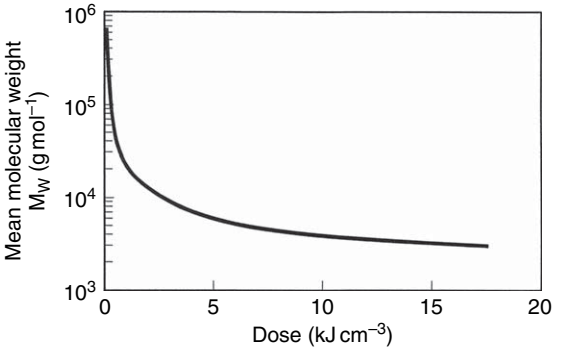


Figure 20 Decrease in molecular weight upon X-ray exposure. (Source: Menz W, Mohr J, Paul O 2001 Microsystem Technology, Wiley-VCH.)

For PMMA, the excitation results in a cleavage of the first side chain. As a result, a radical ester group and a radical carbon atom are available in the main chain of the polymer. The radical carbon atom initiates scission of the chain and results in both a radical and a saturated piece via a double bond with the free carbon atom (Schnabel 1981, Schnabel and Sotobayashi 1983). The result is therefore a short stable chain and two radicals, namely the radical polymer chain and the radical ester chain, which in turn are free to react with each other or with other radicals in the mixture in order to form stable molecules. Further postscission processes exist, namely electron or energy transfers to other molecules. Chains are also known to persist in a metastable excited state.

At higher radiation doses, not only chain scission but also cross-linking of PMMA is observed. At the same time, the molecular weight increases because chemical bonds are formed between different main chains of the polymer. Clearly, chain scission and cross-linking have opposite influences on the molecular weight distribution, even though the influence of the scission mostly dominates the cross-linking; nevertheless, it explains the limiting lower value of molecular weight achievable, as shown in **Figure 20**.

1.11.2.4 X-ray Lithography

During X-ray lithography, synchrotron radiation transfers a mask pattern to a millimeter-thick resist layer. The exposed resist material undergoes chemical modification. The degree of this modification depends on the X-ray sensitivity of the material and on the energy of the absorbed radiation. The quality of the transfer process and hence of the achievable structure geometries depends on the degree of parallelism of the radiation, the degree of radiation smearing on the absorber edges, and on the spectral range of photoelectrons produced in the resist layer.

Furthermore, secondary effects must be considered. These include fluorescence electrons produced in the mask membrane and photoelectrons that are released in the substrate. It is also important to take into account deformation of the mask due to thermal expansion of the mask membrane.

1.11.2.4.1 Requirements on the absorbed radiation dosage

PMMA is distinguished as a resist by its high contrast, i.e., by its excellent image reproducibility (Greeneich 1975). However, its resistance to radiation damage results in the need for fairly high radiation dosages in order to achieve a significant reduction in molecular weight. Before radiation, the molecular weight distribution of PMMA is typified by **Figure 21(a)**. The bimodal distribution has an average molecular weight of 600 000 (El-Kholi *et al.* 1993). Now, at 38 °C, only 50% of the molecules with weight up to 20 000 g mol^{-1} are dissolved. Because the nonradiated PMMA has only a small fraction of its molecules in this weight range, it remains relatively unaffected by the developer.

For PMMA, a radiation dose of 4 kJ cm^{-3} results in a monomodal weight distribution of 5700 g mol^{-1} (**Figure 21(b)**). Because the weight distribution lies well within the 50% dissolution range of the developer, the radiation dose is sufficient for material removal. At lower dosage levels, dissolution is insufficient and residues remain; hence 4 kJ cm^{-3} represents the lower dosage limit for PMMA. Mechanical measures such as stirring or ultrasonic agitation during developing can somewhat lower this limit.

At a radiation dose of 20 kJ cm^{-3}, the average weight distribution is around 2800 g mol^{-1} as shown in **Figure 21(c)**. All the reduced PMMA are relatively quickly dissolved by the developer. Higher dosages should, however, be avoided, as they lead to intrinsic damage such as the formation of bubbles, which, in turn, would hinder the formation of defect-free microstructures.

Thus, in order to process defect-free microstructures with complete resist removal, it is essential that the deposited radiation dosage lies between 4 and 20 kJ cm^{-3}. The dosage may therefore vary by at most a factor of 5. Now we can proceed to deduce the wavelength of the radiation. From **Figure 22** we plot, for monochromatic X-ray radiation and as a function of wavelength, the depth in PMMA at which the absorbed dose has dropped by a factor of 5. For example, for structures with a 500 µm depth, it follows that the X-ray wavelength may not exceed 0.25 nm.

Because the synchrotron radiation has a fairly wide spectral distribution, any long-wavelength components that are not yet absorbed by the X-ray window of the beam tube should be additionally filtered out using an appropriate preabsorber. **Figure 23** demonstrates how, to what extent, and where synchrotron radiation is absorbed for a typical

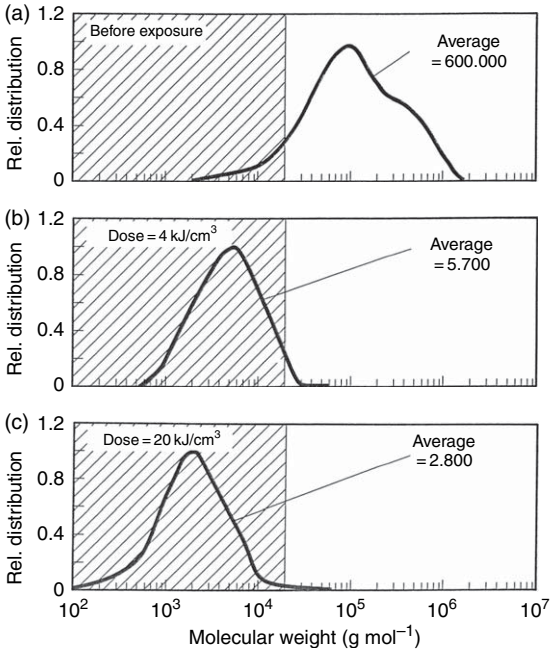


Figure 21 Distribution of molecular weight of polymethylmethacrylate (PMMA) before and after exposure at radiation dose of 4 and 20 kJ cm^{-3}, respectively. (Source: Menz W, Mohr J, Paul O 2001 Microsystem Technology, Wiley-VCH.)

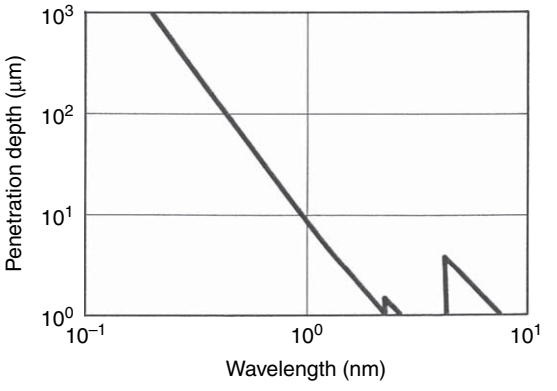


Figure 22 Influence of the X-ray wavelength on the penetration depth in polymethylmethacrylate (PMMA) for the case that the absorbed dose has dropped by a factor of 5. (Source: Menz W, Mohr J, Paul O 2001 Microsystem Technology, Wiley-VCH.)

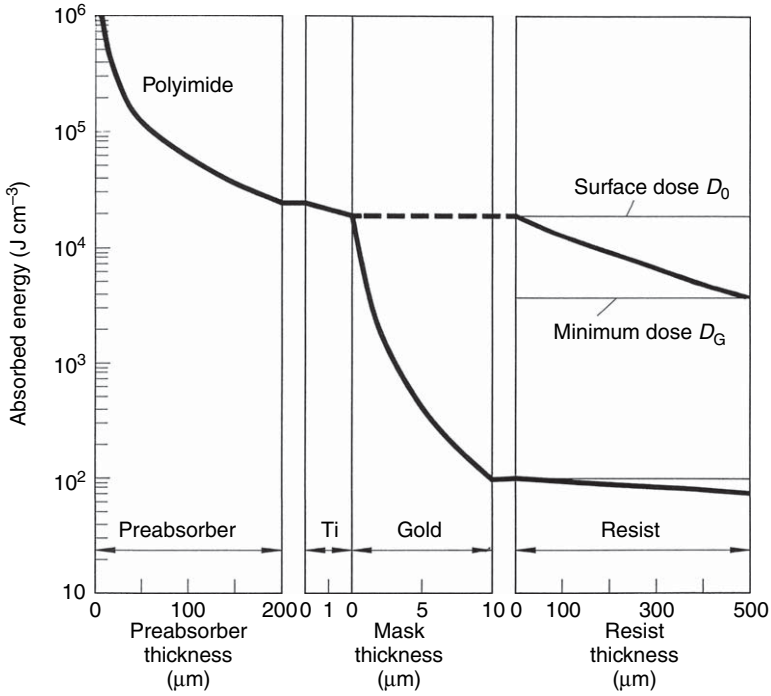


Figure 23 Absorbed energy of a 500-μm-thick polymethylmethacrylate (PMMA) layer with a mask having gold absorbers of 10-μm thickness and with a 200-μm-thick Kapton (polyimide) membrane as preabsorber. (Source: Menz W, Mohr J, Paul O 2001 Microsystem Technology, Wiley-VCH.)

microstructure experiment. The curve is typical for the electron stretcher and accelerator (ELSA), the electron stretcher ring, with a characteristic wavelength of 0.5 nm, situated in Bonn, Germany. Here the experimental conditions, especially for the irradiation time and for the choice of preabsorber, are selected so that the underside of a 500-μm-thick resist layer obtains the minimum dose of $D_G = 4\,\mathrm{kJ\,cm^{-3}}$, yet the dose does not exceed $D_O = 20\,\mathrm{kJ/cm^{-3}}$ on the upper surface. For the upper limit, the long-wavelength component of the incoming radiation is filtered out through a 200-μm-thick preabsorber. It is usually made of Kapton (polyimide), but beryllium or other materials are also appropriate. The key feature is that the material should show a high contrast between soft and hard radiation, with the soft radiation being absorbed.

In order to ensure that resist areas that correspond to shadow regions of the mask are not attacked by the developer, the deposited doses should not exceed the damage dose limit of $D_s = 100\,\mathrm{J\,cm^{-3}}$. This places a minimum contrast limitation on the mask of 200. In order to achieve this high value, the thickness of the mask's gold absorber should be at least 10 μm. For other radiation choices, it may be that mask gold layers should be even thicker. It is imperative to note that the spectrum should be optimized with respect to the radiation source and the resist type and thickness.

1.11.2.5 Development

In order to achieve a predictable production process, the regions of low molecular weight must be selectively soluble in a suitable developer. Furthermore, the specificity of dissolution must be very high, as unexposed polymer at the surface of the substrate will be exposed to the developer for long periods of time as the development process progresses. In addition, a swelling of the polymer during development is not allowed, as this will lead to stress crack formation and propagation.

A measure for the qualification of a good developer is the contrast γ, which should be as high as possible. At the dose D_c the resist will be completely dissoluted, whereas at D_i it will still not be attacked.

$$\gamma = \frac{1}{\log \frac{D_c}{D_i}} \qquad [5]$$

In addition, the development rate is, of course, material-dependent, i.e., from the molecular weight

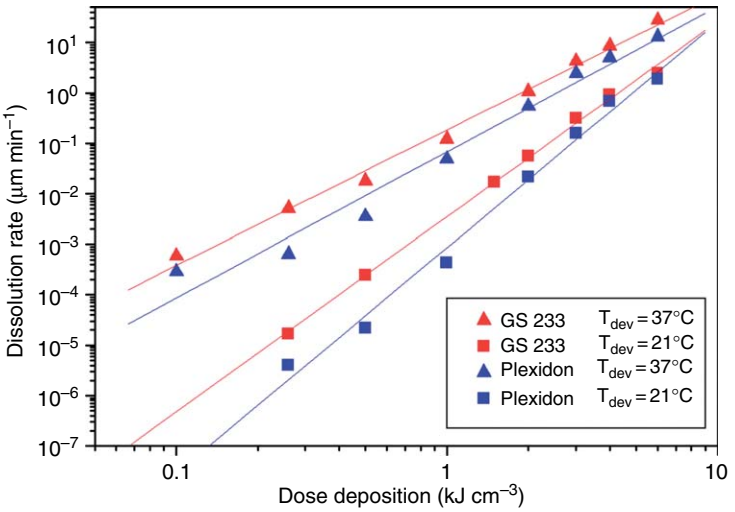


Figure 24 Developer dissolution of two different commercial polymethylmethacrylate (PMMA) brands as a function of the exposure dose.

and the development rate of the unexposed material, and it is crucially dependent on the temperature of the developer.

A developer that satisfies these requirements, suitable for PMMA X-ray lithography, is a mixture of ethylene glycol–monobutyl ether, monoethanolamine, tetrahydro-1,4-oxazine, and water (Ehrfeld *et al.* 1988). The dissolution rate as a function of the deposited dose in the PMMA is shown in **Figure 24**. Two different commercial PMMA types were investigated and the solubility was measured at 37 °C and at room temperature (Achenbach *et al.* 2000).

As expected, the dissolution rate increases with the deposited dose and with the temperature. At lower temperatures, the contrast (in **Figure 24** the slope of the straight line) of the developer increases, i.e., it becomes more selective. The two chosen PMMA brands differ in molecular weight: the Plexidon is cross-linked, whereas the GS233 is not, i.e., it features a lower molecular weight. Consequently, the dissolution rate of the GS233 is larger than that of the Plexidon.

1.11.2.6 Quality of LIGA Structures

One of the outstanding features of the LIGA process is its ability to produce high aspect ratio structures whose sidewalls show vanishing deviation from the vertical and therefore possess constant quality over the entire height of the structure and over all lateral dimensions. **Figure 25** shows a 400-µm-high test structure produced by X-ray lithography. Its width was measured optically as a function of height, and shows a deviation from the perpendicular of only 40 nm per 100 µm height (Mohr *et al.* 1988).

The factors that contribute to this variation are the inclination produced by the scan movements of the substrate through the beam; low selectivity of the developer and direct physical effects brought about by the X-ray radiation (Becker *et al.* 1984); and effects resulting from the construction of masks and samples. The different sources of error are now considered in detail.

1.11.2.6.1 Fresnel diffraction, photoelectrons

Figure 26 shows the effect of Fresnel diffraction at an absorber edge, as well as the influence of photoelectrons produced in the resist. Furthermore, the effect of the divergence of synchrotron radiation is shown.

When an incident light beam is partially shadowed by an edge, then all points along the edge have to be treated as light sources for circularly diffuse light waves. These waves in turn interfere with each other (the Huygens' principle) and lead to the phenomenon of diffraction, i.e., a local variation in light intensity. The variation at a point is a function of the position of the point relative to the edge, the wavelength of the light, and the lateral distance of the point of analysis. As a result of diffraction, instead of a sharp shadow, we obtain regions of lower intensity in the unmasked area and regions of higher intensity in the masked regions. The effect is especially pronounced for monochromatic radiation,

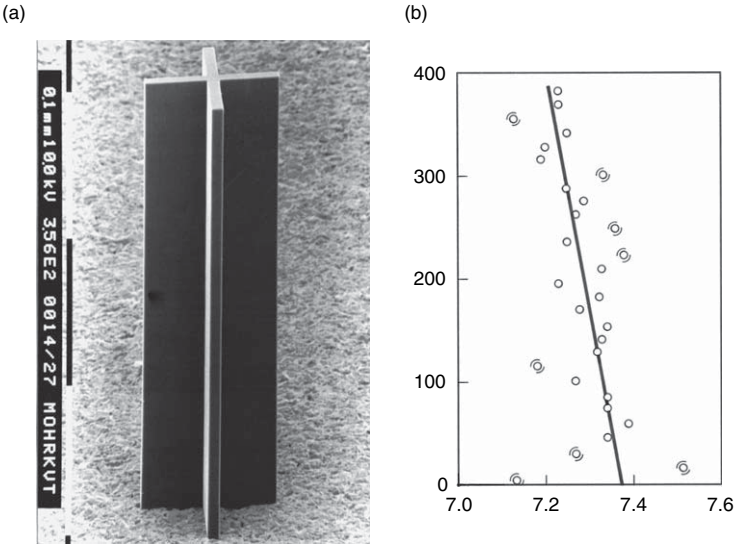


Figure 25 (a) Scanning electron microscope (SEM) picture of the 400-μm-high test structure and (b) measured linewidth in dependence of vertical position. (Source: Menz W, Mohr J, Paul O 2001 Microsystem Technology, Wiley-VCH.)

where the extinction conditions (path length difference of $\lambda/2$) are exactly met. Now, for polychromatic radiation, the entire spectrum must always be considered at every point. Because the exact conditions for extinction only corresponds to exactly one wavelength at each point, the resultant influence of diffraction is relatively small. The ideal shadow limit due to Fresnel diffraction can be calculated and is found to vary linearly with wavelength. In the wavelength region of interest, i.e., between 0.1 and 1.0 nm, this results in an uncertainty of only 0.1 μm as shown in **Figure 27**, which is calculated for a resist thickness of 500 μm.

Within the resist, the X-ray photons release photoelectrons and Auger electrons. Their reaction with the resist material is the cause of the desired chemical modifications of the resist material. The electrons impact with the molecules of the resist,

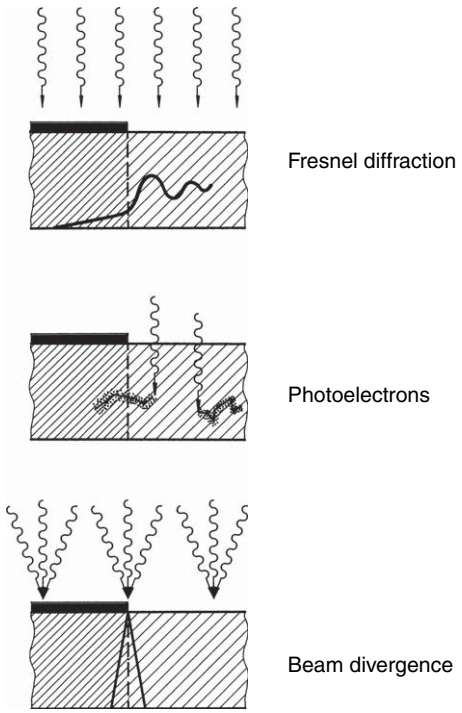


Figure 26 Effects that limit the lateral resolution of X-ray lithography. (Source: Menz W, Mohr J, Paul O 2001 Microsystem Technology, Wiley-VCH.)

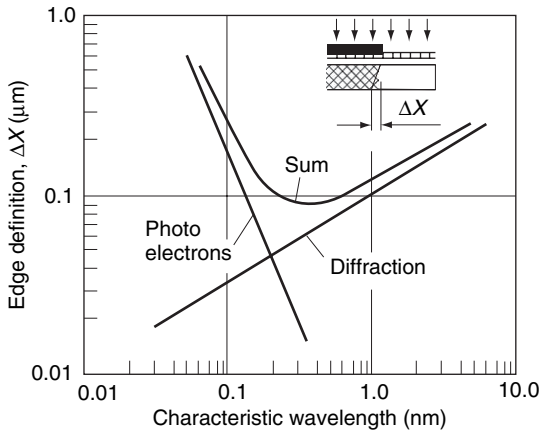


Figure 27 Total error related to Fresnel diffraction and photoelectrons. (Source: Menz W, Mohr J, Paul O 2001 Microsystem Technology, Wiley-VCH.)

thereby losing some of their energy. Due to their initial energy, they potentially have infinite range so that they can end up in unexposed regions of the resist and thereby reduce the sharpness of the image. The effective range of the electrons varies quadratically with their energy, so this effect increases with the shorter wavelengths of synchrotron radiation (see also **Figure 27**). Again, at the wavelengths of interest, the uncertainty lies around 0.1 μm. The observed error is of course a superposition of both effects and is presented as the sum of both in **Figure 27**. The minimum error is obtained with a wavelength of around 0.3 nm.

1.11.2.6.2 Divergence of radiation

The synchrotron radiation does not propagate perfectly parallel. Its finite divergence has two sources. First, there is a natural divergence due to the finite electron energy of the beam and the second source of divergence is due to oscillations of the electron orbits about the ideal reference orbit, also known as betatron orbits (Koch 1987). Neglecting the betatron orbits, an ideal current filament results in an opening angle that depends on the radiation wavelength. The angle decreases with increasing electron energy.

The divergence due to the betatron oscillation is strongly dependent on the specific details of the synchrotron, as well as position of the tangential point of the electronic orbit. Typical divergence values lie between 0.1 and 1 mrad in the horizontal and vertical directions. Now, in order to analyze the influence of beam divergence on the resulting micropart precision, we need to consider the source points in the electron orbit, which are approximately 10 m removed from the substrate. In addition, we need to consider that the sample is scanned across the beam in the vertical direction. In this way, every point is exposed to a multitude of beam angles. Hence, the vertical edges experience the full effect of beam divergence.

This contrasts strongly with the horizontal direction, where the local divergence at worst leads to a slight inclination of the structure with respect to the perpendicular. At a distance of 10 m from the electron orbit, the maximum rotation from the vertical of a 100-μm-wide structure is expected to be about 0.01 mrad, which of course can be neglected.

1.11.2.6.3 Inclination of the absorber walls to the beam

A relative inclination of the absorber walls with respect to the beam can occur in two ways. First of all, the normal of the mask and sample, which should lie parallel to the photon beam, can only be adjusted to this value to be within a finite tolerance. Second, the mask absorber walls may themselves not be perfectly perpendicular. For example, for masks produced by optical lithography, the angle of the sidewalls can be in the region of 85°.

The dose deposition under a strongly sloped absorber varies considerably. This can be used to an advantage when producing high aspect ratio conical structures, as we can achieve a lateral variation of 0.4 μm over a height of 500 μm.

1.11.2.6.4 Fluorescence radiation from the mask membrane

The absorption spectra of most materials are characterized by edges at which the absorption coefficient increases stepwise with increased radiation energy (*cf.* **Figure 10**). Such increases arise because the incoming photons are energetic enough to excite electrons from the inner K or L shells to the continuum. When such excited states recombine, fluorescence radiation is released proportional to the energy difference of the level. This radiation is generated in the mask membrane and propagates isotropically and of course can also be directed under the absorber, damaging the resist in unwanted areas. In practice, this is noticed after development as a rounding of edges of the resist and, in case of long exposure times, also in an attack of the remaining PMMA structures.

The effect will be especially pronounced when the increase of the absorption coefficient lies in a wavelength region, which overlaps with the primary synchrotron spectrum used for lithography. This is in fact the case for titanium, which is mostly employed for the mask membrane and whose absorption edge lies at 0.497 nm. In the example shown in **Figure 28**, a 1-mm-height structure, irradiated with a titanium mask, shows a rounding radius of 150 μm. The effect is very much reduced to about 2 μm when using a beryllium membrane, as it does not possess absorption edges in the wavelength region of interest (Pantenburg and Mohr 1995).

1.11.2.6.5 Production of secondary electrons from the adhesion and electroplating initiating layer

As discussed in detail earlier, the resist is mounted upon a metallic substrate, usually titanium or gold. The substrate has a considerably enhanced absorption coefficient when compared with the resist

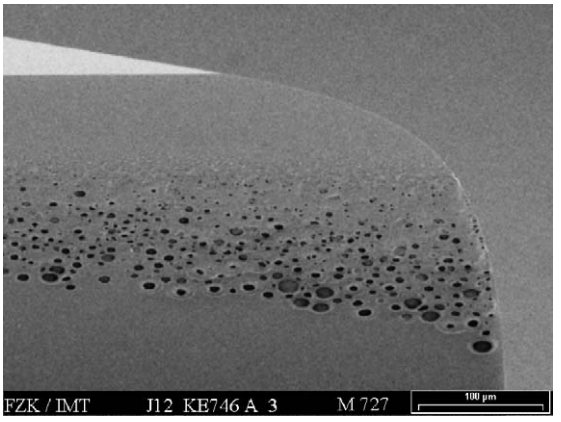


Figure 28 Typical structural loss due to fluorescence radiation on a 1-mm-thick PMMA structure after an exposure through a titanium process mask. (Source: Menz W, Mohr J, Paul O 2001 Microsystem Technology, Wiley-VCH.)

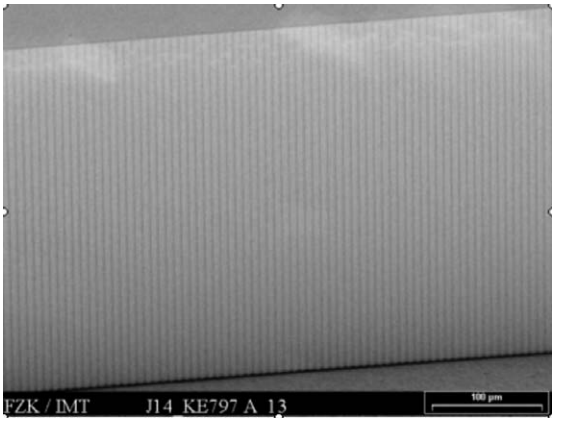


Figure 29 Microprecision component type A, grating of a microspectrometer. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

material. As a result, the effective cross section for secondary electron generation is greatly increased at the interface between the substrate and the resist. Such electrons greatly contribute to resist damage at this interface. Whereas this is useful in areas that are exposed; it has a detrimental effect on masked areas, as the adhesive bond between the resist and the substrate is thereby damaged. Two remedies exist. First, it is wise to raise the thickness of the absorber above the minimum value discussed previously. Second, the fraction of high-energy photons can be reduced by employing special X-ray mirrors.

1.11.2.6.6 Achievable aspect ratios

An important quality measure is the aspect ratio of structures, usually the height of a structure (or feature) versus the smallest achievable lateral dimension. Table 1 provides some examples of the aspect ratios that have been achieved. We distinguish between fairly large type A structures (millimeter to centimeter) that have fine details (nanometer to micrometer), for example, in case of spectrometers, and stand-alone type B microstructure components with lateral dimensions in the micrometer range as in most other cases. The table is illustrated by examples shown in Figures 29–31.

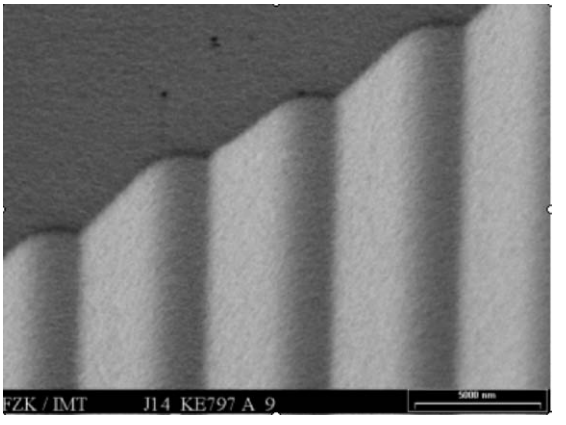


Figure 30 Close-up of reflective grating. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

Table 1 Examples of achieved aspect ratios by deep X-ray lithography

Aspect ratio	*Minimum lateral dimensions (µm)*	*Structure height (µm)*	*Type*	*Microstructure*	*Reference product*
650	0.2	130	A	Grating structure	UV–VIS spectrometer (**Figures 29** and **30**)
250	2	500	A	Grating structure	IR spectrometer
40	Diameter = 26	1100	B	Columns, here in SU-8	Columns (**Figure 31**)
30	100/100	3000	B	Lines and spaces	Lead frames

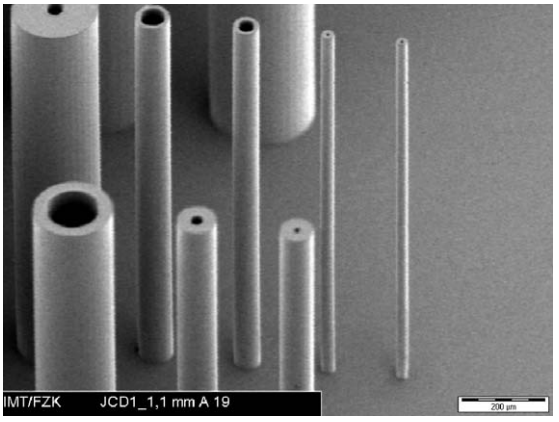


Figure 31 Microprecision component type B: columns in SU-8. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

1.11.2.7 Electroforming of LIGA Structures

X-ray lithography produces plastic microstructures directly as end product, and for micro-optical components made of PMMA this makes sense. In many cases, however, metallic parts are required, and for these the formed resist cavities are filled with metal by electroforming. For example, in order to produce high volumes of plastic microparts economically using any of the replication techniques listed in Section 1.11.2.8, robust and shape-retaining metallic mold inserts are required. Galvanic deposition after X-ray lithography is well suited to this purpose. In addition, electroforming is also an important step in producing the gold absorption masks needed for quality X-ray lithography.

Masks are predominantly electroplated in gold, whereas microstructures are typically formed from nickel. Copper plating has not nearly played the same role in LIGA operations. For special needs, alloys are also produced. For example, for high hardness structures, a nickel–cobalt bath is used and for magnetic cores, a nickel–iron bath is employed.

1.11.2.7.1 Galvanic deposition of nickel for the production of microstructures

For the galvanic production of nickel microstructures, as required for mold inserts, a nickel sulfamate electrolyte is employed. It typically contains 75–90 $g\,l^{-1}$ of nickel in the form of nickel sulfamate, 40 $g\,l^{-1}$ of boric acid as a buffer, and approximately 4 $g\,l^{-1}$ of an anion-active wetting agent. The pH of the bath is between 3.6 and 4, and the bath is operated at a temperature between 50 °C and 60 °C.

Electroplating begins on the exposed substrate areas of a lithographically structured resist–substrate sandwich. The substrate is generally covered by an oxidized titanium adhesion layer. The layer is the result of a compromise between a good resist bonding interface and an ideal seeding layer for electroplating. Recent findings recommend the use of solid gold or copper as a seeding layer, but adhesion onto these layers is not good enough.

Titanium is oxidized under the influence of a hydrogen peroxide solution, which forms a ~40-nm-thick nonstoichiometric oxide layer, considerably different from the naturally occurring oxide layer that is only 3 nm thick. It is experimentally observed that the oxide layer shows excellent adhesion to PMMA and is sufficiently conductive to allow electroplating to initiate. The titanium surface is microscopically rough and contains microchannels in which the resist molecules can anchor itself. As a result, both the metal and plastic regions adhere excellently to the substrate (Bacher *et al.* 1992).

In order to attain the highest possible rate of film growth, a relatively high nickel concentration is used in the electrolyte bath. Through this choice, structures with small cross sections are filled at a rate comparable to those with larger cross sections. The alternative would be to use leveling agents. This technique is, however, strongly geometry-dependent, as metal ion concentration levels must exactly match the particular transport situation as dictated by local feature layout, and hence avoided.

The height uniformity of structures of different cross section depends strongly on the electrolyte flow conditions also (Bade *et al.* 1996). Material transport differences are minimized if the electroplating current density is kept small when compared to that dictated by the diffusion-limited flow thickness. As a result, diffusion of ions only takes place as material transport and convective effects are minimized. This is fulfilled for a Reynolds number less than or equal to 2.

Electroplating should proceed in a uniform manner that ensures that uniformity in thickness and quality is achieved not only across the micrometer dimensions of a few microstructures, but also across the entire substrate surface of a few centimeters diameter. A major cause for inhomogeneity in electroplating is a nonuniform current field distribution across the substrate. One way to compensate for

the nonuniform distribution is to use electrical field blends made of dielectric material between the substrate and electrodes.

An anion-active wetting agent present in the electrolyte supports the penetration of electrolyte into the deep and narrow channels of the PMMA resist so that electroplating initiates uniformly across the metallic base plate. Typically, the influence of the wetting agent on the value of the surface tension of the electrolyte saturates at higher concentrations of agent. The optimum value of wetting agent concentration was determined by measuring the contact angle between the PMMA and the electrolyte. For nickel sulfamate electrolytes at 50 °C, an optimum concentration of 0.5% was determined. At this dosage level, the surface tension drops from the original 75 to 25 mN m^{-1}. The contact angle drops from about 75° to approximately 5° after 10 min of wetting time, which is considered sufficient for the purpose. Penetration is additionally supported by evacuating the microstructure plate.

Galvanic deposition of nickel is typically carried out in the range of 1–10 A dm^{-2}, which results in a film growth speed of 12–120 µm h^{-1}. The current density determines the amount of intrinsic film stress, which should be kept below 20 MPa. If a higher stress occurs, large area structures will bend in response and could even detach from the base plate during plating.

Inner stresses typically vary across the thickness of the deposited film (Harsch *et al.* 1988). **Figure 32** illustrates the influence of layer thickness and current density on internal stresses. With increasing current density, the internal film stress changes from compressive to tensile. In this example, it was possible to deposit stress-free layers with thickness above 50 µm by using 5 A dm^{-3}. The concentration of the wetting agent also has an influence on the level of film stress.

When using nickel sulfamate electrolytes, the internal stresses of deposited films can be also controlled through the temperature of the bath. Higher temperatures result in lower internal stresses. It should be noted, however, that high electrolyte temperatures lead to large evaporation rates and hence reduced exposure times of the electrolyte.

The level of the current density also influences the Vickers hardness value of the resulting structure. **Figure 33** shows that lower deposition current densities result in harder metal parts. For example, the peak hardness of 350 Vickers was achieved at a current density of 1 A dm^{-3}. At higher current densities, the hardness first drops rapidly, then more gradually falls away until reaching a measured value of 200.

The hardness variation is due to the simultaneous deposition of hydrogen, which increases with increased current density. One side effect is that released hydrogen bubbles can settle on the metal structure, thereby screening the surface from the electrolyte and forming pores in the structure. These pores show up only when the electrolyte contains impurities, which in turn serve as nucleation sites for bubble formation. Impurities can be dust from the environment, anode deposits, nickel hydroxide, components of the wetting agent and its

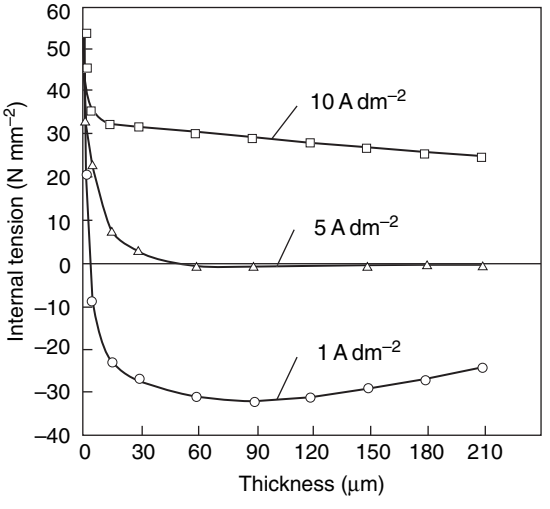


Figure 32 Intrinsic stress on electroplated nickel layer as a function of the layer thickness for different current densities. (Source: Menz W, Mohr J, Paul O 2001 Microsystem Technology, Wiley-VCH.)

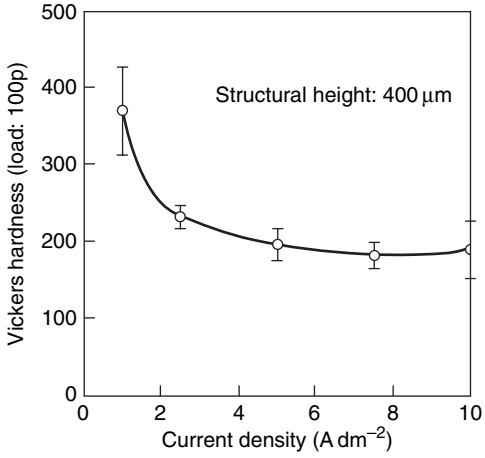


Figure 33 Vickers hardness of electroplated nickel layer as a function of the current density. (Source: Menz W, Mohr J, Paul O 2001 Microsystem Technology, Wiley-VCH.)

decomposition products, and other solid particles. Solid impurities should be continuously removed from the electrolyte through proper circulation through a filter of 0.2 μm pore size.

Defects are caused by an organic decomposition product of the wetting agent in the electrolyte during the course of the electroplating process. By exceeding a critical concentration, an increase in pore formation is observed. The organic impurities are eliminated from the electrolyte via active carbon purification. This has to be carried out regularly, with intervals dependent on the bath load. In addition, over a period of about 14 days, no defects were observed in the electroplated samples while plating with current densities between 1 and 10 A dm^{-3}. However, after this period, the defect rate increases drastically. At this point, the bath is actively carbon purified, which also completely removes the wetting agent. After the addition of fresh wetting agent, the bath is available again for defect-free plating.

Based on the operating parameters presented above, it is possible to completely fill narrow and deep channels in the resist with high-quality metal via electroforming. Surface features of the plastic mold, which lie in the submicrometer range, are copied with high precision. **Figure 34** shows a nickel honeycomb structure that was electroformed as detailed above. The height of the structure is 180 μm, with wall thicknesses of 8 μm. The small nonuniformities visible in the structure are already present in the mask in the form of ridges. As can be seen from the illustration, the resulting top surface of the structure is smooth. The smoothness depends strongly on the roughness of the substrate and the thickness of the deposited layer. In general, for layers above 100 μm we can expect a roughness (R_a) of the top surface at a value somewhat below 1 μm. If required, the roughness can be removed mechanically via polishing, lapping, or milling.

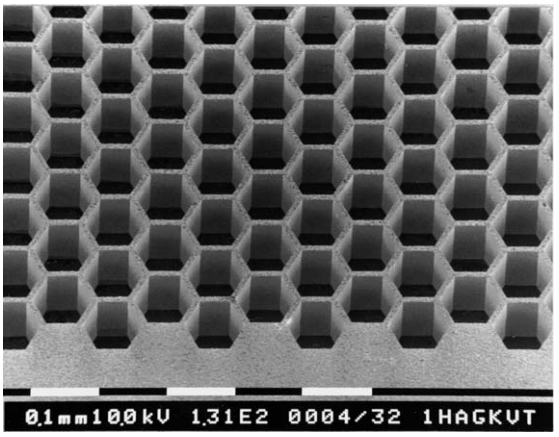


Figure 34 Honeycomb structure electroplated in nickel; the wall thickness is 8 μm, the height of the structure is 180 μm.

1.11.2.7.2 Electrodeposition of further metals and alloys

In principle, LIGA can be used to produce microstructures from any metal or alloy that can be electrodeposited. However, for process reliability, a number of electrolyte requirements need to be fulfilled, including those on microscatterability, wetting, inner stress of the resulting deposited thin film, and the level of gas formation during deposition. Only a few metals can meet these requirements.

Apart from X-ray masks, gold microstructures of several hundreds of micrometers in height have been successfully microstructured. Whereas either cyanide or sulfite gold electrolytes can be used, the sulfitic baths show several advantages as discussed above.

Copper can be plated as microstructures from either sulfamate or fluoroborate electrolytes. Sulfamate electrolytes yield ductile and practically stress-free level layers with high current yield. Organic bath additives are required, though. Because they are difficult to detect, it complicates their regulation in the bath and hence makes the process difficult to control. In addition, copper sulfamate baths are sensitive to impurities. Fluroborate baths also result in high current yield and low film stress, yet do not require any additives and are far less sensitive to impurities. Fluoroborate baths are, however, highly corrosive and the deposited layers are softer on the order of 120 Vickers. When used in conjunction with reactive substrates, a precopper plating layer is necessary. Both baths have been used to successfully produce LIGA microstructures.

Various investigators have produced LIGA structures from metal alloys. A nickel–cobalt alloy was successfully plated from a modified nickel sulfamate electrolyte (Eicher *et al.* 1992). Because the metal obtained is much harder than pure nickel, it is of interest for the manufacturing of mold inserts. Nickel–iron alloys such as permalloy, a mixture containing 80% nickel by weight, are weakly magnetic and feature both a high saturation field strength and a low coercive field strength. These

characteristics are important for the manufacturing of electromagnet cores and hence feature in the design of microactuators.

It is particularly difficult to homogeneously deposit the alloy into microstructures with specified constituents from the appropriate electrolytes. Spatially varying parameters such as the bath temperature and the current density directly influence the local composition of the film. The iron deposition is controlled by diffusion at high current densities. This can lead to iron depletion at diffusion boundary layers and hence to a reduction of iron in the deposited alloy. At lower current densities, the material composition is not influenced by material transport, as ions are replenished sufficiently fast. Nevertheless, the diffusion layer thickness has a strong influence on the material transport, which leads to problems in high aspect ratio cavities, where the diffusion layer thickness can easily reach the same value as the structure's height. Hence, current densities are quite a bit lower than those used for single metal deposition, in the range of only a few amperes per square centimeter (Thommes *et al.* 1995). As a rule of thumb, the highest aspect ratio cavity in a structure will provide the limit on the applicable current density. Based on this rule, uniform composition nickel/iron microstructures have been produced with greatly varying aspect ratio.

The measured magnetic properties of galvanically grown permalloy thin films are comparable to those measured from fused alloys, with the magnetic saturation of 1.1 T being only 5% below the bulk material's value. The difference is accounted for by the mixed crystal structure of the micropart. Whereas the 80–20 alloy showed no difference in properties when measured parallel to and perpendicular to the growth direction, this was not the case for pure nickel and the 50–50 alloy. Anisotropies can be avoided by recrystallization through tempering of the sample. The magnetic saturation can be increased further to about 1.4 T by increasing the iron content to 55%. Beyond this concentration, the saturation drops again. The reason for this change is the transition from face-centered cubic awaruite to body-centered cubic kamacite (Abel *et al.* 1996).

1.11.2.8 Molding

Molding, finally, is the cost-effective process part of the LIGA technology. Molding stands for the innumerous replication of a work and cost-intensive master template whose fabrication starts with the X-ray mask and ends with an electroplated mold insert. This section describes the mold insert fabrication, and then reports very briefly on the applied molding procedures. A detailed description of the fabrication of LIGA mold inserts can be found in (Guttmann *et al.* 2005).

1.11.2.8.1 Mold insert fabrication

Ten-millimeter-thick copper substrates are used instead of silicon or ceramic wafers for the production of LIGA mold inserts. This is a tribute to the electroplating process, which in case of a mold insert does not stop before the metal reaches the upper rim of the PMMA. In order to obtain a rigid and stable metal block with the microstructure incorporated inside, one allows the metal to overgrow the complete PMMA until it forms a closed layer on top of the substrate. The metal is grown to a thickness of at least 5 mm in total. Because of the internal stress that cannot be completely avoided, the metal block tends to bend, and the substrate follows accordingly. The thick copper substrate helps to keep this bending to a minimum. In addition, to reduce the internal stress, the current density has to be kept low even if the metal exceeds the PMMA and has already overgrown the critical structures. Typically, nickel is used for the mold insert, but nickel alloys are also becoming more and more attractive because of better mechanical behavior. **Figure 35** shows such an overgrown nickel block on a copper substrate.

In order to give the mold insert its final shape, the outer contour is defined by wire erosion, as can be seen from **Figure 36**. The first cut planarizes the rear side of the nickel block. Then a step profile is cut into the side of the nickel block. This step profile is

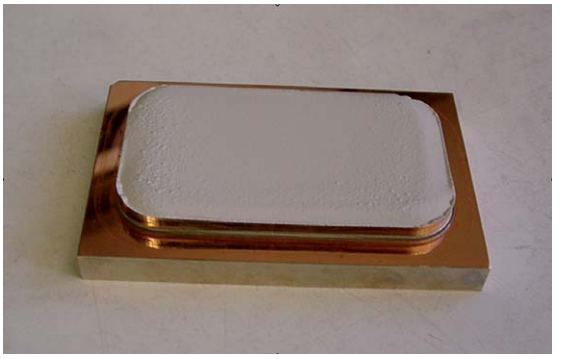


Figure 35 Photograph of a copper substrate with electroplated nickel block on top of it. The polymethylmethacrylate (PMMA) microstructures are completed covered with metal.

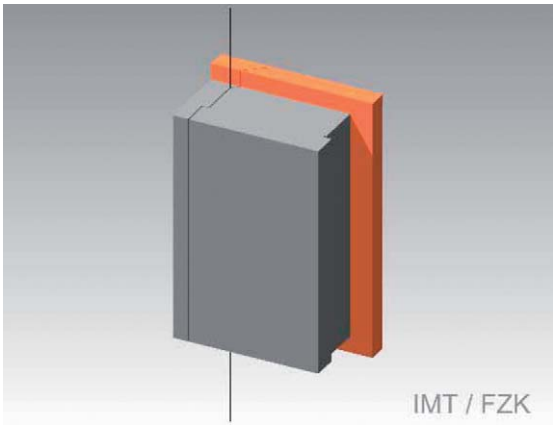


Figure 36 Finishing the outer contour of mold insert by wire erosion.

mandatory for the fixation of the mold insert in a molding machine. Finally, to release the nickel block from the substrate the wire cuts in parallel to the substrate in a depth of 200 µm through the copper. The residual copper layer can be easily wet etched without attacking the nickel.

The surface quality of the front side of the mold corresponds to the surface quality of the original substrate. No surface finishing is needed in the case of sufficient surface quality of the substrate. For optical applications, however, before the removal of the resist, it is possible to process the front of the matrix to the required roughness by milling, lapping, or polishing. After resist stripping, the mold insert features the original PMMA pattern as hollow cavities that can be filled again and again with any polymer that is suitable for replication. **Figure 37** shows an example of a mold insert for the fabrication of microspectrometers (see Section .11.3.2.5).

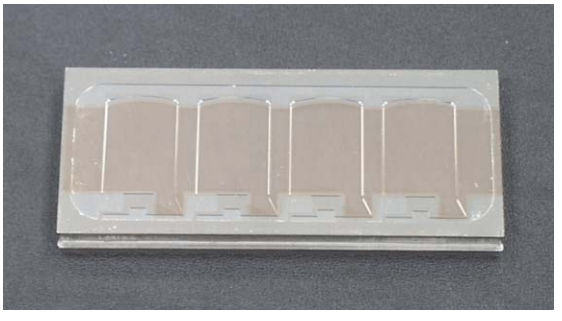


Figure 37 Mold insert featuring the hollow cavities of a microspectrometer pattern.

1.11.2.8.2 Plastic molding in the LIGA process

The mold inserts can be used for reaction and injection molding as well as for hot embossing. They can easily withstand temperatures up to 150 °C and pressures of up to 10 MPa. Even after thousands of molding cycles, the microstructures do not show obvious wear or degradation. Nevertheless, the replication of LIGA microstructures provides some specific challenges that are related especially to the high aspect ratio of the structures. First, the complete filling of the insert with the polymer is a critical process. To fill even the smallest structure details in the depth of the mold, which could be in dimension of submicrometers, the polymer must have a high filling ability, which means a low viscosity. Also during the cooling phase, lunkers and cavities can be formed, which cannot be compensated for later on. Even more difficult is the demolding procedure. On mold release, any tilt of the insert must be avoided, i.e., the mold release should be carried out extremely parallel so that the microstructure will not be deformed or damaged. Finally, measures must be taken to avoid or counteract volume changes due to the shrinkage of the plastic on hardening with process control.

1.11.2.9 Process Variations

In order to cover the widest possible spectrum of applications, the standard LIGA process is extended by numerous process variations. The most important of these are discussed in the following sections. Others will be listed and referenced to.

1.11.2.9.1 Sacrificial layer technology

Micromechanical sensors or actuators often require stationary microstructure parts that remain fixed to substrate in direct neighborhood to moveable microstructures that need to be released from it. In many cases, the movable and the stationary microstructures are mechanically linked so that they form one piece, which makes mounting of separate parts impossible. In addition, many devices need an electromechanical function, i.e., especially the stationary parts have to be electrically connected. These constraints apply to, for example, acceleration sensors, gyrometers, linear actuators, resonators, and similar structures.

In order to overcome these limitations, a sacrificial layer technique that allows for electrical connection on the one hand and partial release of the microstructures from the substrate on the other

hand was developed (Mohr *et al.* 1990) and first applied to LIGA acceleration sensors (Strohrmann *et al.* 1994), microturbines (Himmelhaus *et al.* 1992), and electrostatic motors (Wallrabe *et al.* 1994). **Figures 38–40** explain the process idea. The process starts on an electrically nonconducting substrate, e.g., on a silicon wafer equipped with an isolation layer or on an alumina ceramic substrate. A multifunctional chromium–gold layer is evaporated onto the substrate; during the process it ensures adhesion of the resist and electroplated metal to the substrate, and it serves as a starting layer for the electrodeposition process. Later, it provides electrical connection for the final stationary microstructures. To ensure electrical insulation of the microstructures from each other, the chromium–gold layer is structured by optical lithography and wet etching processes forming contact pads and conductive lines.

In a second step, titanium is sputtered onto the substrate and is also patterned by optical lithography and wet etching. It plays the role of the sacrificial layer, which has to remain underneath the movable structures during the process, and which will be sacrificed at the end in order to release the structures from the substrate.

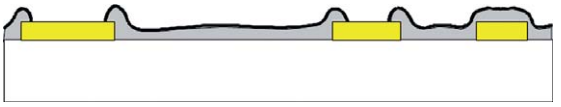


Figure 38 Substrate with patterned chromium/gold tracks for electrical connection and patterned titanium layer as sacrificial layer. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

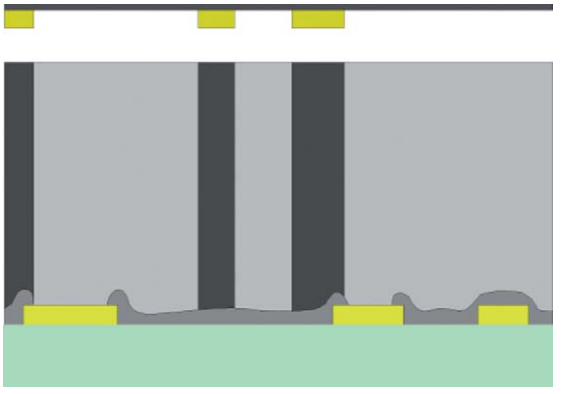


Figure 39 Substrate covered with polymethylmethacrylate (PMMA), X-ray exposure aligned to prepatterned substrate. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

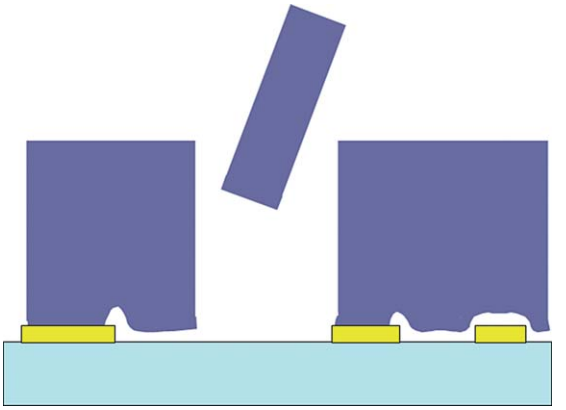


Figure 40 Filling of developed structures with metal and selective etching of titanium in order to release structures from substrate. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

Titanium (Kunz *et al.* 2000) has proven to be a suitable sacrificial layer because it fulfills the following requirements:

- Good adhesion of the resist used in X-ray lithography
- Good initiation and good adhesion of the subsequent electroplating
- Good selective etchability without attacking the other used materials (Cr, Au, Ni, Cu, Al_2O_3, Si)
- Fast underetching underneath structures with large areas

The titanium layer should be thick enough so that the movable structures have a sufficiently large distance to the substrate and are not clamped by etchant residues or impurities. With increasing thickness, however, the precision of patterning with wet etching decreases. A titanium thickness of 3–5 μm is a good compromise between these opposing requirements.

The two masks, the copper–gold and the titanium, that are used subsequently for the optical lithography of the two layers have to ensure an overlap of the etched patterns in order to provide an electrical shortcut among all structures on the substrate for the electroplating during the following LIGA process (**Figure 38**). For X-ray lithography, the substrate is covered in a typical manner with the PMMA. The exposure is performed well aligned to the prepatterned metal layers (**Figure 39**) with the help of

alignment marks on the substrate and on the X-ray mask as described earlier (see Section 1.11.2.2.5).

After electroplating and stripping, the titanium sacrificial layer is selectively etched using hydrofluoric acid, which is shown to be particularly suitable. By undercutting the microstructures that were placed on top of the titanium, they are released from the substrate and become freely movable, while the other parts of the metallic microstructure above the copper–gold layer remain well anchored on the substrate (**Figure 40**). After dicing by laser cutting from the rear side, the electrical connection is made by wire bonding on the prepatterned layer by bond pads and conducting lines.

1.11.2.9.2 Three-dimensional structuring

In principle, the standard LIGA process, whereby the structuring is carried out by a shadow projection, allows only the production of structures with a constant structure height and perpendicular walls. However, many structures require a variation in geometry in the third dimension. This can be achieved in simple forms by the structuring of different levels (Müller *et al.* 1996, Wallrabe *et al.* 2002) (stepped structures), by tilting of masks and substrates relative to the X-ray beam (Feiertag *et al.* 1997) (inclined structures), by careful exposure and melting processes (Göttert *et al.* 1995) (structures with spherical surfaces), or by effective use of secondary radiation (Göttert *et al.* 1995) (conical structures). All these processes are not be described here in detail, and we refer to the mentioned literature.

Another approach for the fabrication of almost fully 3D LIGA structures was pursued at the Ritsumeikan University, Kusatsu, Japan (Tabata *et al.* 2002). During the X-ray exposure, the mask can be moved relative to the substrate in the horizontal and vertical directions, the substrate can be rotated and tilted toward the synchrotron beam, and even a multimask process can be run. The multi-degree-of-freedom X-ray scanner is schematically shown in **Figure 41**.

By combining the exposure features in a clever way, not only inclined structures can be produced but also gray-tone lithography can be performed, resulting in microstructures that have nonconstant lateral dimensions across their thickness. Upon movement of the mask a continuous lateral dose profile can be deposited in the resists in contrast to a stepwise profile as was assumed in this chapter so far. The result becomes clear from **Figure 42**. A simple line pattern can lead to purely inclined lines when the mask and the substrate are just tilted during the exposure. Combined with a lateral mask movement, the dose deposition and the penetration depth

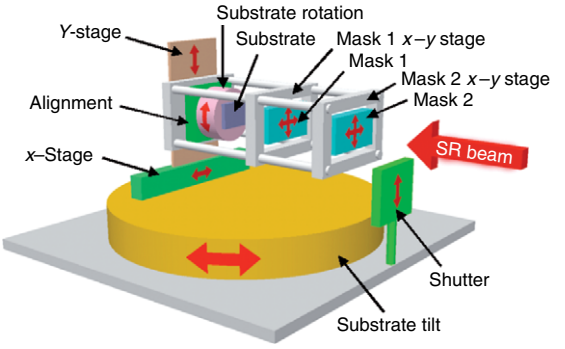


Figure 41 Multi-degree-of-freedom X-ray scanner. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

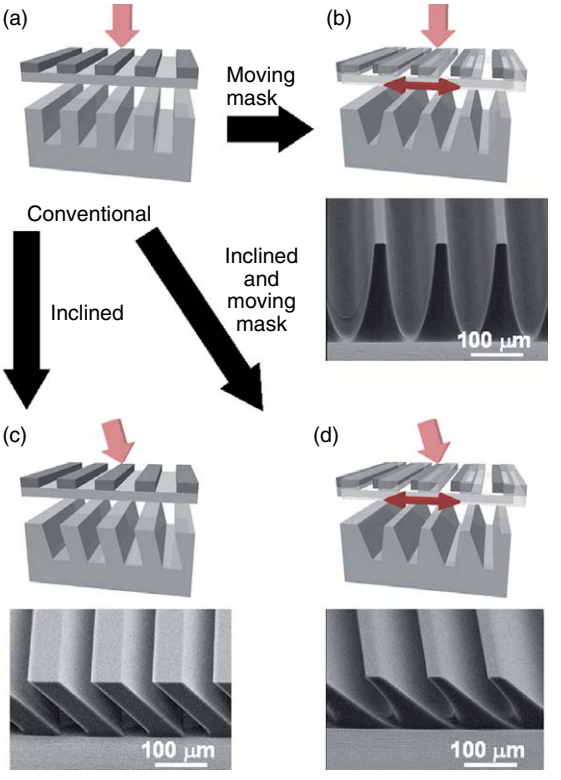


Figure 42 Principle of X-ray gray-tone lithography: upon moving the mask above the resist a continuous dose profile can be deposited. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

vary across the line pattern. After development, this results in a linewidth, which is smaller at the top than at the bottom and which, in addition, is tilted versus the normal axis.

Due to the inhomogeneous X-ray dose deposition in the resist, special care must be taken with the dose profile and the development of the structures. For a better understanding of the process, dedicated simulation tools have been developed that allow an accurate prediction of the process parameters and the feature geometry (Hafizovic *et al.* 2003). **Figure 43** shows as an example of a predicted resist profile at different stages of its development. The exposure was done in two simple steps. First, the overall rectangle was exposed with a dose, which is not sufficient for complete development. Second, only the center part was exposed for a second time. The developed profile does not simply show two levels with vertical sidewalls at the edges but with distinct slopes at the border. These slopes are related to a decreasing development rate with a decreasing deposited dose in the depth from the first exposure.

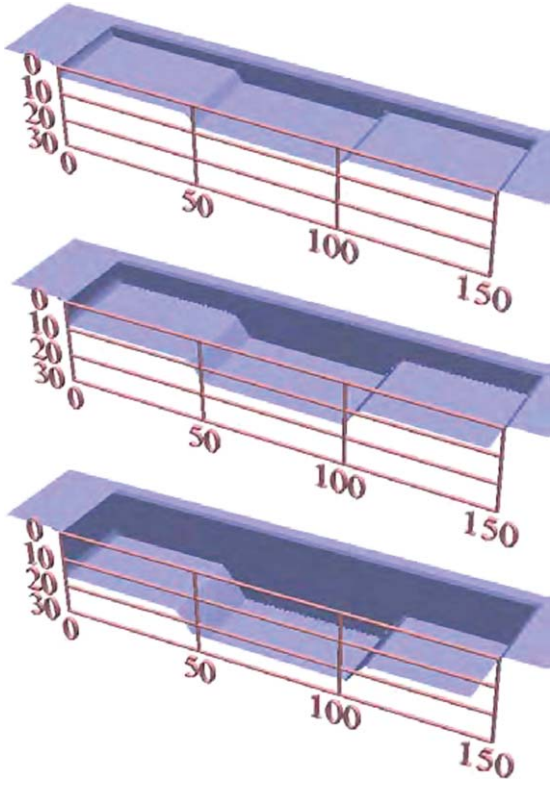


Figure 43 Resist profile at three different development times after a two-step exposure: *x*- and *y*-axes both represent dimensions in micrometers.

1.11.3 LIGA Technology Applications

1.11.3.1 Fields of Application

LIGA structures are typically from polymer or from metal. The commonly used polymer is PMMA because it can be patterned directly by X-ray lithography, and it is also very well suited for molding. Because PMMA also has a very good optical transparency in the visible light range, it is a perfect material – and LIGA a perfect technology – for applications in micro-optics. On the other hand, microstructures from metal are electrically conductive and therefore useful for actuators, as, for example, for electrostatic actuators, but more than that for electromagnetic actuators because soft magnetic FeNi alloys can be easily deposited by electroplating. In addition, metallic microstructures provide a good mechanical stability and can be used for most classical mechanical applications such as for gear tracks. Hence, this part of the chapter that reports on the applications of LIGA technology has four subsections: the first subsection covers the field of micro-optics made from polymer, which are either patterned by direct lithography or by molding. The second subsection describes some optical MEMS (MOEMS) that are made from metal because they incorporate electromechanical actuators, the third subsection gives an overview on mechanical applications also made from metal for stability reasons, and the fourth subsection reports on other applications that do not fit to the three previous subsections.

1.11.3.2 Optical Applications Made from Polymer

Striving for cost reduction through higher integration and on-going miniaturization, there are two basic technological approaches for micro-optical systems: a fully integrated fabrication sequence and a more modular approach. The choice as to which is best suited is a question of the available processes and of the respective yield, of the number of pieces to be fabricated, and most of all of the complexity of the system and the variety of the integrated functions (Ruzzu *et al.* 2003, Wallrabe and Mohr 2003). In the case of lower production numbers, and for the combination of several functional components, a modular design and fabrication are often the better choice. The processes are easier and safer, and one can work with pretested components that help to enhance the yield. This is exactly a case of designing

and manufacturing where LIGA perfectly matches the requirements. Therefore, many of the applications described in this subsection concentrate on micro-optical systems for telecommunication and sensors, which follow the modular approach, or at least a partially modular approach.

A micro-optical system is defined to be modular if its diverse functions are not fabricated in one common fully integrated fabrication sequence and/or if they are not integrated on one and the same substrate. In this section, they typically consist of a micro-optical bench that not only integrates some optical functions but also features alignment aids for other functional components. Examples for pure optical functions are light transfer, refraction and diffraction, imaging, filtering, beam splitting, or image superposition. Electrical or optoelectrical functions can be light generation, light detection or analysis, electronic circuitry, or current/voltage supply to optoelectrical components.

In the following, we start with the description of some elementary optical components such as cylindrical lenses, X-ray lenses, and multifiber connectors. We then introduce the principle of optical bench using a heterodyne receiver as an example. Afterward, we describe the LIGA spectrometer and a distance sensor.

1.11.3.2.1 Cylindrical lenses

Lenses are among the most elementary optical elements. In the macroscopic world, they are made of glass as well as of transparent polymer, in many cases of PMMA. Hence, lenses seem as if they could be easily miniaturized and manufactured by means of LIGA technology. This is true to a limited extent. Like all microstructure technologies, LIGA is a 2D process. Normally, the third dimension is simply the height or the thickness of the structures, whereas geometrical variations in the third dimension are normally restricted to a few different level heights. Only the moving mask technology of the Ritsumeikan University (Tabata *et al.* 2002) (see Section 1.11.2.9.2) allows for fairly free shapes in the third dimension. As a consequence, spherical lenses, as they are common in classical optics, cannot be fabricated. Cylindrical lenses, however, are very well suited for LIGA fabrication, especially for direct X-ray lithography. Because cylindrical lenses can focus the light only in one direction, which results in a focal line instead of a focal spot, it is necessary to combine two of them crosswise. The focal width can be easily determined by the lens radius, of course depending on the refractive index of the PMMA, which is around 1.5.

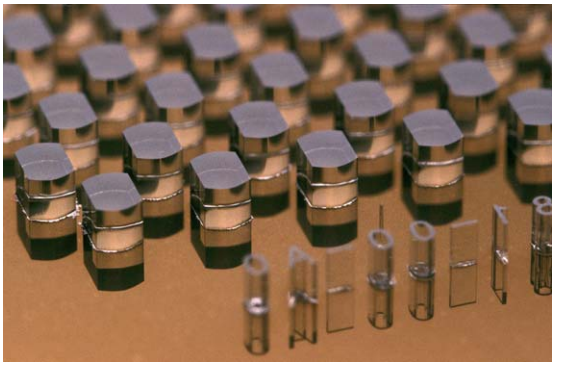


Figure 44 Array of cylindrical lenses with a focal length of approximately 10 mm. The polymethylmethacrylate (PMMA) thickness is 750 μm. The mirror image of the lenses is reflected from the substrate. An identification number together with its mirror image can be seen in the front part.

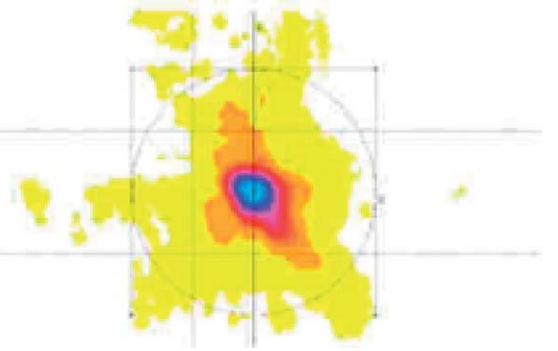


Figure 45 Spot of a horizontally oriented cylindrical lens crosses with a vertical focusing mirror; the full-width at half-maximum is 20 μm.

A photograph of a lens array that was patterned on a silicon substrate in 750-μm-thick PMMA layer by X-ray lithography is shown in **Figure 44**. The substrate is covered by an oxidized titanium layer, which serves as an adhesion layer for the PMMA and also as a sacrificial layer in order to release the lenses from the substrate. **Figure 45** shows a spot intensity distribution from such a single cylindrical lens combined with a focusing mirror (*cf.* Section 1.11.3.2.6). The center of the spot is fairly circular, and the full-width at half-maximum (FWHM) is about 20 μm. The focal length of the lens is 25 mm and the surface roughness is on the order of 20 nm.

Of course, lens arrays as shown in **Figure 44** can also be nicely replicated by molding. However, releasing them from the substrate directly after

X-ray lithography is much easier than after molding. In the first case, sacrificial layer etching is a straightforward and easy to perform process because no residual PMMA layer prevents the etchant to get in contact with the sacrificial layer. Molding such a lens array, however, always results in a residual polymer layer underneath the lenses, which connects the lenses at their bottom side. This residual layer has to be either dry-etched, which would attack the optical lens surface too and is therefore not recommended, or mechanically removed by back side polishing. Another advantage of sacrificial layer etching after pure X-ray lithography is the fact that one may stop etching shortly before the complete release of the lenses from the substrate. In this case, the lenses remain slightly attached to the substrate but maintains their position and, even more important, their orientation. This is of tremendous help for the subsequent lens assembly in any optical system because knowing the orientation enables the assembling person or the assembling machine to touch the lens where it is not risky, i.e., where it has no optically important surface. Hence, if one finally summarizes the effort to obtain individual high-quality lenses, the choice of using direct X-ray lithography and abandoning molding is often the better option.

1.11.3.2.2 Ultra-high X-ray lenses in SU-8

A very new optical application for LIGA cylindrical lenses arises from the development and fabrication of X-ray lenses. Research in X-ray optics is triggered by a large number of potential applications ranging from space-based X-ray telescopes and focusing or imaging systems for synchrotron radiation analysis to medical imaging. A recent approach, first proposed (Tomie 1996) and demonstrated (Snigirev *et al.* 1996) in the 1990s, is the application of lenses for X-rays. Such lenses function as their counterparts for the visible spectral range by refraction of electromagnetic radiation. For X-rays, however, materials have refractive indices of slightly less than 1, typically 0.99999. This small difference between air and materials is generally exploited for total reflection of X-rays at mirror surfaces when irradiated under small glancing angles. X-ray lenses are based on the fact that air (or vacuum) is the medium with a higher refractive index than the surrounding matter. A hole in a piece of metal therefore acts as a cylindrical lens (Snigirev *et al.* 1996). However, the hole shows a very large focal length due to the small differences in refractive indices. For achieving useful focal lengths with X-ray lenses, many individual lenses are arranged along the optical axis.

Modified LIGA techniques are employed to produce X-ray lenses with superior performance (Nazmov *et al.* 2004a). The lithographic process allows for producing several different lens systems in parallel, e.g., for different wavelength ranges and/or focal lengths. The lateral geometries can be chosen absolutely free for optimization of optical properties and correction of aberrations.

Two materials have been evaluated for the matrix in which the air lenses are embedded: SU-8 epoxy for lower photon energies at 5–50 keV (Nazmov *et al.* 2004b) and electroplated nickel for higher energies >40 keV (Nazmov *et al.* 2005). A major breakthrough for X-ray lenses was the use of negative tone SU-8 resist that can be patterned to high accuracy with DXL. The cross-linked areas are rather stable upon further irradiation. Hence, they can be directly used for X-ray lenses even when exposed to high-intensity X-ray beams.

Various different lens designs (**Figure 46**), for linear as well as double focusing, have been tested

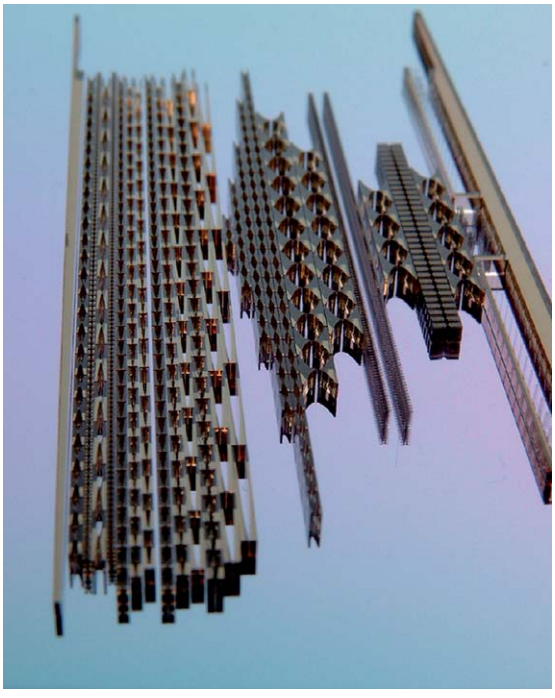


Figure 46 Twenty X-ray lens systems: the optical axis is along the individual rods. Lens details are displayed in the inset. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

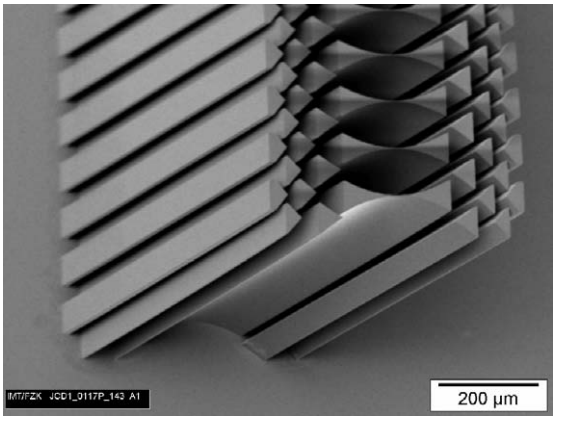


Figure 47 Part of a mosaic lens for 15–20 keV photon energies. Height of the structures is 1000 µm. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

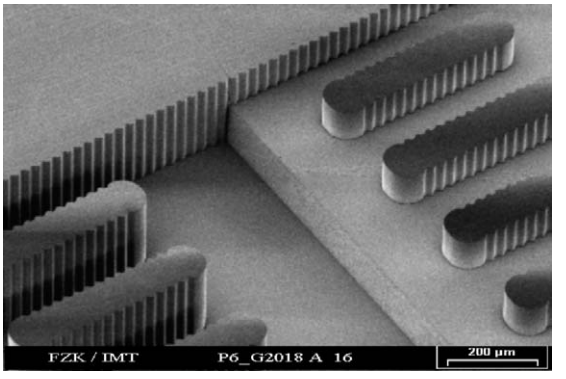


Figure 48 Scanning electron microscope (SEM) of the rippled alignment structures for optical fibers and detail of the fiber connector. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

at various synchrotron radiation facilities and the results have been compared to theoretical calculations. For the European Synchrotron Radiation Facility (ESRF) in Grenoble, dimensions of the focal spot are as small as 300–900 nm limited only by the demagnified X-ray source size, by diffraction, and by fabrication inaccuracies. **Figure 47** shows the respective lens system. The lenses are tilted with respect to the surface. A second lens system following this is tilted to the other side. By such an arrangement, a point focus can be achieved rather than a line focus. The gain in intensity as compared to a pinhole with the same dimensions was found to be approximately 5000.

1.11.3.2.3 Multifiber connectors

A device between a simple optical component and an optical bench is a multifiber connector that allows the parallel coupling of 16 optical fibers (Wallrabe *et al.* 2002). It consists of two plastic pieces out of PMMA: one for the alignment of fibers and guide pins with five rows of highly precise alignment structures and the other for their fixation and protection. Elastic ripples in the sidewalls of the alignment structures facilitate fiber and pin insertion, as can be seen in **Figure 48**, and make the connector insensitive to variations in fiber diameter and dust. The gap between the alignment structures decreases successively from the last row to the first row facing the opposite connector. This enables a very straightforward passive alignment and assembly of the fibers, without the need for micropositioning. Because the gap at the front face is 2 µm smaller than the fiber diameter, the fibers are softly clamped between the alignment structures, thus allowing an easy handling during the on-going assembly. UV curing adhesive is filled into the device through a hole in the cover piece and spreads into the coupler by capillary forces alone. The possibility of using fast UV-induced curing glue instead of epoxy resin that requires longer processing times in an oven is a positive trade-off of the fact that the couplers are made from transparent polymer. Finally, the front face is polished. The pins of both connectors to be coupled are inserted into a similar precise coupling adaptor as shown in **Figure 49**. If the pins damage their respective holes in the adaptor, e.g., due to frequent

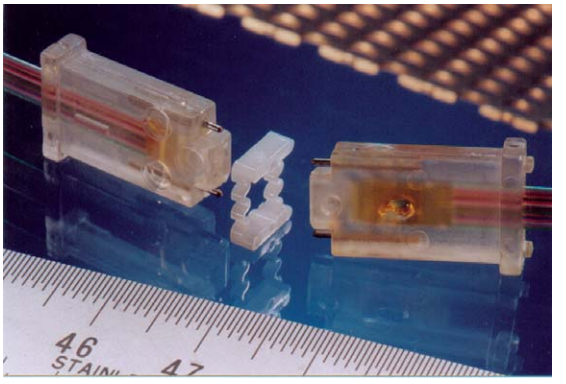


Figure 49 Fully assembled multifiber connector with common coupling adaptor. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

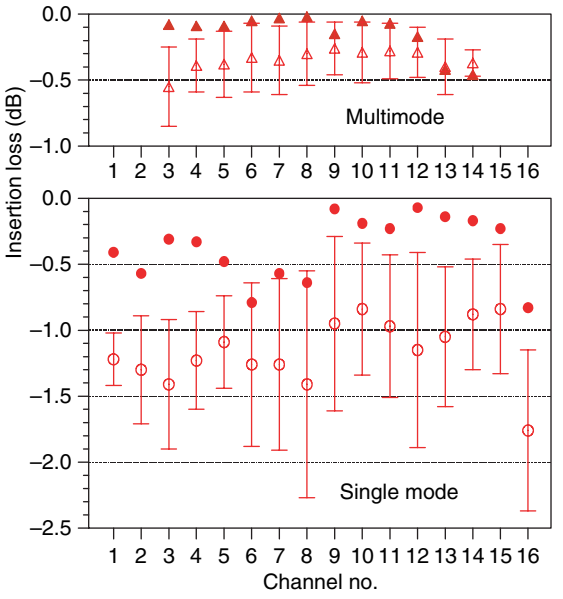


Figure 50 Insertion losses obtained with multifiber coupler for single- and multimode fibers. The open symbols show mean values and the closed symbols show the best connector pairing. (Source: Wallrabe U, Dittrich H, Friedsam G, Hanemann Th, Mohr J, Müller K, Piotter V, Ruther P, Schaller Th, Zißler W 2002 Micromolded easy-assembly multi fiber connector : RibCon. *Microsyst. Technol*. **8**, 83-87.)

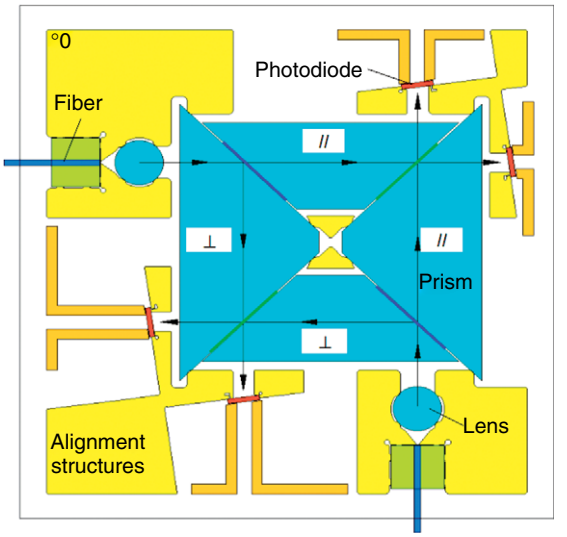


Figure 51 Schematic view of a heterodyne filter for infrared light with beam splitters and resuperposition on the diodes; the LIGA structures are exclusively used to position the optical components. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

reconnection, the adaptor can be replaced easily, whereas the fibers and connectors can be maintained in their assembled form.

For single-mode fibers, the coupling losses are in the range between 1 and 2 dB, on an average at 1.5 dB, and for multimode fibers the reach is 0.5 dB, as can be seen from **Figure 50**. The solid symbols indicate the best connection between the two connectors that was observed from a set of 50 connectors. The connectors are fabricated by means of plastic microinjection molding from PMMA or PC. The tool was fabricated by LIGA technology in combination with microprecision engineering as described in detail in (Wallrabe *et al.* 2002). The connectors are commercialized under the name RibCon® by Spinner Company (www.spinner.de).

1.11.3.2.4 Heterodyne receiver

The optical heterodyne receiver is a perfect demonstrator example for the optical bench approach because it combines various optical and optoelectrical functions. The heterodyne receiver works like a radio for light; hence, it is a wavelength filter for telecommunication. The needed functions are light input, beam collimation, beam splitting, polarization splitting, and finally detection by photodiodes. These totally different tasks call for a modular approach in order to avoid complex, difficult, and expensive processes. **Figure 51** shows the setup of a heterodyne receiver. In this case, two incoming light beams need to be split and superimposed (Ziegler *et al.* 1999). The signal beam and the beam from a local reference laser are coupled into the system using monomode fibers. The light is collimated by ball lenses and then split into the two polarization states. Upon reaching the next optical surface, the beam of each polarization state is again split by 50% and is simultaneously superimposed with the respective beam from the opposite light source. Each of the four final superimposed beams is detected by a photodiode. The system consists of a ceramic chip on which exclusive alignment structures out of polymer are patterned using direct X-ray lithography. The fibers, the ball lenses, the prisms for the polarization beam splitters and 50/50 beam splitters, and the diodes are separately manufactured components. All these additional components are assembled in a fully passive manner on the chip. They are simply pushed toward the alignment structures and subsequently fixed by UV curing glue. The electrical connection of the diodes is achieved through gold tracks, which are

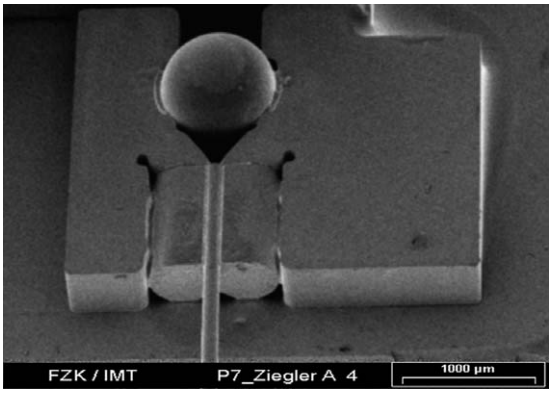


Figure 52 Scanning electron microscope (SEM) of an optical fiber on top of a fiber alignment mount in front of a ball lens; the parts are assembled into the LIGA positioning structures. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

prepatterned on the substrate by optical lithography and wet etching.

Another quite often-used concept is the fixation of the optical fibers. As the altitude of the optical axis is defined by the radius of the ball lens (here 450 µm), the fiber with a diameter of only 125 µm needs to be levered to the same height. This is done by a fiber mount, as shown in **Figure 52**. The fiber mounts are also LIGA devices; however, they were fabricated on a separate substrate with a suitable resist thickness (1 mm) in a vertical orientation, in analogy to the cylindrical lenses that was explained before. Afterward, the fiber mounts are released from the substrate by sacrificial layer etching, and they are assembled in a flipped orientation, i.e., with the axis parallel to the substrate.

From the superposed intensity pattern that was depicted with a camera at the position of one of the diodes, it follows that the assembly accuracy in all directions is on the order of 1 µm across the substrate of size 1 cm^2. This very good result proves the efficiency of the modular approach on the one hand and of the ease of passive assembly on the other hand.

1.11.3.2.5 Spectrometer

The LIGA-made microstructures of the heterodyne receiver have a purely mechanical function because they are used for alignment. In many other cases, they may also have a very distinct optical function, as, for example, in microspectrometers, one of which is illustrated in **Figure 53**. The LIGA spectrometer

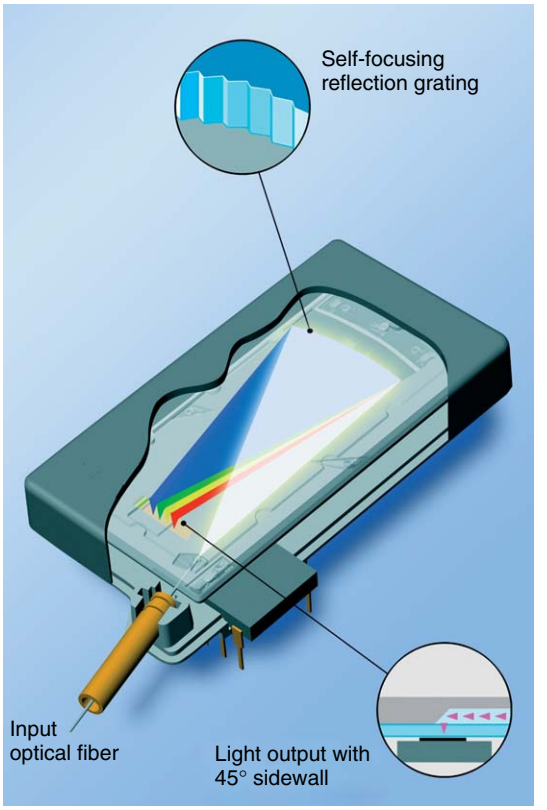


Figure 53 Principle of microspectrometer for visible light: the device is molded in polymethylmethacrylate (PMMA) or polycarbonate (PC) and coated with reflective layers.

was developed for visible and infrared light. The light is coupled into the device by an optical multimode fiber. In a former design, the light was guided using total internal reflection in a three-layer polymer waveguide with adapted refractive indices. Today's design features a hollow waveguide with reflective coating on the top and the bottom surface (Krippner *et al.* 2000). The in-coupled light spreads into the free space of the spectrometer toward a reflecting and self-focusing grating, which diffracts the light and thus separates it spatially into its constituent wavelengths. The grating is typically blazed to the first- or second-order refraction, which means that the simply reflected part of the light will also be reflected to the respective angle, thus enhancing the light intensity there tremendously. In addition, for the self-focusing, the grating is arranged on a Rowland circle, as becomes clear from **Figure 54**. This allows the light to be focused at least horizontally in order to strike just a minimum number of pixels on a photodiode array that is aligned on top of the waveguide. In order to deflect the light out of

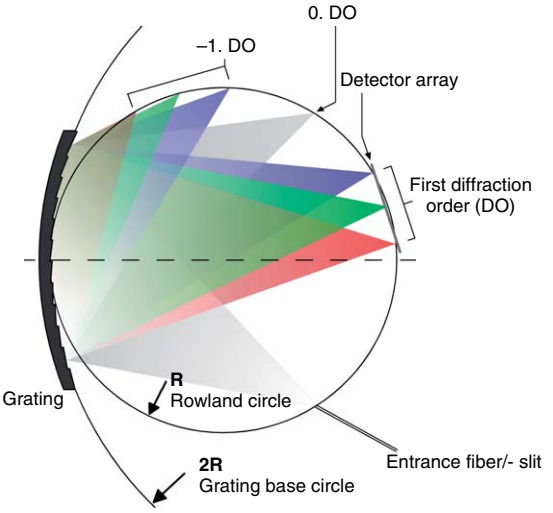


Figure 54 Arrangement of the grating on a Rowland circle leading to a horizontal self-focusing of the light.

Figure 55 BiliChek system for the detection of bilirubin (a jaundice indicator) in newborn babies. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

plane toward the diode array a 45°, mirror is patterned at the focal line. The overall size of a visible light spectrometer is about 20 mm × 13 mm, i.e., it is fairly large. The grating constant, however, is only 1.5 µm with a depth of the teeth of only 0.2 µm. In the case of infrared spectrometers, the grating constant and the overall size increase correspondingly. The spectral resolution is greater than 10 nm for visible light and greater than 15 nm at a wavelength of 1.7 µm. The spectral accuracy is 2 nm and the signal-to-noise ratio greater than 5000.

The spectrometer is produced by hot embossing or injection molding in PMMA or in PC using a nickel or an iron–nickel molding tool. After the molding step, a gold layer is sputtered onto the polymer components in order to make the optical surfaces reflective. The photodiode array and some electronics are assembled upside down on top of the spectrometer. A variety of different spectrometer designs for different application purposes are produced in volume by the company Böhringer Ingelheim microParts (www.boehringer-ingelheim.de/produkte/mikrosystemtechnik/Microtechnology/index.jsp). These modules form the key components of portable measurement systems for various applications. Three examples are given here: the BiliChek system (http://bilichek.respironics.com/), which is an instrument for the noninvasive measurement of the bilirubin concentration of newborn babies (**Figure 55**), the HemoNIR system (www.nirdiagnostics.com), which measures a number of blood parameters such as total-Hb, Met-Hb, and carboxy-Hb from a blood sample of less than 10 µl, and the IdentaColor system (www.identa.dk), which is a small portable tooth color measuring instrument allowing dentists to select the optimum color of the new ceramic insert.

1.11.3.2.6 Triangular distance sensor

Another application that relies on a head-over assembly of the optoelectrical devices is a microdistance sensor that works on the basis of the triangulation principle (Oka *et al.* 2001), i.e., the position of the spot on a sensor's detector is dependent on the distance to the object being measured. The distance sensor is divided into two functional units: a passive optical chip, again fabricated by LIGA technology, including curved mirrors, 45° mirrors, and alignment structures for cylindrical lenses, and an electro-optical chip with a laser diode, a photodiode, and a position-sensitive detector (PSD). Both chips are again mounted head-over to form an optoelectromechanical sensor system. For easy assembly, two microspheres of glass are used as position aligners of the LIGA and the silicon chip (**Figure 56**). First, the microspheres are placed into pyramidal grooves on the silicon base, thus centering themselves. Second, the optical chip with two respective cylindrical holes for the spheres is positioned head-over above the silicon chip and is carefully laid down. Starting from this moment, the assembly is a blind process because of the opaque substrates. The only requirement is that the cylinders are placed somewhere above the glass spheres, i.e., with a prealignment accuracy of several tens to several hundreds of micrometers. The upper chip is then

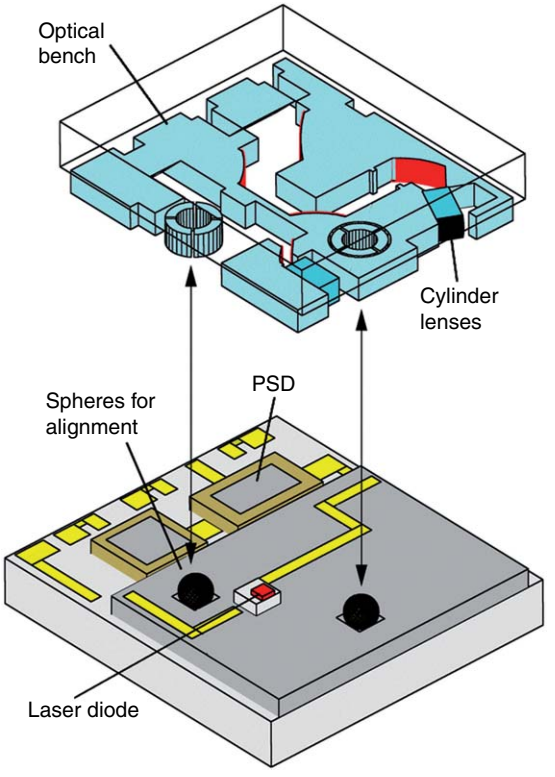


Figure 56 Two chips of a distance sensor are mounted head-over using glass balls as alignment aids. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

slightly pushed downward and aligns itself, while the cylinders slip above the spheres.

The optical components on the LIGA chip are made from PMMA either by direct lithography or by hot embossing and are coated with an evaporated gold layer in order to make the mirrors reflective. The two chips are shown in **Figure 57** before the head-over assembly is performed. Cylindrical lenses that were patterned on a separate substrate and released from it by sacrificial layer etching (*cf.* Section 1.11.3.2.1) are positioned in the illumination and detection light path of the optical bench after the gold evaporation. The light emitted from the edge-emitting laser diode is then focused by a lens and an additional curved mirror that is located directly in front of or behind the lens.

Due to the free lateral shape options of LIGA technology, the optical bench, especially the shape of the curved mirrors, is designed to allow for a linear output signal of the sensor. A 45° mirror at the end of the detection path guides the light toward the PSD,

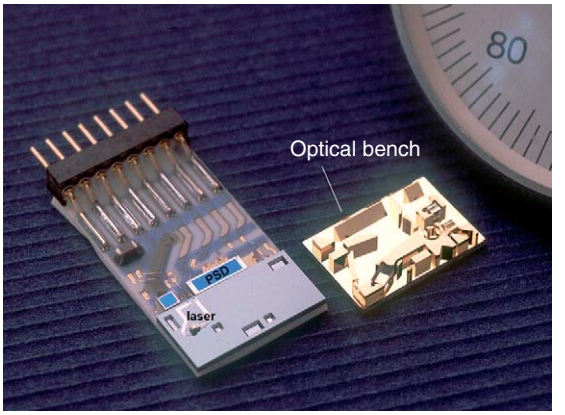


Figure 57 The LIGA-made optical bench and the electro-optical chip of a distance sensor before head-over assembly. PSD, position-sensitive detector.

which is a part of the electro-optical chip. The measured signal of the PSD is normalized and compared to a linear fit as shown in **Figure 58**. The linearity error that is well below 1% is also plotted in **Figure 58**.

1.11.3.3 Optical MEMS

In contrast to the optical systems presented so far, MOEMS, as they are sometimes called, typically combine optical function with a mechanical or an electromechanical function. In many cases, the latter is represented by an actuator in order to move mirrors or to control the optical path in any way that may be required. Generally speaking, one could postulate that mechanical function is provided by all kinds of alignment structures or positioning aids and by fixing and clamping structures. Electromechanical function is provided by actuators, heaters and coolers, and also by mechanical sensors. The electrical part of it mostly requires electrical conductivity of the devices. Hence, in contrast to the earlier part of the chapter, we will now deal with LIGA microstructures that are electroplated in metal.

1.11.3.3.1 Optical cross connector with rotating mirrors

A first example for this kind of LIGA microsystems is an optical cross connector. A fiber switch array combines a LIGA-fabricated micro-optical bench, as described above, with rotating mirrors that are arranged in a matrix scheme (Ruzzu *et al.* 2000). **Figure 59** shows the concept for an $N \times N$ switch. The light from the fibers is collimated onto the mirror surfaces and refocused onto the output fibers

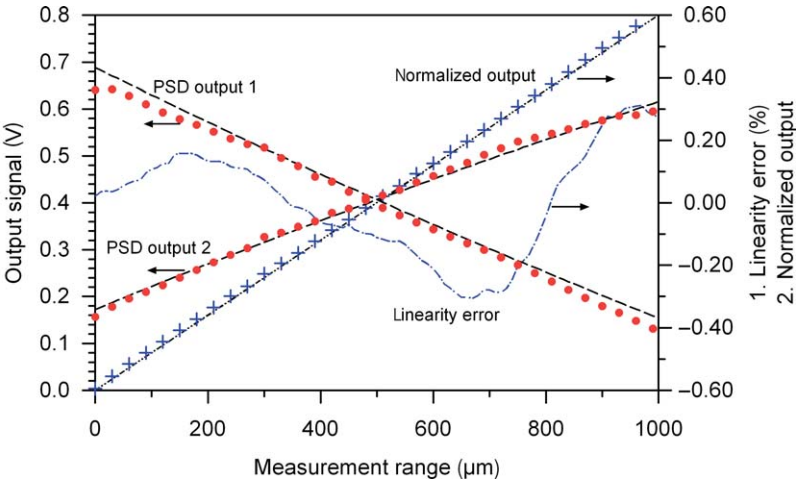


Figure 58 Normalized position-sensitive detector (PSD) output signal as a function of the distance of the measured object. The red line represents the measured results and the blue line a linear fit to the data. The right axis of the graph shows the linearity error of the sensor.

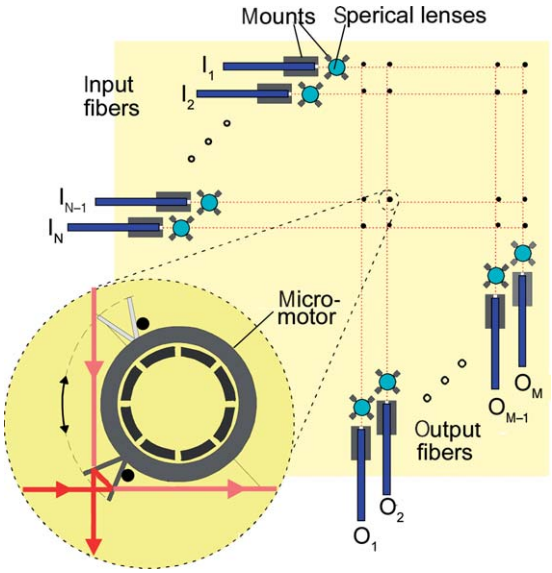


Figure 59 Schematic view of an optical cross connect; rotating mirrors at the beam junctions can be switched either to reflect the beam or to let it pass. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

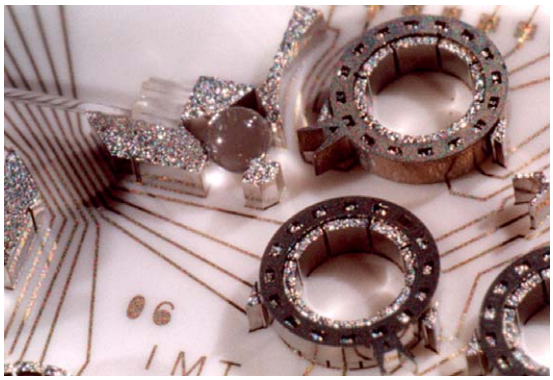


Figure 60 Optical bench with lens and fiber alignment structures and electrostatic micromotors made from nickel. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

by spherical glass lenses. Fibers and lenses are passively aligned with the help of fiber mounts, and the alignment stops in the optical bench. In order to achieve uniform insertion losses for all channels, the fibers are arranged in such a way that all optical path lengths are equal. The mirrors are attached to the outer side of the electrostatic micromotors that swivel the mirrors in and out of the beam path. Mechanical stops define the correct end position of the mirrors. The mirrors were designed as double mirrors in order to ensure a precise 90° reflection. The optical benches, the static motor parts, and the mirror stops are fabricated simultaneously using X-ray lithography and subsequent electroplating of nickel on a sacrificial layer. The rotors with the attached double mirrors are manufactured similarly on a separate substrate. They are released from it and are manually placed above a ring-shaped bearing surrounding the stators. The fiber mounts out of PMMA are patterned on a third substrate. **Figure 60** shows a close-up view of a fully assembled 2×2 switch matrix. On the left side, the fiber on the

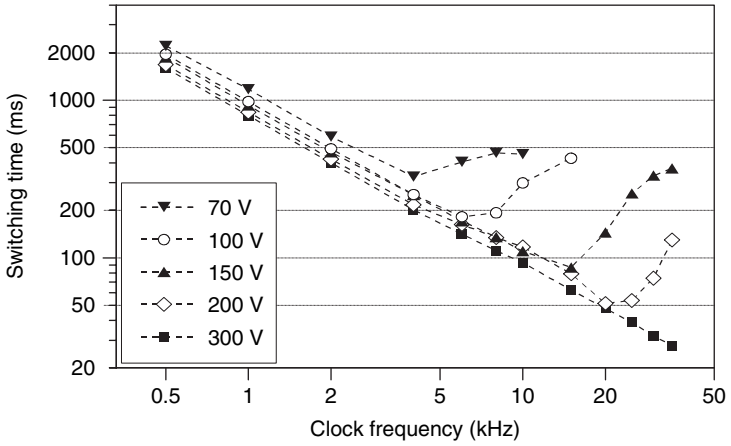


Figure 61 Switching time as a function of the clock frequency for various driving voltages. (Source: Ruzzu A, Haller D, Mohr J, Wallrabe U 2003 Optoelectromechanical Switch Array with Passively Aligned Free-Space Optical Components. *J. Lightwave Technol.* **21(3)**, 664–671.)

fiber mount and a lens can be distinguished. They are pushed toward the stop structures of the optical bench. The mirror in front of the lens is in the on position and the mirror below in the off position.

The electrostatic motors are designed as wobble motors. In contrast to other wobble motors as described (e.g., in Samper *et al.* 1999), the stator poles do not surround the rotor, but they are placed inside the ring-shaped rotor in order to allow the attachment and movement of the mirror (see **Figure 59**). The motor is actuated at voltages ranging from 70 to 300 V. **Figure 61** gives the switching time as a function of the clock frequency at which the voltage is applied successively to the stators. The lower the driving voltage the lesser the clock frequency at which the motor gets out of step. The shortest switching time was 30 ms at 300 V. An optical characterization of the system results in a crosstalk greater than 90 dB between the channels, in a minimum insertion loss of 3 dB with a repeatability of 0.5 dB.

1.11.3.3.2 Oscillating modulator for infrared light

Besides using electrostatic actuation, a unique feature of LIGA technology is to use electromagnetic forces. In this case, the structures are electroplated in permalloy, which is a soft magnetic alloy of 80% nickel and 20% iron. A first application example is a modulator, i.e., a chopper for infrared light and used for the suppression of noise in infrared spectrometers by using lock-in techniques (Krippner *et al.* 1999). The working principle becomes clear from the demonstrator, which is shown in **Figure 62**. Alignment structures for input and output fibers are patterned onto a ceramic substrate in parallel with a movable shutter, looking like a little hammer. The hammer is a part of a magnetic circuit, which is constituted by the coil in the upper part of the picture, and the yoke that guides the magnetic flux generated through the current in the coil toward the hammer. Due to the decrease in the resistance of the magnetic circuit the shutter is pulled backward between the pole shoes of the circuit, thus enhancing the metallic cross section for the magnetic field. Such an actuator is often called a variable reluctance actuator. The yoke for the coil was manufactured

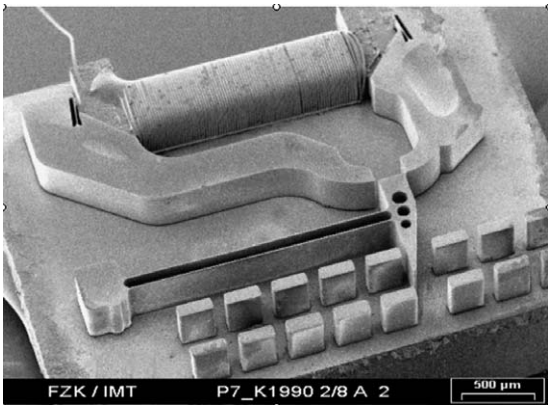


Figure 62 Electromagnetic actuator made from NiFe and operated as a chopper to modulate infrared light. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

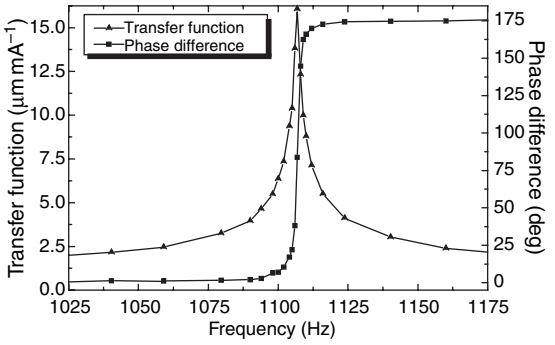


Figure 63 Normalized displacement of the shutter and phase shift as a function of excitation frequency.

on a separate substrate and released. The coil is wound on a turning machine and then assembled onto the substrate using alignment and clamping structures. A sinusoidal current generates a periodic force field and hence an oscillation of the shutter. Figure 63 shows the amplitude per milliampere current in the coil and the phase shift at an excitation close to the resonance frequency, which is about 1100 Hz.

For integration of the chopper in a spectrometer, the fiber alignment structures are skipped. The chopper is inserted head-over in the light input path of the spectrometer like in the assembly approach discussed earlier with distance sensors. This is facilitated by four stops on the edges of the yoke and the fixing block of the shutter, which were patterned simultaneously for this purpose. Figure 64 shows a close-up view of the very precise position of the shutter tip directly in front of the input fiber of a spectrometer in its alignment groove.

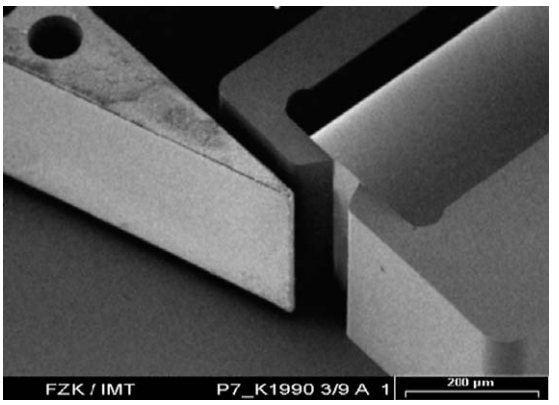


Figure 64 Close-up of Figure 62 with assembled fiber. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

1.11.3.3.3 Fourier transform infrared spectrometer for infrared light

Another example using a variable reluctance actuator is a Fourier transform infrared spectrometer (FTIR). The design is based on a Michelson interferometer that consists of an optical bench for the passive alignment of the optical components and an integrated actuator (Wallrabe *et al.* 2005). The optical bench and the actuator are again made of permalloy of height 380 µm. The dimensions of the complete system are 11.5 mm × 9.4 mm (Figure 65). The light to be analyzed is delivered to the system via an optical fiber. It is collimated with a fused silica ball lens of 650 µm diameter. At the beam splitter, the light is separated into two rays, one reflects off the fixed mirror and the other travels toward the movable mirror. Both rays are reflected and travel back to the beam splitter where they interfere with each other. The interference signal is recorded with an InGaAs PIN photodiode that is sensitive in the wavelength range of 850–1700 nm. For recording the signal, the position of the movable mirror has to be determined very precisely. Hence, monochromatic laser light is used. The laser light is also delivered to the spectrometer through an optical fiber and propagates in a second optical channel parallel to the white light to be analyzed.

Figure 66 shows the movable plunger of the variable reluctance actuator, which is attached to a set of four folded cantilever beams and which is surrounded by fixed core structures thus forming two magnetic circuits that are driven by two coils. Because variable reluctance motors deliver only pulling forces,

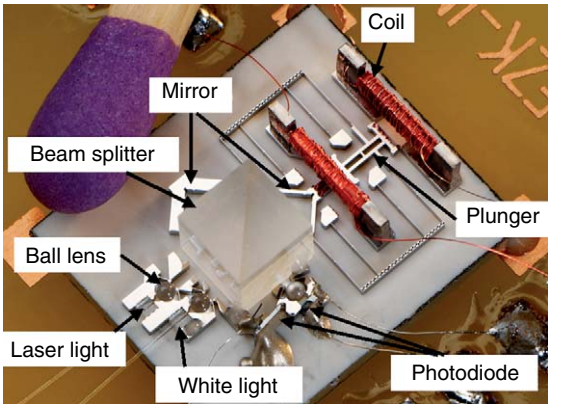


Figure 65 Three-dimensional view of spectrometer including optical bench (*left side*) and actuator (*right side*). (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

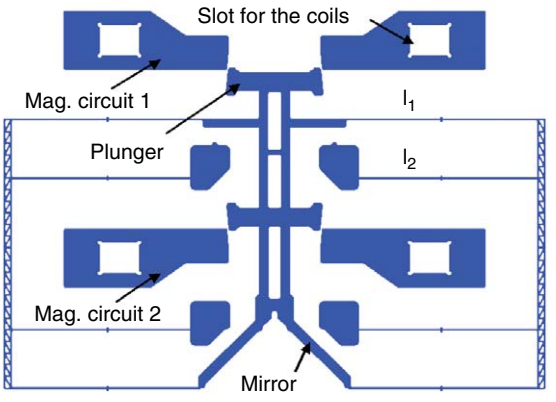


Figure 66 Schematic of actuator and magnetic circuits; the coils are placed into the respective slots. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

the displacement length was increased by using two coils, one for pulling the plunger toward the beam splitter and the other to pull it in the opposite direction. The separate coils are assembled into slots prepared in the fixed magnetic circuit.

For spectral analysis, the optical bench was equipped with all-optical components (see Figure 65), and the laser light at a wavelength of 1540 nm was delivered through a monomode fiber that ensures a good collimation of the beam. In this first experiment, only one coil was activated for which a displacement of the actuator of 54 µm was achieved leading to a spectral resolution of 25 nm. A measured white light spectrum is given in Figure 67. It shows a distinct minimum at a wavelength of 1.5 µm, which is related to the spectrum of the optical fiber used for the light input, as can also be seen from Figure 67.

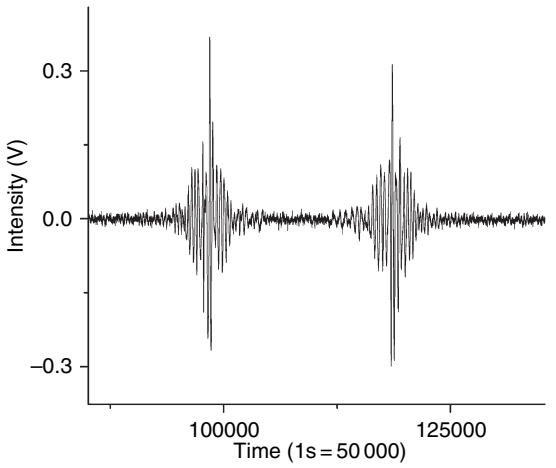


Figure 67 White light spectrum in comparison to absorption spectrum of optical fiber.

1.11.3.4 Mechanical Applications

Besides integration of electromechanical actuators in optical microsystems, another more and more upcoming field of application of LIGA technology is of pure mechanical nature. The accuracy of the X-ray lithography combined with the option of a copy of the pattern in a variety of metals represents an outstanding technology feature for mechanical applications. Currently, perhaps the most important application is concerned with gears. Three different approaches on LIGA-made gears are discussed in the following sections.

1.11.3.4.1 Micro Harmonic Drive®

Microgears are key elements in microdrive systems. They allow to reduce the speed and enhance the torque of miniature electromagnetic motors, which deliver typical output torques of several microNewton meters at speeds up to 50 000 rpm (www.faulhaber-group.com, www.maxonmotor.com). Only when combined with a gearbox the motors become applicable for many applications in microrobotics and automatization.

The Micro Harmonic Drive®, which is commercialized by Micromotion GmbH in Germany (www.mikrogetriebe.de), represents a high-precision microgear with extremely high reduction ratio. It is an assembly of a set of six separate gears in total (Slatter *et al.* 2006) that can be seen in Figure 68. The basic components are the so-called wave generator featuring a sun gear wheel and two planetary gear wheels. The sun gear is the one that will be connected to the driving unit, i.e., to the motor. The wave generator is placed in a set of three other gear wheels: the flexspline, the circular spline, and the dynamic spline. The dynamic spline serves as the torque output of the system. Hence, it is connected to the device that can be driven by the gear system, e.g., a microrobot. The flexspline features teeth at its inner and outer surface, whereas the others have teeth at only their inner surfaces. In addition, the flexspline has two teeth less than the others.

The working principle becomes clear from Figure 69. The overall diameter of the wave generator is slightly larger than the inner diameter of the flexspline gear, which is thin enough to allow for flexible deformation. Once the wave generator

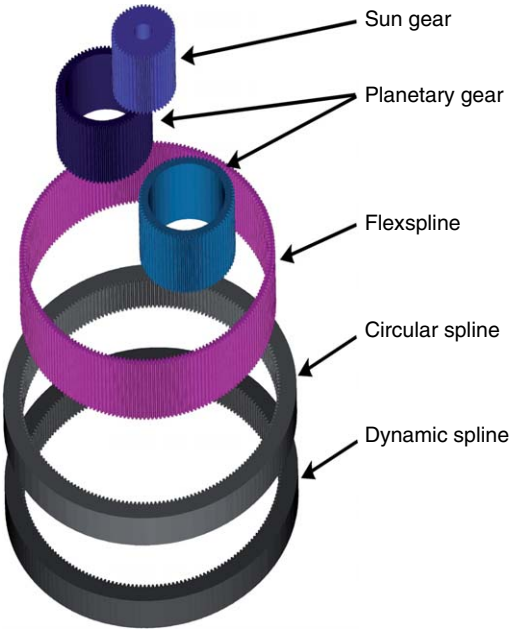


Figure 68 Gear components before assembly. (Source: Slatter R, Degen R, Burisch A 2006 Micro-mechatronic actuators for Desktop factory applications. *ISR/ROBOTIK 2006 - Joint conference on robotics*, Munich, Germany, May 2006. VDI-Bericht Nr. 1956, VDI Wissensforum IWB GmbH, Germany, pp. 137–150.)

is inserted into the flexspline, the flexspline is deformed to an elliptical shape. Thus, the outer teeth at the long axis of the ellipse are engaged with the respective teeth of the circular spline. With the rotation of the wave generator, the deformation and the tooth engagement travel along the circular spline. After a full rotation of the wave generator, the circular spline has moved by exactly the difference of the two teeth. The dynamic spline that is tightly connected to the circular spline moves correspondingly.

The large number of teeth on the outer gears compared to the travel of only two teeth during one rotation results in very large reduction ratios of the Micro Harmonic Drive system. Depending on the size of the system (8–10 mm diameter) reduction ratios between 160:1 and 1000:1 can be achieved, and the rated torque that can be loaded to the system varies from 7 to 24 mN m (Slatter *et al.* 2006). Due to the fact that the planetary wheels of the wave generator are also a bit deformable, the Micro Harmonic Drive exhibits excellent accuracy and almost zero backlash. Thus, the repeatability is 10 arcsec.

All gears are manufactured by X-ray lithography and subsequent electroplating of nickel–iron and final polishing of the surface. The NiFe has a fairly low Young's modulus of 135 kN mm^{-2} and a good fatigue behavior (Abel 1996). The lifetime is more than 2500 h at rated torques and speeds corresponding to many millions of cycles. **Figure 70** shows a section of a silicon wafer with flexspline and planetary gear wheels after electroplating, polishing, and stripping of the resist. Finally, they will be released from the substrate by sacrificial layer etching.

1.11.3.4.2 Cycloid gear system

A simple shadow casting by DXL results in patterns with high-quality vertical walls. However, the versatility requirements for current micropart applications include more complex 3D geometries and multilevel

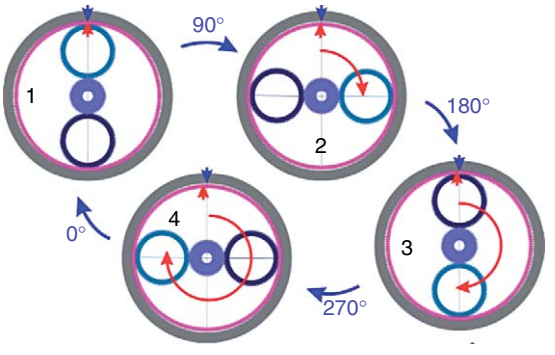


Figure 69 Operation principle of the Micro Harmonic Drive®. (Source: Slatter R, Degen R, Burisch A 2006 Micro-mechatronic actuators for Desktop factory applications. *ISR/ROBOTIK 2006 - Joint conference on robotics*, Munich, Germany, May 2006. VDI-Bericht Nr. 1956, VDI Wissensforum IWB GmbH, Germany, pp. 137–150.)

Figure 70 Flexspline and planetary gear wheels electroplated in NiFe on a Si wafer. (Source: Slatter R, Degen R, Burisch A 2006 Micro-mechatronic actuators for Desktop factory applications. *ISR/ROBOTIK 2006 - Joint conference on robotics*, Munich, Germany, May 2006. VDI-Bericht Nr. 1956, VDI Wissensforum IWB GmbH, Germany, pp. 137–150.)

configurations. One example of a 3D LIGA mechanical structure is another cycloid gearbox. A cycloid gear formation remains an advantage for miniaturization, because it does not need a planar arrangement of several gears. However, the layered formation of gear components requires an accurate planarization and height controllability of the microparts, for which sequential planar formation by the multiexposure LIGA technique can be quite fitting.

Figures 71 and **72** show a prototype of such a 2-mm-diameter microcycloid gear system (Hirata *et al.* 1999). The entire gear train consists of a casing and three vertically stacked disks and gears. Each part is composed of three different levels. The first level, 40 μm high, is fabricated by UV lithography, and the second and the third level, 195 and 250 μm high, respectively, are processed by aligned X-ray lithography. The alignment error between two X-ray lithography processed layers is within the ±5 μm range. The thickness of the layers is controlled by mechanical surface machining with an accuracy of ±3 μm for the UV lithography-processed structures and of ±10 μm for the DXL-processed structures.

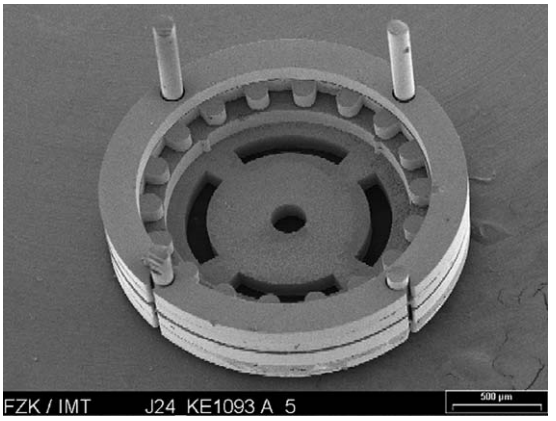


Figure 71 The casing of a microcycloid gear system fabricated by multiexposure X-ray lithography combined with UV lithography. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

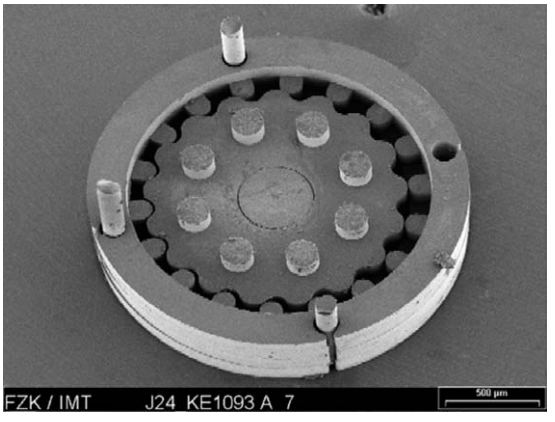


Figure 72 Microcycloid gear system with assembled central gear; the thin cylinders on the outer ring are pieces of optical fibers simply used as alignment aid. (Reprinted with permission by William Andrew Publishers from MEMS: A Practical Guide to Design. Analysis and Applications, Edited by Jan Koruink and Oliver Paul, 2006.)

For the assembly of the gears, three disks were initially positioned just one on top of each other using a vacuum gripper under a stereomicroscope. The correct alignment was done afterward by introducing short pieces of optical fiber (diameter 125 μm) as alignment pins into the respective holes (diameter 135 μm) in the disks. Finally, dynamic tests of the gear system were successfully conducted with a mechanical torque input by an electrical motor. A proper rotational speed reduction was observed in the operational input range of 3–1500 rpm with the designed gear ratio of 18.

1.11.3.4.3 Gears for watch movements

Gears become attractive to luxury watch industry for use in mechanical movements if they fulfill the extreme accuracy requirements so that lubrication becomes unnecessary and if they can be made simultaneously from precious materials. Hence, watch gears are fabricated from hard gold in small series, i.e., a few thousand pairs per annum (Meyer *et al.* 2005), at the Forschungszentrum Karlsruhe in cooperation with the synchrotron source ANKA (ankaweb.fzk.de/). **Figure 73** shows a set of gears and levers that have been patterned by X-ray lithography and electroplating. Beryllium masks are used to ensure the required accuracy in several hundred thick PMMA layers, and a new electrolyte for hard gold, which is an alloy of 99% gold and only 1% additives, mainly Cd and As (Dambrowski *et al.* 2006), was developed. It leads to almost pure (24 carat) golden microstructures with a Vickers hardness of already 160 HV. Further hardening of the sidewalls by ion implantation of beryllium increases the hardness by an additional factor of 3 (Meyer *et al.* 2005). Meanwhile, the implantation process has been modified and less toxic elements are used. The top surface is polished by fly cutting. Besides the esthetic motivation, golden gears benefit from the facts of being nonmagnetic and noncorrosive.

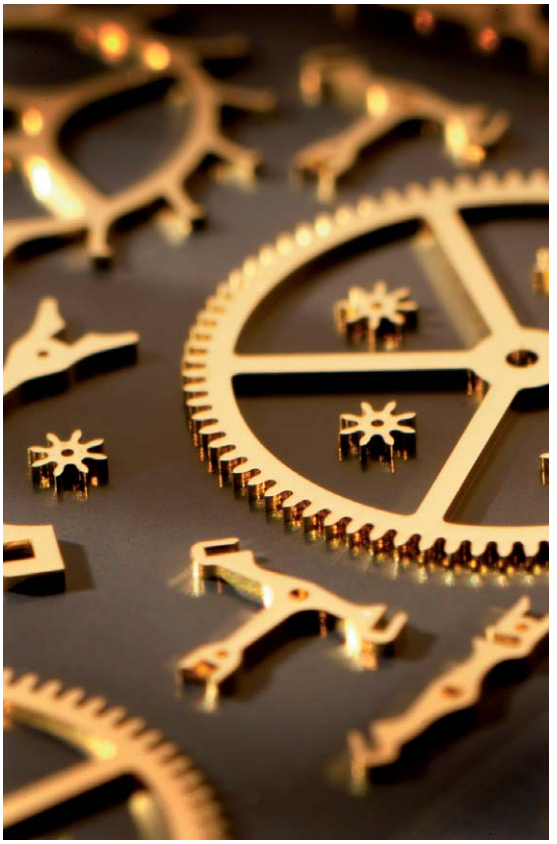


Figure 73 Watch gears and levers electroplated in hard gold.

The reproducibility of the optimized process is excellent as can be seen from **Figure 74**. It shows a trend chart of dimensional deviations of a test pattern, which was patterned on 200-μm-thick PMMA after the development. The variation is on the order of ± 0.5 μm.

1.11.3.4.4 Turbines

We conclude the section on mechanical application of LIGA technology with a hydrodynamic rotating microdrive, which provides an alternative to electrostatic or electromagnetic micromotors when electricity has to be avoided either for medical reasons or due to the risk of explosion hazards. Therefore, a microturbine has been developed, which is normally operated with water but may also run on gas. It was originally developed for use on cardiac catheters to drive a cutter for plaque removal.

In contrast to classical macroscopic turbines, microturbines feature a planar design. The turbine shown in **Figure 75** is assembled from various microcomponents (Wallrabe 1998): an electroplated metallic rotor with radial blades is mounted on a shaft, a nozzle plate made from PMMA by X-ray lithography surrounds the rotor, and a housing also made from PMMA, however using injection molding, forms the distal tip of the cardiac catheter. The fluid is delivered from the back side through the outer annular tube of the housing. It is injected into the rotor through the nozzle plate and is drained back through the inner annular ring. The hollow cylinder in the center of the turbine housing is a tribute to their application on catheters. It allows for positioning the catheter using a guide wire, which is a commonly used technique in catheter labs.

The nozzle geometry (minimum width 15 μm) is one of the parameters, which define the power and the torque as well as the operation pressure. The number of rotor blades and nozzles is kept coprime, thus guarantying a mostly steady torque generation.

The assembled turbine system is shown in **Figure 76**. For better visibility of the nozzles it is coated with gold. The turbine is operated with water at 37 °C. At 88 ml min^{-1} the pressure drop across the turbine is about 7 bar. The turbine achieves an idling speed of 1250 rps and a maximum power of 45 mW corresponding to a torque of 8 μN m at 900 rps.

1.11.3.5 Other Applications

This subsection describes two very new applications of LIGA technology. One deals with radio frequency (RF) MEMS for RF devices, which has been a main playground for silicon technology so far. The other concentrates on very small structures with still fairly high aspect ratios in the submicrometer range.

1.11.3.5.1 RF resonators

A commonly used device in RF technology is a microcapacitor with variable capacitance. Various variable capacitors have been patterned in silicon technology, in many cases with the capacitor plates positioned horizontally, i.e., in parallel to the substrate. A possible design for a vertical plate capacitor as it can be fabricated by means of LIGA technology is sketched in **Figure 77**. The capacitor consists of a cantilever, which is connected to a ground electrode, and a neighboring capacitance electrode. The cantilever is released from the substrate and can move back and forth. In order to vary the gap width of the capacitor, an additional actuator electrode is placed on the opposite side of the cantilever. If the cantilever

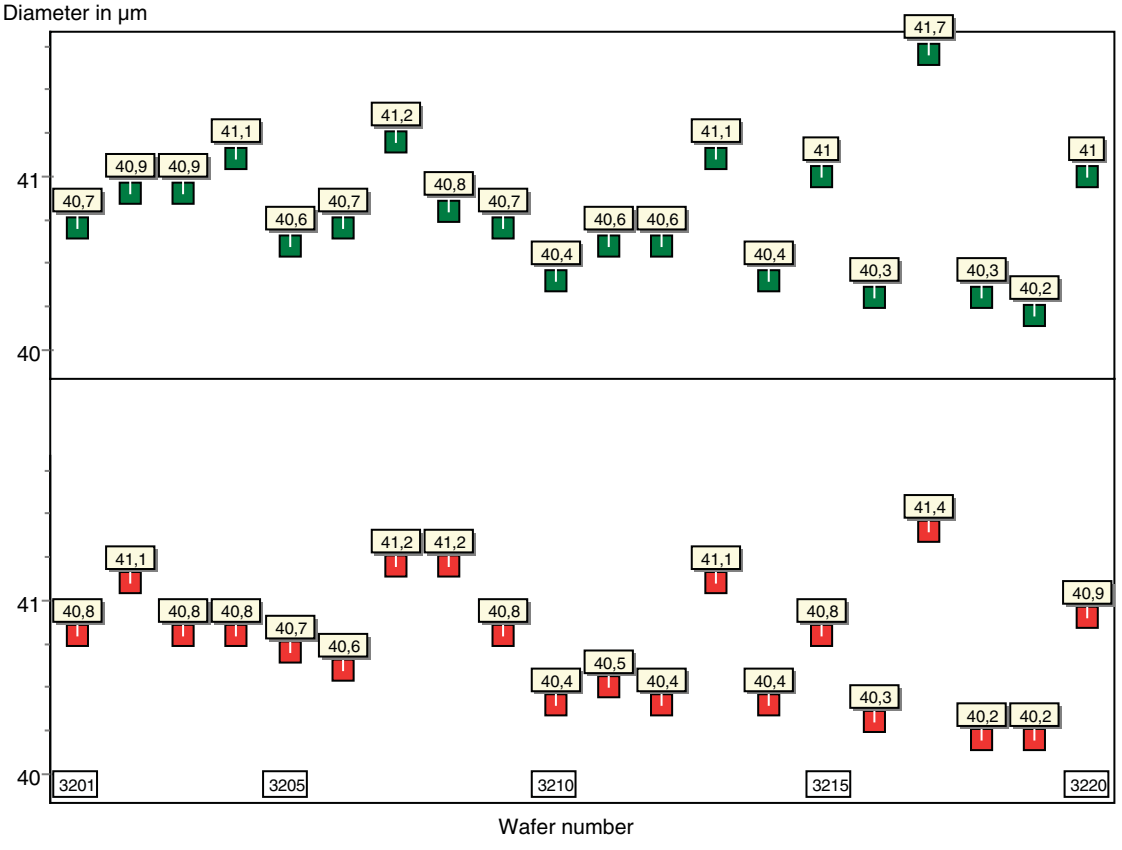


Figure 74 Trend chart of size reproducibility on 200-µm-thick polymethylmethacrylate (PMMA) test patterns after development.

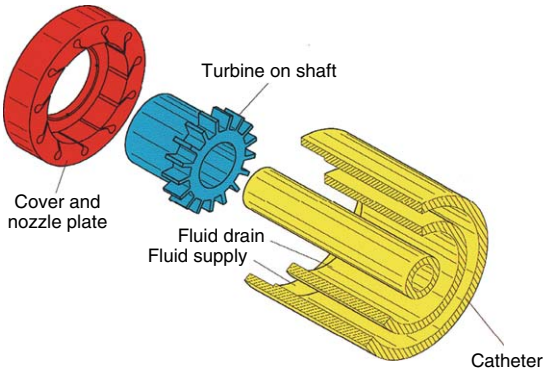


Figure 75 Schematic view of a turbine for cardiac catheters; the outer diameter is 2.5 mm.

Figure 76 Assembled turbine system; for visualization the sealing cover plate is still missing.

is actuated toward the actuator electrode, an increase in the tuning range over a two-plate capacitor can be achieved only if the actuator gap is made larger than the capacitance gap (Haluzan and Klymyshyn 2004). This increase in the tuning range at the cost of higher tuning voltages keeps the RF signal and the control signal separate from each other.

The capacitors are made by X-ray lithography with subsequent electroplating of metal and, as a final step, time-controlled sacrificial layer etching to release the cantilever locally from the substrate.

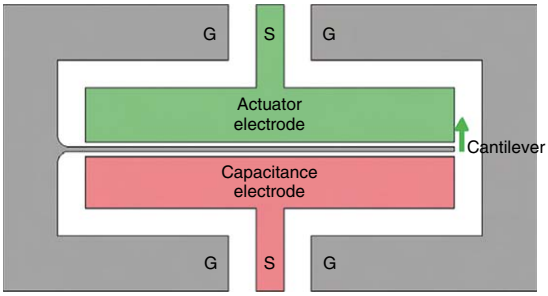


Figure 77 Sketch of three-plate variable capacitor. The additional ground structure on the right-hand side facilitates testing using standard three-pin wafer probes.

The benefits of this approach are the following: RF applications require good, preferably metal, conductors. The plating process of LIGA allows for the use of not only nickel, which is the most commonly plated metal in LIGA but also high-conductivity metals, such as copper and gold. The typical smooth sidewalls reduce those losses that originate from the skin effect. The penetration depth decreases with increasing frequency and is normally less than 1 µm for high-conductivity metals. Hence, when the penetration depth is of the same order as the surface roughness, high losses are induced in the rough surface. Additionally, LIGA can be performed on substrates with low dielectric losses such as quartz or alumina ($\epsilon_r = 9.8$ at 10 GHz).

Figure 78 shows a variable capacitor that is electroplated in 100-µm-thick nickel (Achenbach *et al.* 2007). The capacitance gap width is 1.6 µm and the actuator gap width is 6 µm. The length of the cantilever and of the electrodes is 1 mm. In the original PMMA form that was then filled with metal, the very small capacitance gap appeared as an extremely thin and long PMMA line. In order to improve the stability of such a fragile structure, which, in addition, has a respective small contact surface to the substrate and thus a very reduced adhesion, supporting auxiliary structures have been patterned, which can be figured out in the center of **Figure 78**.

Figure 79 shows the measured Q factor of the capacitor with a very low series resistance from **Figure 78** (Achenbach *et al.* 2007). At 3 GHz, it is still 33 and reaches a value of 175 at a frequency of 1 GHz. These already competitive results can be still improved by further reduction of the series resistance. Copper and gold show higher conductivity by a factor of 3–4 and are currently electroplated instead of nickel.

Other LIGA-made RF components are also subject to research and development. These are 3-dB line couplers with a center frequency of 18 GHz (Kachayev *et al.* 2003) or cavity resonators with center frequency at 24 GHz (Ma *et al.* 2005). In the latter case, very thick resist layers up to 2 mm need to be patterned with the specific advantage of still smooth surfaces. For details we refer to given references.

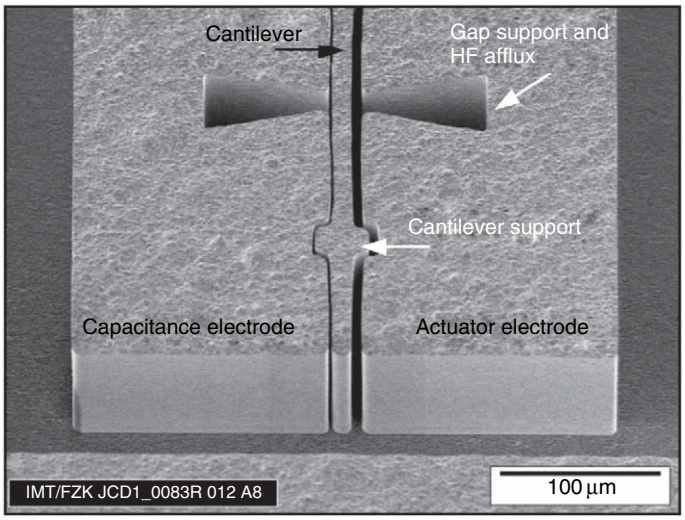


Figure 78 Scanning electron microscope (SEM) of a 100-µm-thick variable capacitor electroplated in nickel. The capacitance gap is 1.6 µm and features supporting auxiliary structures. (Source: Achenbach S, Klymyshyn D, Haluzan D, Mappes T, Wells G, Mohr J 2007 Fabrication of RF MEMS variable capacitors by deep X-ray lithography and electroplating, *Microsyst Technol*. **13(3–4)**, 343–7.

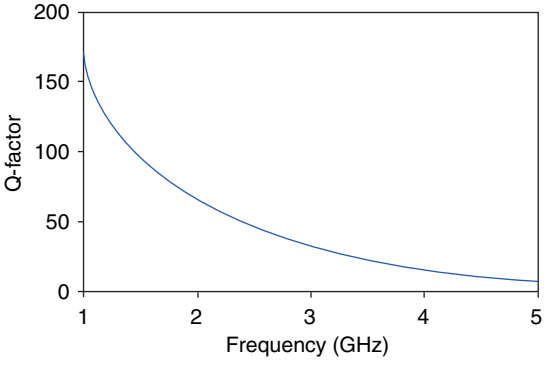


Figure 79 Measured average *Q* factor as a function of frequency for the 100-μm-thick nickel variable capacitor.

1.11.3.5.2 Submicron structures

A new approach for LIGA technology is driven by structures with feature sizes well below 1 μm. The feasibility of nano-LIGA is still under evaluation. First, submicron structures have been patterned successfully in polymer as well as in metal. The obtained aspect ratios are not yet the same as for the LIGA structures as discussed above, but aspect ratios of about 10 have been proven so far. The limiting factor that occurs during the development and subsequent drying of submicron polymer structures is an insufficient stability of those structures with respect to capillary forces. **Figure 80** shows the effect of the capillary force acting on a set of PMMA columns of 550 nm diameter and 5 μm resist thickness (Achenbach *et al.* 2004). Although the adhesion of the columns to the substrate is sufficient to keep

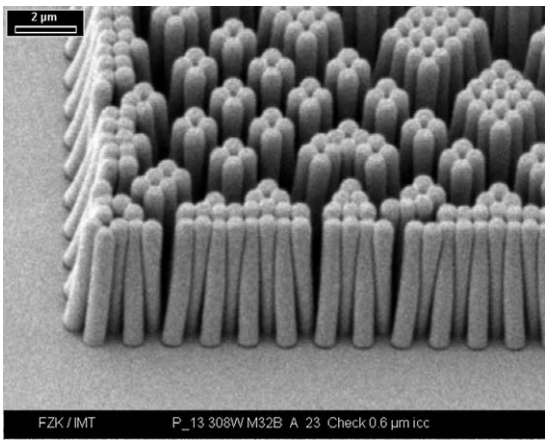


Figure 80 Scanning electron microscope (SEM) of polymethylmethacrylate (PMMA) columns of 550 nm diameter and 5 μm thickness. The deformation is a result of capillary forces during drying.

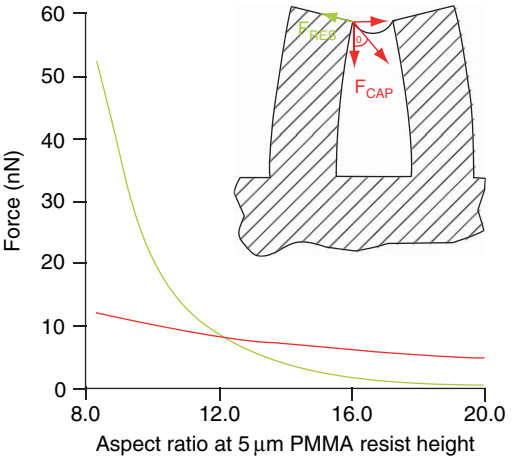


Figure 81 Comparison of capillary force and reset force as a function of aspect ratio for infinite resist lines (polymethylmethacrylate (PMMA), thickness 5 μm). The inset shows the capillary force and the induced opposing reset force.

them in place, they are bent and clustered forming a relative regular crystalline pattern. **Figure 81** compares the capillary force with the opposing elastic reset force given by the geometry of the microstructure as a function of the aspect ratio (Mappes *et al.* 2007). In this case, an infinite PMMA line was considered for a resist thickness of 5 μm. At an aspect ratio of approximately 12, the capillary force becomes larger than the reset force. Hence, the structure will become unstable and will show respective deformations.

In addition, diffraction of the X-rays at the edges of the absorber structures on the X-ray mask during the exposure needs to be considered for submicron structures. This is typically not a severe task in DXL. Even large proximity gaps between mask and resist surface during the exposure do not cause dramatic damage in case of larger LIGA structures. For submicron structures, however, diffraction becomes visible, as can be seen from **Figure 82**. Here, an array of PMMA columns approximately 1 μm diameter shows small pits in the center of the columns that appear after the development. They resemble the Poisson's spot observed behind a circular obstacle, as is known from the optics lecture. **Figure 83** shows a predicted resist profile after a simulation of dose deposition under consideration of diffraction as derived in (Nazmov *et al.* 2000) and subsequent development.

One prospective application of these regular submicron structures with high aspect ratio are filters for

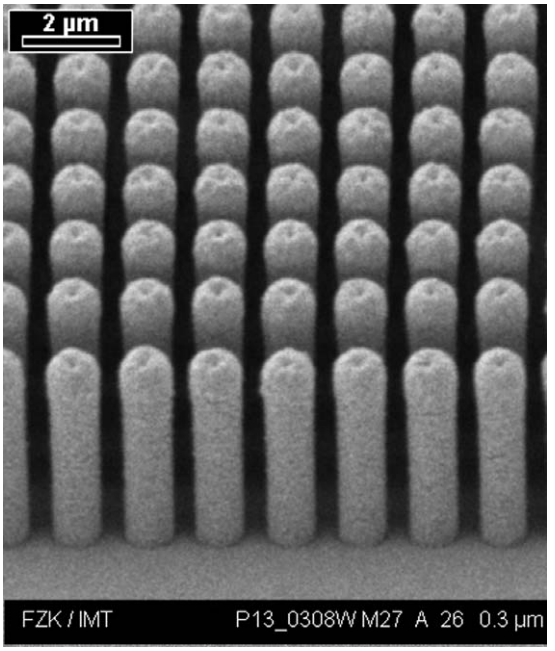


Figure 82 Array of polymethylmethacrylate (PMMA) columns of approximately 1 µm diameter and resist thickness 5 µm. The proximity gap during exposure was 125 µm.

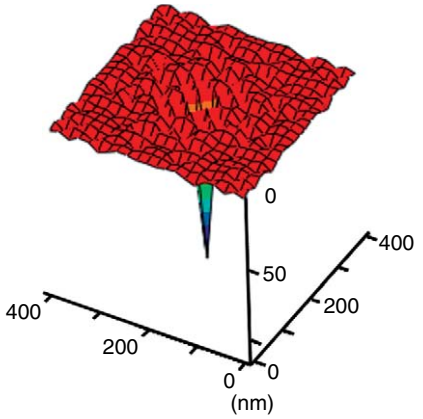


Figure 83 Simulated surface profile of a circular column after development.

infrared light. For an infrared filter, a fairly thick metal mesh with a grating constant on the order of 1 µm is required (Mappes *et al.* 2006). The filter's quality and efficiency is directly dependent on the metal thickness, hence it calls for high aspect ratio structures. An array of PMMA columns of 5 µm thickness, which are similar to those above, is therefore electroplated in gold. After stripping the PMMA, the gold mesh features a thickness of 4 µm with a linewidth of 720 nm, as can be seen from **Figure 84**. Similar meshes have been made free of defects with more than 5600 columns with a grating constant of 1.34 at a column diameter of 1 µm, as shown in **Figure 85**.

1.11.4 Outlook

Originally, LIGA was developed in Germany for a very specific purpose, namely, to manufacture nozzles for the separation of uranium isotopes in large enrichment facilities. The requirements of the technology were high aspect ratio and minimum feature sizes in the micrometer range. Its potential became evident in the 1990s when numerous examples for structures and devices were presented. The lead in MEMS technologies, however, was assumed by silicon micromachining, and not by LIGA, in spite of its technical superiority for many applications. One main reason for this is the required technical infrastructure, process know-how, and cost issues. Silicon-based technologies on the one hand could exploit the vast technology base developed for chip making with an existing investment worth billions of dollars. LIGA, on the other hand, was new and confined to research laboratories. Besides the process itself, the laboratories invested a huge amount of money and brainwork to develop suitable machines, such as X-ray scanners and micromolding machines. Furthermore, as a key process step, LIGA required access to a synchrotron radiation facility, again a research laboratory, and thus often unacceptable for industries establishing manufacturing plants. Most of the synchrotrons offer worldwide access to their facilities. An hour of exposure time is currently on the order of €200 per beamline, which – to be honest – does not take into account the investment of the equipment, i.e., the beamline and the scanner. For the implementation of a new exposure setup with its own beamline, overall costs of approximately €2 000 000 may arise. Nevertheless, this is a demonstration of the strength and superiority of LIGA; several industrial LIGA products have been launched by industries and synchrotron radiation facilities and are currently widely used for microfabrication (**Table 2**).

As a consequence of the weaknesses of LIGA, the developers of LIGA have focused their efforts over the past few years on quality management, stabilization of processes, standards, and cost. Small volume production was demonstrated and reliability of processes was dramatically increased. The concept of deliver mold inserts on demand was verified, and at least small- and medium-sized companies are increasingly accepting that they can now simply purchase a mold insert for their production in their plant.

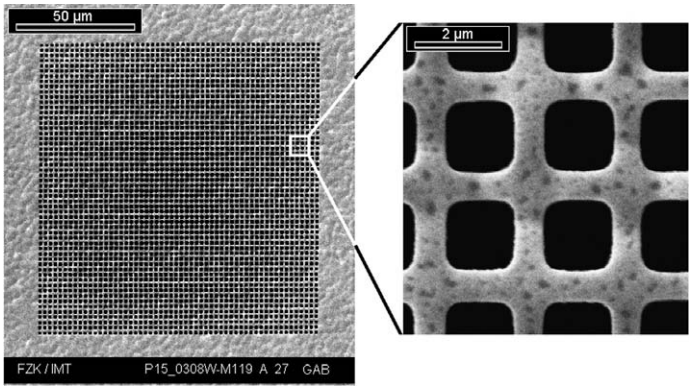


Figure 84 Infrared filter made from electroplated gold. The trench width is 720 nm and the thickness is 4 µm.

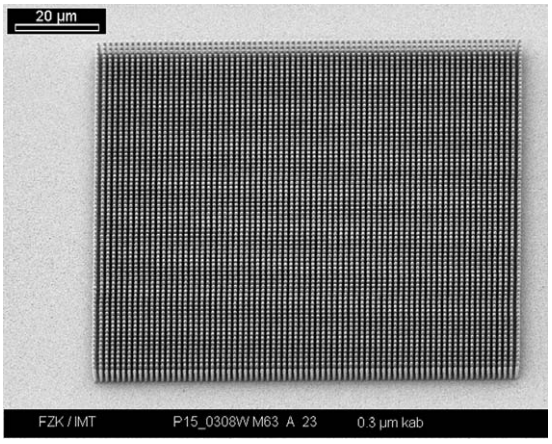


Figure 85 Array of more than 5600 defect-free polymethylmethacrylate (PMMA) columns of 1 µm diameter.

Furthermore, a new so-called Direct-LIGA concept is currently tested where the components are produced directly by X-ray lithography with or without subsequent electroplating, similar to semiconductor manufacturing, at an acceptable cost. This approach uses only the first (LI) or the first two (LIG) steps of LIGA, which can become cost-effective in terms of throughput, especially for a moderate number of pieces. Two highly encouraging approaches are pursued: new resists such as SU-8 reduce the exposure times by one to two orders of magnitude and the ability to expose very large areas in one exposure step.

Besides these efforts in making LIGA acceptable as a manufacturing technology for various industrial products, cutting edge research in and with LIGA remains a hot topic. The goals include research in new materials, new replication techniques, and new lithography approaches exploiting the short wavelength nature of X-rays. Questions such as How small can we really get? Can we overlap our top-down technologies with the typical bottom-up approach in nanotechnology? and at which dimensions? are currently being addressed. Making devices much smaller than is possible today will open up entirely new fields in research and applications.

Table 2 Synchrotron radiation centers with facilities for deep X-ray lithography and LIGA and commercial providers of LIGA services and products

Synchrotron radiation facilities with LIGA activities	*Commercial LIGA providers*
Australia: Australian Synchrotron – under construction	ANKA GmbH, Germany
Brazil: LNLS	
Canada: CLS – under construction	Axsun Corp., USA
France: SOLEIL – under construction	
Germany: ANKA, BESSY-II	MEZZO Corp., USA
Italy: Elettra	
Japan: NewSubaru, AURORA	
Korea: PSL	Anwenderzentrum at BESSY, Berlin; Germany
Russia: VEPP-III	
Singapore: SSLS	
Sweden: MAX-II	
Taiwan: NSRRC	
USA: ALS, APS, CAMD, NSLS-II, SPEAR-III, SRC	

References

Abel S 1996 Charakterisierung von Materialien zur Fertigung elektromagnetischer Mikroaktoren in LIGA-Technik. PhD thesis, University of Kaiserlautern

Abel S, Ehrfeld W, Lehr H, Möbius H, Schmitz F 1996 Charakterisierung von Materialien zur Fertigung elektromagnetischer Mikroaktoren in LIGA-Technik. *Proc. Micro Mat '95', Deutscher Verband für Materialforschung e. V*, Berlin, Germany, S. 413–20

Achenbach S, Klymyshyn D, Haluzan D, Mappes T, Wells G, Mohr J 2007 Fabrication of RF MEMS variable capacitors by deep X-ray lithography and electroplating. *Microsyst. Technol.* **13(3–4)**, 343–7

Achenbach S, Mappes T, Mohr J 2004 Structure quality of high aspect ratio sub micron polymer structures patterned at the Electron Storage Ring ANKA. *J. Vacuum Sci. Technol. B* **22(6)**, 3196–201

Achenbach S, Pantenburg F J, Mohr J 2000 Optimierung der Prozessbedingungen zur Herstellung von Mikrostrukturen durch ultratiefe Röntgenlithographie (UDXRL). FZKA 6576, Forschungszentrum Karlsruhe

Bacher W, Bley P, Hein H, Klein U, Mohr J, Schomburg W K, Schwarz R, Stark W 1991 Herstellung von Röntgenmasken für das LIGA-Verfahren. *KfK-Nachrichten* **23**, 76–83

Bacher W, Ruprecht R, Michaelis A, Schultze J W, Thies A 1992 Titanoxid als Startschichten in der Mikrogalvanik, Dechema-Monographienband 125. VCH Verlagsgesellschaft, pp. 459–71

Bade K, Leyendecker K, Thommes A, Bacher W 1996 Electroplating at high aspect ratio micropatterned electrodes-influence of mass transfer. *Proc. 4th Int. Symp. Magnetic Materials, Processes, and Devices. Applications to Storage and Microelectromechanical Systems (MEMS)*. Eletrochemical Society, Pennington, NJ, USA, pp. 697–708

Becker E W, Ehrfeld W, Hagmann P, Maner A, Münchmeyer D 1986 Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming, and plastic moulding (LIGA process). *Microelectr. Eng.* **4(1)**, 35–6

Becker E W, Ehrfeld W, Münchmeyer D, 1984 Untersuchung zur Abbildungsgenauigkeit der Röntgentiefenlithografie mit Synchrotronstrahlung bei der Herstellung technischer Trenndüsenelemente. KfK Report 3732, Kernforschungszentrum Karlsruhe

Becker E W, Ehrfeld W, Münchmeyer D, Betz H, Heuberger A, Pongratz S, Glashauser W, Michel H J, Siemens Rv 1982 Production of separation nozzle system for uranium enrichment by a combination of X-ray lithography and galvanoplastics. *Naturwissenschaften* **69(11)**, 520–3

Boerio F J, Williams J W 1981 Structure and properties of organosilane primers for adhesive bonding. *Appl. Surface Sci.* **7(1–2)**, 19–31

Dambrowski N, Schulz J, Bade K 2006 Mikrogalvanoformung von Goldbauteilen für mechanische Applikationen. *Galvanotechnik* **1**, 188–92

Ehrfeld W, Bley P, Götz F, Mohr J, Münchmeyer D, Schelb W, Baving H J, Beets D 1988 Progress in deep-etch synchrotron radiation lithography. *J. Vac. Sci. Technol. B* **6(1)**, 178–82

Eicher J, Peters R P, Rogner A 1992 Produkte und Einsatzgebiete der LIGA-Technik, VDI Report No. 960. VDI-Verlag, Düsseldorf, pp. 683–97

El-Kholi A, Bade K, Mohr J, Pantenburg F J, Tang X-M 2000 Alternative resist adhesion and electroplating layers for LIGA process. Microsyst. Technol. **6**, 161–4

El-Kholi A, Bley P, Gottert J, Mohr J 1993 Examination of the solubility and the molecular weight distribution of PMMA in view of an optimised resist system in deep etch X-ray lithography. *Microeletr. Eng.* **21(1–4)**, 271–4

Feiertag G, Ehrfeld W, Freimuth H, Kolle H, Lehr H, Schmidt M, Sigalas M, Kiriakidis G, Pedersen T, Soukoulis C 1997 Fabrication of photonic crystals by deep X-ray lithography. *Appl. Phy. Lett.* **71(11)**, 1441–3

Göttert J, Fischer M, Müller A 1995 Mikrolinsen und ihre Anwendung in der Medizin- und Informationstechnik, FZKA 5670. Forschungszentrum, Kalrsruhe, pp. 169–74

Greeneich J S 1975 Developer characteristics of poly-(methyl methacrylate) electron resist. *J. Electrochem. Soc.* **122(7)**, 970–6

Guttmann M, Schulz J, Saile V 2005 Lithographic fabrication of mold inserts. In: Baltes, Brand, Fedder, Hierold, Korvink, and Tabata (eds.) *Adavanced Micro and Nanosystems, Vol. 3 Microengineering of Metals and Ceramics*. Wiley-VCH, Weinheim, pp. 187–219

Hafizovic S, Hirai Y, Tabata O, Korvink J G 2003 X3D: 3D X-ray lithography and development simulation for MEMS. *Proc. Transducers '03*, Boston, MA, USA, pp. 1570–3

Haluzan D, Klymyshyn D 2004 High-Q-LIGA-MEMS vertical cantilever variable capacitors for upper microwave frequencies. *Microwave Opt. Tech. Lett.* **42**, 507–11

Harsch S, Ehrfeld W, Maner A 1988 Untersuchungen zur Herstellung von Mikrostrukturen großer Strukturhöhe durch Galvanoformung in Nickelsulfamatelektrolyten, KfK-Nachrichten, Vol. 4455.

Hein H, Bley P, Goettert J, Klein U 1992 Elektronenstrahllithographie und Simulationsrechnungen für die Herstellung von Röntgenmasken beim LIGA-Verfahren. Congress Gerätetechnik und Mikrosystemtechnik 1992, Chemnitz, Germany. Report No. 960, VDI-Verlag, Düsseldorf, pp. 75–86

Himmelhaus M, Bley P, Mohr J, Wallrabe U 1992 Integrated measuring system for the detection of the number of revolutions of LIGA microturbines in view of a volumetric flow sensor. *J. Micromech. Microeng.* **2**, 196–8

Hirata T, Chung S J, Hein H, Akashi T, Mohr J 1999 Micro cycloid-gear system fabricated by multi-exposure LIGA technique. *SPIE Int. Symp. Micromachining and Microfabrication 99*, Santa Clara, CA, USA, Vol. 3875, pp. 164–71

Kachayev A, Klymyshyn D, Achenbach S, Saile V 2003 High vertical aspect ratio LIGA microwave 3-dB coupler. *ICMENS 2003, Int. Conf. MEMS, NANO, and Smart Systems*, Banff, CA, USA, pp. 38–43

Koch E E (*ed*.) 1987 *Handbook on Synchrotron Radiation.* North-Holland, Amsterdam, Vol. 1A, ISBN 0-444-86709-0

Krippner P, Kühner T, Mohr J, Saile V 2000 Microspectrometer system for the near infrared wavelength range based on the LIGA technology. In: Mariella R PJr. (ed.) *Micro- and Nanotechnology for Biomedical and Environmental Applications.* SPIE, Vol. 3912, pp. 141–9

Krippner P, Mohr J, Saile V 1999 Electromagnetically driven microchopper for integration into microspectrometers based on the LIGA technology. In: Motamedi M E and Göhring R (eds.) *Miniaturized Systems with Micro-Optics and MEMS.* SPIE, Vol. 3878, p. 144

Kunz T, Kohl M, Ruzzu A, Skrobanek K, Wallrabe U 2000 Adhesion of Ni-structures on Al_2O_3 ceramic substrates used for the sacrificial layer technique. *Microsyst. Technol.* **6**, 121–5

Ma Z, Klymyshyn D, Achenbach S, Mohr J 2005 Microwave cavity resonators using hard X-ray lithography. *Microwave Optical Tech. Lett.* **47(4)**, 353–7

Maner A, Ehrfeld W, Schwarz R 1988 Electroforming of absorber patterns of gold on masks for X-ray lithography. *Galvanotechnik* **79(4)**, 1101–6

Mappes T, Achenbach S, Mohr J 2006 X-Ray lithography for devices with high aspect ratio polymer submicron structures. *Book of Abstr., 32nd Int. Conf. Micro- and Nano-Engineering 2006, MNE 2006*, Barcelona, Spain, September 17–20, 2006, pp. 375–6

Mappes T, Achenbach S, Mohr J 2007 Process conditions in X-ray lithography for the fabrication of devices with sub-micron feature sizes. *Microsyst. Technol.* **13(3–4)**, 355–60

Meyer P, Lange J, Arendt M, Dambrowsky N, Saile V, Schulz J 2005 Launching into a golden age (2) – High precision parts for luxurious Swiss watches. *Commercialization of Micro and Nano Systems (COMS 2005): 10th International Conference*, Baden-Baden, Germany, August 21–25, 2005. MANCEF, Albuquerque, NM, USA, pp. 241–4

Mohr J, Burbaum C, Bley P, Menz W, Wallrabe U 1990 Movable microstructures manufactured by the LIGA-process as basic elements for microsystems. In: Reichl H (ed.) *MICRO SYSTEM Technologies '90, 1st Int. Conf. Micro, Electro, Opto, Mechanic Systems and Components,* Berlin, Germany. Springer-Verlag, Germany, pp. 529–37

Mohr J, Ehrfeld W, Münchmeyer D 1988 Requirements on resist layers in deep-etch synchrotron radiation lithography. *J. Vac. Sci. Technol.* **B6(6)**, 2264–67

Müller A, Goettert J, Mohr J, 1996 Aufbau hybrider mikrooptischer Funktionsmodule für die optische Nachrichtentechnik mit dem LIGA-Verfahren. FZKA 5786, Forschungszentrum Karlsruhe

Nazmov V P, Mezentsva L A, Pindyrin V F, Petrov V V, Yakoleva E N 2000 Modelling and manufacture of regular microstructures with high aspect ratio in acrylic plastic. *Nucl. Instrum. Meth. A* **448**, 493–6

Nazmov V, Reznikova E, Mohr J, Saile V, Snigirev A, Snigireva I, DiMichiel M, Drakopoulos M, Simon R, Grigoriev M 2004a Refractive lenses fabricated by deep SR lithography and LIGA technology for X-ray energies from 1 keV to 1 MeV. *AIP Conf. Proc.* **705**, 752–5

Nazmov V, Reznikova E, Mohr J, Snirgirev A, Snirgireva I, Achenbach S, Saile V 2004b Fabrication and preliminary testing of X-ray lenses in thick SU-8 resist. *Microsyst. Technol.* **10**, 716–21

Nazmov V, Reznikova E, Snirgirev A, Snirgireva I, DiMichiel M, Gregoriev M, Mohr J, Matthis B, Saile V 2005 LIGA fabrication of X-ray nickel lenses. *Microsyst. Technol.* **11(4–5)**, 292–7

Oka T, Nakajima H, Shiratsuki A, Tsugai M, Wallrabe U, Hollenbach U, Krippner P, Mohr J 2001 Development of a micro optical distance sensor with electric I/O interface. *11th Int. Conf. Solid-State Sensors and Actuators, TRANSDUCER '01*, Munich, Germany, pp. 536–9

Pantenburg F J, Mohr J 1995 Influence of secondary effects on the structure quality in deep X-ray lithography. *Nucl. Instrum. Meth. B* **97(1–4)**, 551–6

Ruzzu A, Haller D, Mohr J, Wallrabe U 2003 Optoelectromechanical switch array with passively aligned free-space optical components. *J. Lightwave Technol.* **21(3)**, 664–71

Samper V D, Sangster A.J, Reuben R L, Wallrabe U 1999 Torque evaluation of a LIGA fabricated electrostatic micromotor. *J. Microelectromech. Syst.* **8(1)**, 115–23

Schenk R, Halle O, Müllen K, Ehrfeld W, Schmidt M 1997 Highly sensitive resist material for deep X-ray lithography. *Microelectr. Eng* **35(1–4)**, 105–8

Schnabel W 1981 *Polymer Degradation*. Hanser International, München, ISBN 3446132643

Schnabel W, Sotobayashi H 1983 Polymers in electron beam and X-ray lithography. *Prog. Polymer Sci.* **9(4)**, 297–365

Schnabel W. 1978 In: Jellinek H H G *(ed.) Aspects of Degradation and Stabilization of Polymers.* Elsevier, New York, pp. 153–6, 176.

Schomburg W K, Baving H J, Bley P 1991 TI- and BE-X-ray masks with alignment windows for the LIGA process. *Microelectr. Eng.* **13(1–4)**, 323–6

Schulz J, Mono T, Chung S J, Mohr J 1996 The effect of residual solvent on the profiles of thick positive DNQ-photoresist for microsystem technologies. *Microsyst. Technol.* **2**, 50–5

Skrobis K J, Taylor J W, Engelstad R L 1995 Bond mechanisms and bond strengths of solvent-welded PMMA sheets for use as deep X-ray resist in microfabricated. *Proc. 188th Meeting of the Electrochemical Society*, October 8–13, 1995, pp. 337–46

Slatter R, Degen R, Burisch A 2006 Micro-mechatronic actuators for desktop factory applications. *ISR/ROBOTIK 2006 – Joint Conference on Robotics*, Munich, Germany, May 2006. VDI-Bericht Nr. 1956, VDI Wissensforum IWB GmbH, Germany, pp. 137–50

Snigirev A, Kohn V, Snigireva I, Lengeler B 1996 A compound refractive lens for focusing high energy X-rays. *Nature* **384**, 49–51

Strohrmann M, Eberle F, Fromheim O, Keller W, Kromer O, Kijhner T, Lindemann K, Mohr J, Schulz J 1994 Smart acceleration sensor systems based on LIGA micromechanics. In: Reichl H (ed.) *Microsystem Technologies '94, 4th Int. Conf. Micro Electro, Opto, Mechanical Systems and Components,* Berlin, Germany. Springer-Verlag, Germany, pp. 753–62

Tabata O, Matsuzuka N, Yamaji T, Uemura S, Yamamoto K 2002 3D fabrication by moving mask deep X-ray lithography (M^2DXL) with multiple stages. *IEEE MEMS 2002, 15th Int. Conf. Micro Electro Mechanical Systems*, Las Vegas, NV, USA, pp. 180–3

Tabata O, Matsuzuka N, Yamaji T, Uemura S, Yamamoto K 2002 3D fabrication by moving mask deep X-ray lithography (M^2DXL) with multiple stages. *IEEE MEMS 2002, 15th Int. Conf. Micro Electro Mechanical Systems*, Las Vegas, NV, USA, pp. 180–3

Thommes A, Stark W, Bacher W 1995 Die galvanische Abscheidung von Eisen-Nickel in LIGA-Mikrostrukturen. Scientific Report FZKA-5586, Forschungszentrum, Karlsruhe

Tomie T 1996 X-ray lens. *Japanese Pat.* 1 994 000 045 288, priority 18.02.1994

Wallrabe U 1998 *Mikroturbinen als hydrodynamischer Kleinstantrieb, F&M, Feinwerktechnik, Mikrotechnik, Mikroelektronik 9*. Carl Hanser Verlag, München, Germany, pp. 646–9

Wallrabe U, Mohr J 2003 Modular microoptical systems for sensors and telecommunication.In: Baltes H, Korvink J, and Fedder G (eds.) *Sensors Update.* Wiley-VCH, New York, Vol. 12, pp. 143–74

Wallrabe U, Bley P, Krevet B, Menz W, Mohr J 1994 Design rules and test of electrostatic micromotors made by the LIGA process. *J. Micromech. Microeng.* **4(1994)**, 40–5

Wallrabe U, Dittrich H, Friedsam G, Hanemann Th, Mohr J, Müller K, Piotter V, Ruther P, Schaller Th, Zißler W 2002 Micromolded easy-assembly multi fiber connector: RibCon. *Microsyst. Technol.* **8**, 83–7

Wallrabe U, Solf C, Mohr J, Korvink J G 2005 Miniaturized Fourier transform spectrometer for the near infrared wavelength regime incorporating an electromagnetic linear actuator. *Sens. Actuators A* **123–124C**, 459–67

Ziegler P, Wengelink J, Mohr J 1999 Passive alignment and hybrid integration of active and passive optical components on a microoptical LIGA bench. *3rd Int. Conf. Micro Opto Electro Mechanical Systems MOEMS'99*, Mainz, Germany, pp. 186–9

Biographies

Jürgen Mohr, born in 1957 received his diploma in physics in 1983 and his Ph.D. in mechanical engineering in 1987 from the University of Karlsruhe with an emphasis in micro system technologies. He has been with the Forschungszentrum, Karlsruhe since 1987, first as a research associate working in the field of X-ray lithography for micro fabrication (LIGA process). Since 1992 he is leading the X-ray lithography and micro optics group at the Institute of Micro Structure Technology of the Forschungszentrum. His interests are in the field of micro-fabrication by Deep X-ray lithography (LIGA process) as well as in micro optical components and systems based on polymers for sensor and telecommunication applications. He is a member of the IEEE-LEOS society. He has published more than 100 papers on various topics in micro technology and applications in mechanics, optics and the sensor field. Jürgen Mohr is co-author of two text books on micro technology.

Ulrike Wallrabe studied physics at Karlsruhe University, Germany. In 1992 she received her PhD degree for mechanical engineering of microturbines and micromotors. From 1989 to 2003 she was with the Institute for Microstructure Technology (IMT) at Forschungszentrum Karlsruhe where she focused on microactuators and MOEMS made by the LIGA technique. Since 2003 she holds a professorship for microactuators at the Department of Microsystems Engineering, IMTEK, at the University of Freiburg, Germany. Her focus lies in magnetic microstructures, coils, and adaptive optics. Ulrike Wallrabe has published more than 60 journal and conference papers, as well as three book chapters.

Dr. Jan G. Korvink obtained his M.Sc. in computational mechanics from the University of Cape Town in 1987, and his Dr. sc. techn. in applied computer science from the ETH Zurich in 1993. After his graduate studies, he joined the Physical Electronics Laboratory of the ETH Zurich, where he established and led the MEMS Modelling Group. This was followed by a move in 1997 to the Albert Ludwig University in Freiburg, Germany, where he holds a Chair position in microsystem technology and runs the Laboratory for Microsystem Simulation. Currently, Dr. Korvink is a vice dean of the Faculty of Applied Science and a member of the board. He has written more that 180 journal and conference papers in the area of microsystem technology, and co-edits the review journal *Applied Micro and Nanosystems*, see http://www.wiley-vch.de/books/info/amn. His research interests cover the modelling and simulation and low cost fabrication of microsystems.

1.12 Low-Cost MEMS Technologies

Bruce K. Gale, Mark A. Eddings, Scott O. Sundberg, Andrew Hatch, Jungkyu Kim, and Tammy Ho, Department of Mechanical Engineering, Salt Lake City, UT, USA

1.12.1 Introduction

While microelectromechanical systems (MEMS) technologies have demonstrated ever more amazing capabilities, the costs associated with MEMS processing have also grown. MEMS and microfabrication technologies have traditionally been dominated by expensive, complicated, cutting-edge equipment. Technologies such as photolithography, dry etching, physical vapor deposition (PVD), chemical vapor deposition (CVD), and related tools that have dominated the MEMS field since its inception are notoriously expensive and rely on processes and components that are inherently costly. Vacuum components, complicated electronics, and precision stages are required for most of these processes and all carry significant price tags. Typical research grade tools usually cost more than US$100 000 and normally a suite of tools is required for any MEMS chip to be fabricated. A basic MEMS research lab often costs more than US$10 million and can be much more expensive depending on the tooling and infrastructure (clean rooms, electrical, and gas systems) required. Manufacturing tools typically run over US$1 million and entire facilities can cost more than US$1 billion. For products where the cost of this equipment and tooling could be spread over millions of units each year, the use of these tools can be easily justified. For potential products where the number of parts each year is less, these large start-up costs can often be prohibitive. In the research arena, the high cost of setting up MEMS facilities often limits research in this field to those institutions with the deepest pockets. In addition, the traditional MEMS technologies are best suited to mass production, but for prototyping, they are often difficult to work with and can again be overly expensive. Thus, over the last decade, extensive efforts have been devoted to developing technologies for MEMS and for microfabrication that are accessible to all researchers, not just those focused on MEMS. Note that the technologies discussed in this work are primarily for generating microscale features between 1 and 100 μm or larger.

One of the first approaches to reducing costs in MEMS was to move to materials less costly than single-crystal silicon wafers, which is the traditional MEMS substrate. Glass, plastics, and some ceramics have proven to be suitable substrates for many MEMS applications and have also proven to be less expensive. Some of the earliest work in this field combined low-cost materials with some of the traditional MEMS processes that were also low cost, such as wet etching and spin coating to produce glass microchannels (Manz *et al.* 1990). Some of these materials have also been found to have better properties than silicon for some applications in biology, microfluidics, optics, chemistry, and medicine. As researchers began to explore new materials, new technologies had to be developed in many cases, since the traditional MEMS processes were optimized for silicon or were specific to silicon. Thus, many of the techniques discussed in this chapter are specific to nonsilicon substrates, while others can be used with any material. Many of these new processing techniques have proven to be capable, yet inexpensive. Unfortunately, for chips requiring low linewidths (<10 μm), traditional (or expensive) MEMS techniques have proven difficult to replace, but for MEMS devices where lateral dimensions are not as critical or relatively large (>10 μm) and it is primarily vertical dimensions that are very small (<1 μm), the techniques discussed in this chapter have proven their mettle.

Low-cost MEMS and microfabrication techniques typically fall into one of the following four categories:

(1) Low-cost photolithography techniques
(2) Photolithography replacements
(3) Rapid prototyping techniques
(4) Low-cost mass production technologies

The first category, low-cost photolithography techniques, is primarily directed toward producing inexpensive photolithography masks. A standard chrome mask on a glass substrate typically costs more than US$5 per square centimeter and can easily be double or triple that amount depending on the

needed quality. For a device that requires multiple masks, the mask-making costs can multiply quickly. In a prototyping situation where several design cycles may be required, the mask cost can be the primary cost driver for a project and can exceed US$1000 per mask. Techniques that may be included in this category are laser and ink-jet printed masks, xurography, and shadow masking using masks made with traditional machining tools. The second category, photolithography replacements, is characterized by techniques designed to eliminate the need for masks, UV exposure systems, optical aligners, and expensive chemicals. The least expensive UV exposure systems of any quality (providing collimated, uniform light) cost at least US$20 000 and systems that provide the ability to align different layers often exceed US$250 000 for basic systems. Techniques in this category include xurography, screen printing, and laser cutting. The third category, rapid prototyping techniques, is designed to eliminate the difficulties and costs associated with producing small-volume prototypes using traditional MEMS tools. Most MEMS processes are fantastically well developed for mass production, since they operate as batch processes, but that also means that it often costs just as much to make one part as it does to make hundreds of parts. Since there is a limited benefit to multiple parts during prototyping, the cost cannot be spread out and can be significant. In addition, microfabrication techniques are typically relatively slow, and even relatively simple designs can take weeks to manufacture and outsourced designs are often not delivered for 6 months or more. Thus, many of the inexpensive prototyping techniques are also much faster, saving money as well as time. Some of the techniques that can be used for low-cost prototyping of MEMS include lamination, xurography, laser cutting, soft lithography, and screen printing. The fourth category, low-cost production technologies, is primarily aimed at technologies for MEMS devices that do not require the substantial infrastructure required for traditional MEMS, but can still produce similar volumes at a relatively lower cost. Techniques in this category typically involve molding, embossing, or stamping. Each of these categories will be addressed in detail in this work. The reader should note that the low-cost techniques can usually be combined with more traditional MEMS techniques to maximize productivity while minimizing cost.

An important consideration in MEMS fabrication is the packaging of the completed MEMS devices. Conventional wisdom is that MEMS packaging can account for up to 80% of the total cost of a MEMS device, primarily because many MEMS packages must be custom designed. To overcome these packaging costs, a number of approaches have been proposed. The most successful involve modifying or using traditional integrated circuit (IC) packages, but this is often not possible for fluidic or optical systems. Some MEMS designers have managed to essentially eliminate packaging. A few of these methods as well as a few methods for minimizing packaging costs will be reviewed with their related manufacturing techniques.

1.12.2 Low-Cost Photolithography Techniques

Many engineers and scientists have worked to reduce the cost of performing basic photolithography processes or steps. Two primary approaches have been taken. The first approach is to reduce the cost of making the photolithography masks (see Section 1.12.2.1). The second approach is to eliminate the processing steps after photolithography by creating the desired microstructures using the photolithography process (see Section 1.12.2). This approach can be combined with the first approach. A third approach is to eliminate photolithography entirely, which will be discussed in Section 1.12.3.

As noted earlier, standard photolithography masks can be a substantial expense and outsourced masks often take significant time to receive. Most photolithography masks are typically made of precisely patterned UV-absorbing materials such as chrome (most common), iron, or photopatternable emulsions on a glass substrate. Chrome masks are patterned using a tool called a pattern generator, which exposes a layer of photoresist deposited on the metal layer. The photoresist is then developed and the underlying metal layer etched to produce the mask pattern, which can then be reproduced multiple times. The pattern generators are robotic instruments with precise movement and optics to produce high-quality patterns, and are thus relatively expensive. A typical mask can take several hours of machine time to produce, and complex masks can easily take days. The multiple steps required to produce the masks, the precise tooling, and long tool-use times lead to substantial mask fabrication expenses. Thus, to reduce these costs, several alternatives to metal masks have been explored.

1.12.2.1 Low-Cost Photolithography Mask Fabrication

Although typical chrome mask generation with e-beam lithography is very accurate (<1 μm feature sizes) it is very expensive (>US$500 per 4 × 4 in mask) and can have a long turnaround time. For this reason, alternatives to chrome mask generation have been evaluated in depth by many groups. Laser printing, microfiche and photoreduction, microscope projection, microfluidics, xurography, and direct laser writing have all been implemented as creative photomask alternatives (Narasimhan *et al.* 2004).

1.12.2.1.1 Laser printed masks

Fabricating photolithography masks by printing the positive/negative mask onto a transparency using laser printing is a common technique. Creating masks with this technique is appealing because it can significantly reduce the cost and time of fabrication (Qin *et al.* 1996). High-resolution laser printers can have resolutions from 5000 to 20 000 dpi and can print linewidths down to 10 μm with a local accuracy of 2 μm. The equipment is expensive but there are many outsourcing locations for printing on transparencies (<US$40 per transparency). Laser printers use a process where a photoreceptor, which is a photosensitive drum or belt that rotates, is electrostatically charged. A laser diode is then directed at a moving mirror, which aims the laser beam onto the photoreceptor and reverses the charge wherever it hits. The surface is then exposed to a toner, a fine plastic powder with coloring, and the toner is attracted to the areas on the photoreceptor where the laser beam hit. The photoreceptor is rolled over the paper in order to transfer the image and the paper is then passed through rollers, which provide heat and pressure to bond the toner to the paper. Following toner bonding, the photoreceptor is cleaned and the charge is removed. Therefore, laser printers rely on the spot size of the laser beam generated to determine the minimum feature size, which is determined by the laser quality used. The feedthrough mechanism is also important in providing the local accuracy. The two different methods used are mechanical feedthrough and a fixed internal drum. It has been found that the fixed internal drum is capable of a higher accuracy dot placement (Suleski and O'Shea 1995).

1.12.2.1.2 Photographic pattern reduction

Photographic reduction of prints, using 35-mm film or microfiche, can be used to further reduce the feature size capabilities for mask generation (Deng *et al.* 1999). This reduction allows the use of a regular office printer to create linewidths down to 15 μm, making mask generation very cost-effective (Deng *et al.* 2000). Using a high-resolution laser printer, photoreduction can yield minimum linewidths of 5–10 μm (about US$20 per 10 reductions) (Dotson *et al.* 2004). To further reduce feature sizes, microscope projection has been used. A printed photomask is placed in the field stop, diaphragm, of the microscope. Light passes through the photomask and objective and exposes photoresist on a substrate located on the stage. As the light passes through the objective (typically 40× or 100×), the pattern is reduced allowing feature sizes down to 1 μm to be achieved. The main disadvantage of this reduction technique is the limited area that can be exposed, although step and repeat operations could be used (Love *et al.* 2001, Whitesides *et al.* 2001).

1.12.2.1.3 Microfluidic gray-scale masks

Microfluidics has been used to create photomasks and even gray-scale masks, in order to create 3D structures. The microfluidics are fabricated by first printing a mask with the desired pattern and then exposing photoresist on a silicon wafer. Once the photoresist is developed, poly(dimethylsiloxane) (PDMS) is molded using the photoresist structures and then bonded to a spun PDMS thin layer. Dye is injected into the PDMS channels with varying concentrations. UV light is then passed through the PDMS channels onto a substrate with photoresist. Then, depending on the dye concentration, the developed structures vary in thickness, thus creating 3D structures (Chen *et al.* 2003).

1.12.2.1.4 Masks generated using xurography

Photomasks and shadow masks can be fabricated using xurography, a fabrication technique explained in depth in Section 1.12.3.5. For photomasks, a UV-opaque material is chosen to be patterned and the undesired material is peeled away, without removing the backing material. This mask is then transferred and adhered to a glass substrate, thus creating a photomask. For shadow masking, a material is cut with the desired pattern and transferred to a substrate, using adhesive or electrostatic interactions for adhesion. Material can then be deposited over the substrate (i.e., sputtering, PVD, e-beam evaporation). Once the material is deposited, the cut pattern is peeled away, removing the deposited material

where the pattern was adhered. This fabrication technique can provide minimum linewidths of ~25 μm (Bartholomeusz *et al.* 2005).

1.12.2.1.5 Laser electrodeposition for masks

Photomasks have also been developed using direct laser writing on an electroless Ni deposited substrate, an inexpensive deposition method requiring little equipment. This technique can use an inexpensive Ar^+ laser, rather than an expensive pulsed deep UV light source, but takes advantage of a high numerical aperture objective to increase the power density and decrease the spot size of the laser. The smallest feature size generated by this setup was 0.5 μm (Lorenz *et al.* 2004). This technique can provide a relatively simple, economical, and rapid alternative to e-beam lithography while still providing comparable feature size capabilities.

1.12.2.2 Direct Microfabrication Using Photolithography

A number of techniques have been developed to eliminate some of the steps after photolithography, and thereby reduce the total cost of the microfabrication process. These processes include direct fabrication of the microstructure during the exposure process. Examples of this type of work have been completed with both polymers and glass. In both cases, the only step after UV exposure is development using a solvent for the polymers and an acid for the glass.

1.12.2.2.1 Photopatternable polymers

The development of MEMS and microfluidic components often require the use of complex 3D geometries. The rapid fabrication of 3D curves, vias, porous membranes, and tubes have been made possible through the implementation of photoactivated polymers. The basic operating principle of photopatternable polymers is the introduction of a UV source onto a photosensitive polymer, initiating a cross-linking reaction across exposed regions. Excess polymer and reactants are generally washed away leaving the microstructure. Two of the more common techniques used in photopatternable polymers include liquid-phase photopolymerization and SU-8 lithography.

1.12.2.2.1.(i) Liquid-phase photopolymerization

The technique often described as liquid-phase photopolymerization provides a rapid, inexpensive prototyping option for fabricating microstructures. Molds and layered structures are possible in minutes without the use of costly clean room facilities. Variations of this method have been widely explored by Beebe *et al.*, making possible the swift creation of 3D microchannels, filters, fibers, and tubes.

For basic mold or structure patterning, a generic cartridge fixed to an adhesive gasket with the thickness of the desired structure is placed upon a substrate such as a microscope slide and is filled with UV-sensitive polymer through sample inlet ports (Beebe *et al.* 2000, Khoury *et al.* 2002). A photomask is then placed over the cartridge and a UV source exposed to pattern the polymer within the cartridge. The exposed polymer is then flushed with solvent, removing any unexposed polymer waste, thus revealing the patterned structure. The quality of the resulting structures is dependent upon the resolution of the equipment used and the chemical properties of the polymer. Typically, PDMS is poured over the polymer structures and microfluidic channels are created by direct molding. Three-dimensional channel networks can be created by using slight variations on this technique and stacking multiple layers of polymerized material (Mensing *et al.* 2005).

Other complex structures such as fibers and tubes that utilize the microscale characteristics of multiple combined laminar flows and photopolymerization are possible under different configurations (Jeong *et al.* 2004, 2005). The fabrication apparatus consists of a PDMS substrate containing a microchannel, with an inlet and outlet pipette inserted into the ends of the microchannel to provide sample delivery. A separate channel that intersects the inlet pipette is cored into the PDMS to provide sheath flow around the sample flow. A UV source is placed above the outlet pipette to polymerize the sample flow inside the sheath flow in order to produce microfibers. The addition of another smaller pipette inside the inlet pipette enables the addition of a nonpolymerizable core flow that results in microtubes as the sample flow is polymerized at the outlet pipette (**Figure 1**).

The fabrication of microfilters and porous membranes is accomplished through randomly generated patterning (Moorthy and Beebe 2003). A photosensitive polymer is prepared consisting of monomers, a photoinitiator, a cross-linker, and additives such as water. After the UV source is applied and polymerization has occurred the additive is removed, leaving behind the porous network within the polymer.

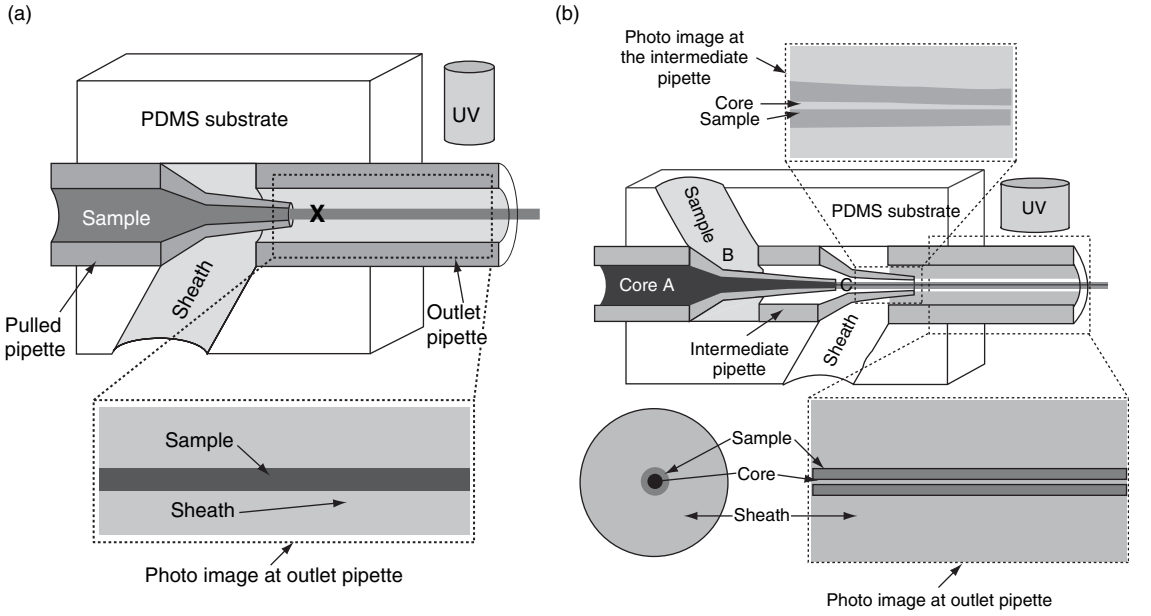


Figure 1 A representation of the fabrication apparatus: (a) microfibers; (b) microtubes. PDMS, poly(dimethylsiloxane). (Source: Jeong W J, Kim J Y, Kim S J, Lee S H, Mensing G and Beebe D J 2004 Hydrodynamic microfabrication via 'on the fly' photopolymerization of microscale fibers and tubes. *Lab Chip* **4**, 576–80; reproduced with permission from the Royal Society of Chemistry.)

The technique is advantageous for microsystems as the size of the pores is easily controlled by the concentrations of the reagents and the membrane can be created in any location such as a chamber or a microchannel located within a device. The ability to create filter regions locally offers many possible applications in microfluidics.

1.12.2.2.1.(ii) SU-8 lithography The negative photoresist Epon SU-8 or SU-8, which was first developed by IBM, has found wide use in MEMS thick-film-resist applications (Despont *et al.* 1997, LaBianca and Gelorme 1995, LaBianca *et al.* 1995, Leek *et al.* 1995, Shaw *et al.* 1997). The unique optical properties, harsh environment durability, and lithographic contrast have made SU-8 a popular photoresist for high aspect ratio structures. The use of SU-8 has enabled researchers to produce high-resolution structures for direct molding or protective coatings for electroplating. Drawbacks to using SU-8 include several baking sessions to condition the photoresist and clean room facilities to reduce dust and particles that may affect film quality. However, with proper equipment and facilities, high-resolution, durable processing is possible in an afternoon for a relatively low cost (**Figure 2**).

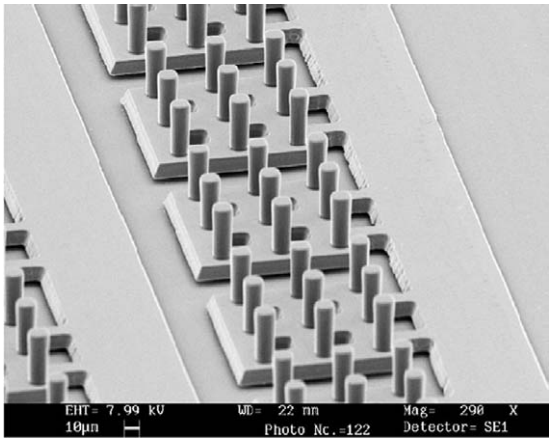


Figure 2 SU-8 pillars fabricated on an array of silicon cantilevers. The pillars are $10 \times 10\ \mu m^2$ in size and 30 μm high. (Source: Conradie E H, Moore D F 2002 SU-8 thick photoresist processing as a functional material for MEMS applications. *J. Micromech. Microeng.* **12**, 368–74; reproduced with permission from the Institute of Physics.)

There are many varieties of SU-8 that have been developed by varying the solvent contained in the photoresist. The solvent used will generally affect photoresist viscosity and wettability for different types of substrates such as silicon and glass. These

properties will determine the quality and thickness of the spun layer.

The ability to control layer thickness is an import feature of SU-8 rapid prototyping. Direct PDMS molding and optical waveguides are the most common applications for SU-8 lithography. Tight control of structure width and height has elevated SU-8 to the photoresist of choice for creating microchannels in PDMS (Anderson *et al.* 2000, Duffy *et al.* 1998). Also, when producing molded microchannels, there is usually only one mask step, so alignment is eliminated along with the expensive equipment associated with the alignment process. The process of using SU-8 or some other type of mold for PDMS casting is called soft lithography (see Section 1.12.4.1). The SU-8 and substrate surface is chemically coated to allow clean, unhindered peeling of the PDMS microchannel replica. An additional slab of PDMS is placed on top of the grooves to form microchannels. The PDMS layers will combine through an oxygen plasma treatment or partial curing of layers. Following a final bake a single PDMS slab remains with embedded microchannels. Other PDMS microcomponents such as pumps, actuators, and sensors are also possible through SU-8 molding (Unger *et al.* 2000). While the SU-8 processing is somewhat time-consuming, once a master mold is made, repeated PDMS molding is possible.

SU-8 can be used to make microchannels directly or by a subsequent bonding step. SU-8 exposure without development followed by a series of SU-8 application and exposure steps can lead to the production of buried SU-8 channels or 3D SU-8 networks (Gracias *et al.* 2005). Variations of this process using laminated layers of SU-8 can achieve a similar result (Abgrall *et al.* 2005). Simple patterning of SU-8 followed by a bonding step using adhesives or other materials can be readily used for creating microfluidic channels, though the bonding process is not always simple (Gale *et al.* 2002).

Direct use of SU-8 in MEMS components has also been shown to form optical waveguides. Because of the excellent optical properties of SU-8 waveguides, they have been used as biosensors and in other detection schemes (Chang-Yen and Gale 2003). For rapid and efficient prototyping, direct integration of the optical waveguide reduces coupling losses and interference from the microsystem platform. The direct fabrication of SU-8 into MEMS devices facilitates this integration by removing additional packaging and fabrication steps.

1.12.2.2.2 Photosensitive glass

Photosensitive glass allows for direct fabrication of glass microstructures, which can eliminate a number of later processing steps. In most cases the elimination of the later steps allows for an overall reduction in the cost of device fabrication, which is necessary to offset the relatively high cost of obtaining photosensitive glasses.

Photosensitive glass has a unique set of characteristics that allow the etch rate of the glass to be modified using common lithographic techniques, creating a 20-fold increase in the etch rate of the modified areas. This physical change allows for direct fabrication of glass structures and makes it possible to create anisotropically etched features in the glass without the need for masking the glass during the chemical etch itself. **Figure 3** shows diagrams for this process.

The unique chemical behavior of photosensitive glass is due to its special composition. Each brand of this type of glass has a proprietary recipe. The important constituents of all types of photosensitive glass are the introduction of metallic ions – in this case silver (Ag^{+}) – and cesium oxide (CeO_2) into the glass matrix (Stookey *et al.* 1978). When the photosensitive glass is exposed to UV light or other high-energy sources such as a proton beam (Gomez-Morilla *et al.* 2005), infrared femtosecond laser (Hongo *et al.* 2005), or X-rays (Shimizugawa *et al.* 2003), the cesium

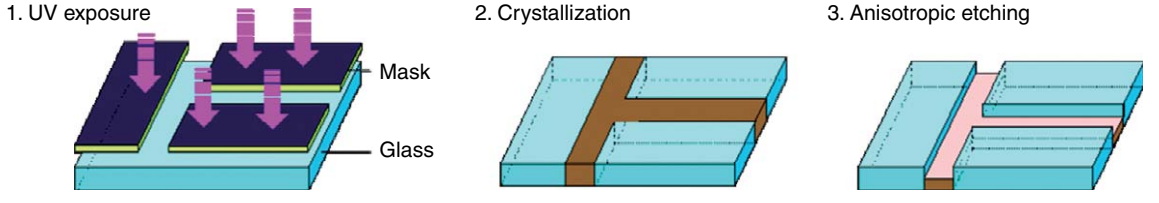


Figure 3 Microfluidic structures can be fabricated in photosensitive glass by first masking and exposing the glass to UV light, then heat treating the glass, and finally covering the entire surface with an HF-containing solution.

causes the metals to precipitate out of the glass matrix. The nucleation of the metals and other additives during heat treatment causes a visual discoloration of the exposed areas and a local weakening of the molecular network. It is this alteration of the glass matrix that causes the exposed areas to etch at rates much higher than the rates for the unexposed regions. Because of these unique properties, photosensitive glass has some distinct advantages and disadvantages. Since the chemical etch is anisotropic in nature, virtually straight sidewalls are formed. This contrasts with the curved etch profile associated with isotropic etching. In addition, photosensitive glass can be etched with high aspect ratios, while the typical isotropic etch can never be more than half as deep as it is wide. **Figure 4** shows a structure that was formed by anisotropically etching photosensitive glass.

Photosensitive glass is often used to create microfluidic channels, which requires another layer to seal or close off the channel. Multiple layers of photosensitive glass may be bonded using fusion bonding techniques (Olsen and Serpa 1978) and other traditional glass-bonding methods such as pressure-assisted bonding, chemical bonding, water bonding (Ito *et al.* 2001), and glass–metal bonding. Anodic bonding is not typically used with photosensitive glass due to mismatch problems in the thermal expansion coefficient between the glass and silicon. Other channel sealing methods using tapes, PDMS, and other polymers have been demonstrated. Embedded 3D channels have also been created in photosensitive glass using femtosecond lasers (Masuda *et al.* 2003), but this technology is certainly not inexpensive.

1.12.3 Photolithography Replacements

A number of techniques have been developed to circumvent the need for any photolithography steps in the production of microscale devices. These techniques typically use some sort of cutting or machining process to produce microstructures that can then be reproduced using molding. For those techniques not amenable to later molding, the processes, even though they are serial in nature, are sufficiently fast and inexpensive that they can easily compete with photolithographic processes. Some of these processes, such as ink-jet printing and screen printing can be direct replacements for photolithography since they can be used to deposit a resist material directly. Microcontact printing (μCP) can also be used to deposit resists and other materials as a replacement for photolithography, but as it is a subset of soft lithography, which is discussed elsewhere (Section 1.12.4.1.2), it will not be discussed here. Techniques such as xurography and laser cutting of laminates produce microdevices directly (or masks

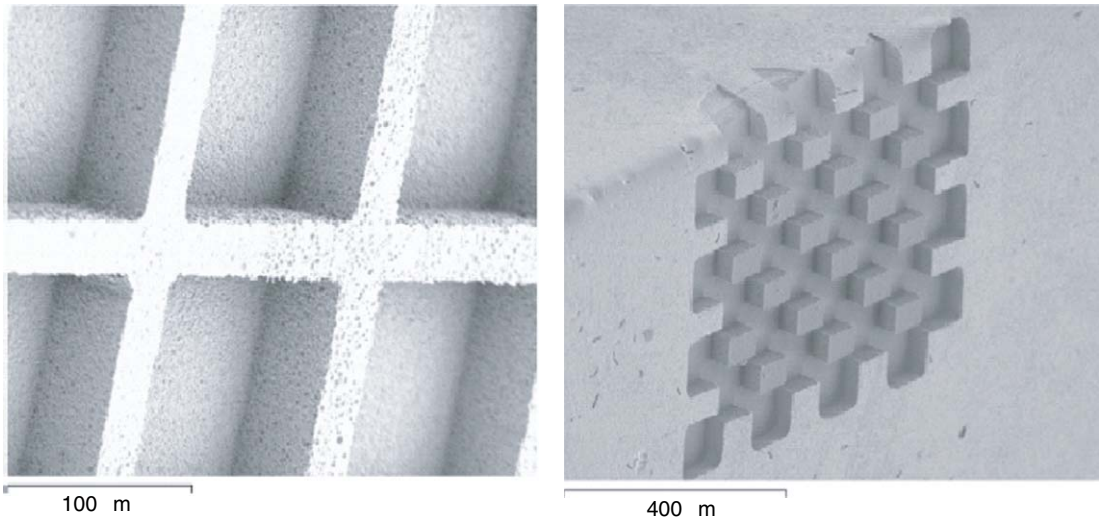


Figure 4 Features made using an anisotropic etch in photosensitive glass. (Source: Gomez-Morilla I, Abraham M H, de Kerckhove DG, Grime G W 2005 Micropatterning of foturan photosensitive glass following exposure to MeV proton beams. *J. Micromech. Microeng.* **15**, 706–9; reproduced with permission from the Institute of Physics.)

directly) and can eliminate not only the photolithography process, but all of the subsequent etching and deposition processes.

Traditional machining techniques are also beginning to enter the microdomain and impressive demonstrations of efficient, relatively inexpensive devices fabricated using micromilling, microelectro discharge machining (microEDM), and related technologies are becoming more common (Hupert *et al.* 2006). These techniques, though, tend to be serial and very slow, so they are typically only useful for making molds that are later used to produce microdevices. Thus, they will not be explored further here, especially since there is significant literature already available on these traditional techniques.

1.12.3.1 Ink-Jet Printing

In general, two types of deposition methods are used with ink-jet printing: continuous-mode technology and the more frequently used drop-on-demand (DOD) technology. The continuous-mode print head uses a pressurized fluid chamber to create a continuous fluid stream, which is divided into droplets by pulsing the fluid with a transducer. These droplets are electrostatically charged and are able to be deflected in order to either hit or miss a moving target substrate. The DOD method uses a volumetric change, at ambient pressure, in the fluid to create droplets either by the displacement of a piezoelectric element (Wallace 1989) or by forming a bubble in the ink by heating a resistive element (Aden *et al.* 1994).

1.12.3.1.1 Feature size capabilities

For ink-jet printing, the properties of the print head, the substrate material, the dispensed ink, and the mechanical resolution of the printer are key factors in determining the minimum feature size capabilities. The dispensing head of ink-jet printing determines the amount of fluid dispensed per deposition. The DOD method is preferred because it is able to form much smaller droplets (~20–100 μm in diameter) than the continuous-mode system (diameters up to 0.5 mm may be produced). Also, the continuous mode has relatively high priming volumes and is not very consistent for droplet placement. The piezoelectric print head is usually chosen over DOD with microfabrication because it allows a larger range of ink compatibility when compared to the thermal print heads, which are generally limited to water-based inks, unless the head is redesigned for each differing solvent that is desired to be used within the ink.

The temperature and hydrophobicity/hydrophilicity of the substrate material are important in providing minimum microfeatures. Heating the substrate that the ink will be deposited on, generally in the range of 100–300 °C, improves the deposition quality. One reason for this is that the evaporation of the ink is almost immediate once in contact with the substrate, thus eliminating many wetting problems such as surface tension causing bulging or pinching, uneven surface wetting due to contamination, dust particles that can wick the ink, and scratches that may be on the surface, which creates a capillary effect causing the ink to spread. Also, with rapid ink evaporation, multiple layers can be deposited relatively easily and quickly (Fuller *et al.* 2002) to form 3D structures (Gothait 2003). The surface chemistry (hydrophobic or hydrophilic properties) is also important in defining the minimum feature size. Hydrophilic and hydrophobic regions can be defined on the substrate surface to help control where the ink is deposited by having the ink attracted to certain regions and repelled from others (Sirringhaus *et al.* 2000). These surface regions can be fabricated on the substrate surface using common microfabrication techniques or inexpensive alternative methods.

The dispensed ink composition is another important attribute. Viscosity, evaporation rates, surface tension, solvent used, and adhesion promoters are important properties for ink. High viscosity and low evaporation rates are generally recommended for dispensing because the print head is less likely to get clogged. The surface tension and ink-base chemical makeup plays a key role in determining how much the ink will spread once in contact with the substrate. Several solvents that change the homogeneous film formation on the substrate (water, terpineol, acetophenone, anisole, butyl acetate, ethyl acetate, methyl benzoate, toluene, and *p*-xylene) can be used (Tekin *et al.* 2004). The adhesion promoters within the ink determine how well the ink will bind to the substrate. For conductive inks, the size, dispersion, and stability of the metallic nanoparticles in the ink are very important. Clogging can occur if the particles are too large or if aggregation occurs among the metallic nanoparticles. These physical properties also determine the conductivity performance of the metallic ink once deposited on the substrate.

The mechanical resolution of the ink-jet printer plays a role in the quality of microfeatures

achievable. This depends on the mechanical resolution of the print axes. Accurate positioning of the ink-jet head allows for less dimensional errors when fabricating devices and determines minimum linewidth and minimum spacing between features.

1.12.3.1.2 Ink-jet printing applications

Applications for microfabricating devices using ink-jet printing technology includes solder printing, conductor trace printing, printing of adhesives and dielectrics, printing of optical polymers, and multiple-layered structures.

Ink-jet printers have been found capable of depositing conductive (Lee *et al.* 2005, Mei *et al.* 2005) and magnetic inks (Voit *et al.* 2003) efficiently and, therefore, ink-jet printer speed and resolution have been used for addressing specific electronic fabrication needs and creating magnetic mass storage patterns. The deposition of solder bumps on circuit boards and even fabrication of the entire circuit board – the traces, dielectric coating, and solder bumps – all using an ink-jet printer has been demonstrated. It has been shown that up to 600 bumps per second can be achieved using this technology without needing to have contact with the board during fabrication, thus providing the electrical interconnects necessary. Solder bumps as small as 24 μm have been printed (Hayes *et al.* 1999a). Furthermore, it has been shown that the entire circuit board interconnects can be fabricated using solely ink-jet printing. **Figure 5** shows a diagram of how the traces, dielectric coating, and solder bumps could all be fabricated with an ink-jet printer (Hayes *et al.* 1999b).

Printing of thermosets, thermoplastics, and UV-curable adhesives can also be deposited (Meyer 2001) for component bonding or even for the creation of waveguides. For this application, wettability between the surface and the deposited material is important. If the surface energy of the substrate is too high, the adhesion to the subsequent material may not work well. A low-wetting coating can help the flow of adhesive on the substrate without adversely affecting the adhesion.

Three-dimensional ink-jet printing is now regularly used to do rapid prototyping of all kinds of devices (some rather large). On the microscale, some of the most interesting applications include fabrication of microfluidic devices (Cooley *et al.* 2001) and tissue engineering (Saunders *et al.* 2005).

1.12.3.2 Screen Printing

Screen printing is able to provide an inexpensive and versatile manufacturing method for depositing material on almost any surface with precision. This fabrication technique is able to minimize the amount of material waste during deposition by directing it to specified locations, unlike the traditional techniques of spin coating and PVD, which are incapable of this material control. It has been found that screen printing can achieve linewidths and line spacings down to 50 μm (Hughes and Ernster 2003).

The basic principle of ink deposition using screen printing is to pass a squeegee over a screen, forcing ink through the screen onto a substrate. Off-contact printing and contact printing are the two main

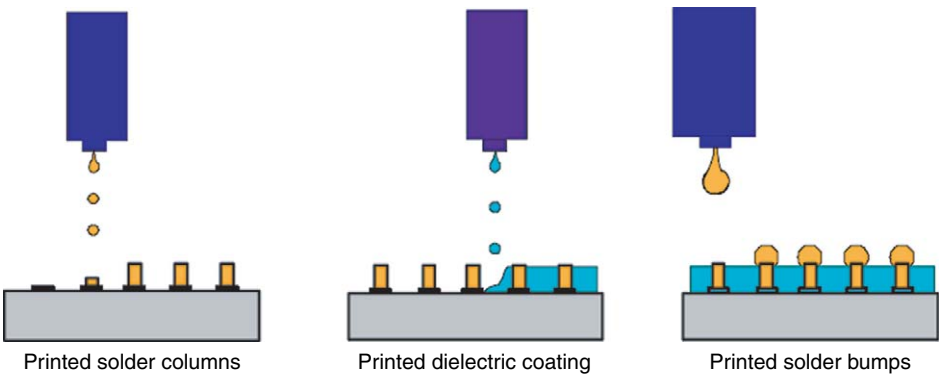


Figure 5 Process diagram for building chip-scale packages by printing solder columns, followed by a dielectric coating between the solder columns, and finally the solder bumps on top of the solder columns. (Source: Hayes D J, Wallace D B, Cox W R 1999 Microjet printing of solder and polymers for multi-chip modules and chip-scale packages. *Proc. IMAPS Int. Conf. High Density Packaging and MCMs*, Denver, CO, USA, pp. 242–7. © 1999, reproduced with permission from IMAPS – International Microelectronics and Packaging Society.)

methods used. Off-contact printing allows the screen to be in contact with the substrate only where the squeegee is located and is the most commonly used technique. Stretching the screen mesh, by applying pressure to the squeegee, causes the screen to come in contact with the substrate and forces the ink particles to pass through the screen. When the screen is released from tension it snaps back, leaving the ink adhered to the substrate, as illustrated in **Figure 6**. Contact printing allows the entire screen to be in contact with the substrate during the printing process and the ink is deposited on the substrate by the force of the squeegee applied to the ink particles. Following ink deposition using one of these methods, the ink is allowed to settle and is then dried, generally between 100 °C and 150 °C, in order to evaporate the solvent out of the ink. Heat treatment may then be given to the dried ink by firing it in a kiln at a temperature between 600 °C and 900 °C (Jeong *et al.* 2001).

Screen printing components include the ink, squeegee, stencil, screen, and substrate. Each of these components introduces variables into the system that will need to be optimized:

(1) The properties of the ink: The ink viscosity affects the printed thickness of the substrate – the lower the viscosity the thinner the print. Drying and firing cause the ink to shrink and must be accounted for when designing a stencil.
(2) Screen printer settings: If the screen distance is small and the screen tension is less, ink will remain in the screen; if it is too large, a thicker pattern that can damage the screen is generated. The faster the squeegee passes over the screen the thinner the film, but if this is too fast it could result in gaps or inconsistencies within the printed pattern. Increasing the squeegee pressure can result in a thicker print. The increased angle and differing geometry of the squeegee can also increase pressure on the ink, yet it does not increase pressure on the screen.
(3) Geometry of the stencil: The tolerances and feature sizes of the stencil generated on the screen determine the minimum feature sizes available for the printed pattern.
(4) The screen: The mesh size, mesh material, and wire diameter/geometry help determine the thickness of the film printed – the smaller the mesh the thinner the print.
(5) The substrate: If the substrate is not flat or not parallel to the screen, film thickness will vary across the substrate (Jabbour *et al.* 2001, Parikh *et al.* 1991).

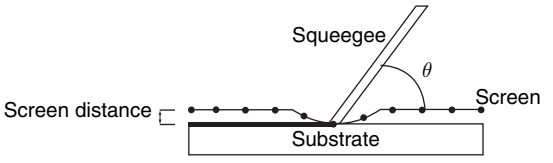


Figure 6 Illustration of off-contact screen printing method. The screen distance and squeegee angle are defined. As the squeegee passes, the screen is stretched to contact the substrate, while forcing ink particles through the screen simultaneously. Once the screen is released from tension it snaps back to its original position, leaving the ink adhered to the substrate.

Many applications of screen printing for microfabrication have been demonstrated. One common use is in generating solar cells that are up to 15% efficient in using screen printing fabrication (Doshi *et al.* 1996, Rohatgi *et al.* 2003, Sakai *et al.* 2005). Screen printing has been used to produce disposable electrochemical sensors (Xu *et al.* 2004). A thick-film screen printing has been used to create diaphragms within ceramic capacitive pressure microsensors (Sippola and Ahn 2005). Flip chip packaging has used screen printing of solder as an inexpensive fabrication method (Chow *et al.* 2000). Fabrication of photodiodes with a response time of 0.32 s and a rise time of 0.48 s has also been achieved using screen printing (Yahaya *et al.* 1998).

1.12.3.3 Stamping

Stamping relies upon the fabrication of a master mold that can generate replica patterns. Molds for creating stamps are typically etched structures in silicon wafers, lithographically patterned SU-8, or electroplated metals. Polymeric materials that are soft and elastic are typically used for the molding of the stamp contact head. PDMS works well for most stamping applications due to its transparency, flexibility, and biocompatibility. For high-resolution applications, more rigid materials are required for the stamp head. However, a trade-off occurs as more rigid materials do not provide the same levels of surface contact that flexible materials provide. In these cases a combination or hybrid material can be used to attain a high degree of contact and resolution. Other important parameters to consider when using

stamping processes are the aspect ratios of the patterning features. For high aspect ratio structures, a tendency to buckle is possible depending on the flexural strength of the material and the applied pressure necessary. Low aspect ratio structures, if spread far apart, may experience bowing from the bulk material between structures that may impede clean stamp patterning. Stamping in microfabrication is typically referred to as microcontact printing and is discussed in much greater detail in Section 1.12.4.1.2.

1.12.3.4 Laminate Microfabrication

Laminate, or layered, fabrication approaches have been developed to enable device assembly from thin sheets of polymer plastics, SU-8 films, photopolymer dry films, PDMS, and other materials. Because of their relatively inexpensive and easily machinable characteristics, polymers have moved to the forefront as materials used for rapid prototyping and fabrication. Polymer laminates have much to offer in the field of MEMS because of their ability to make 3D microfluidic devices that can be passively or actively driven by interfacing with other components. Laminate fabrication techniques allow for quick fabrication times, reduced costs, and 3D structures that are capable of incorporating hybrid elements such as electrodes, filter membranes, pumps, and sensors. Although the use of hybrid elements complicates fabrication and design since multiple patterns and a difficult assembly process are required, rapid fabrication methods, cheap material selection, and simple bonding techniques make laminate fabrication possible and useful (Sasserath and Fries 2002). Laminate MEMS fabrication has been demonstrated commercially as the primary fabrication method used by a company called Micronics based in Washington, USA. The company uses laminate technologies to do both rapid prototyping and commercial production of microfluidic chips for researchers and businesses. Presented here are methods and techniques demonstrating how laminate microfluidic chips can be quickly patterned and bonded together to make 3D microfluidic devices. The limitations of the different techniques are discussed and the major applications presented.

1.12.3.4.1 Laminate materials

Laminate microfluidic chips can be made using a variety of materials and fabrication methods. The two most common materials are thin polymer plastics, which can be bonded using adhesives or thermal processes, and PDMS, which can be bonded using surface plasma treatments and pressure, or thin layer cross-linking and baking. Other laminate layering methods include SU-8 and dry film photoresists that can be photopatterned using photolithography and bonded together to form multilayer laminate structures.

1.12.3.4.1.(i) Polymers Thin polymer sheets, such as Mylar or polystyrene, are commercially mass produced and can be purchased to fabricate polymer laminate chips. These materials are typically ordered as large rolls of thin sheets or can be purchased as smaller precut sheets. Sheet thicknesses can range anywhere from 10 μm to hundreds of microns if desired. Sheets can be made of different polymers and can be coated with various adhesives depending on the desired application of the laminate card and the intended use of buffers and samples. The thin polymer sheets and adhesives are assembled as a five-layer sandwich consisting of the polymer, adhesive layers, and siliconized release layers, as shown in **Figure 7**. The polymer and adhesive materials are generally compatible with aqueous solvents and their selections may vary depending on the solutions with which it will interface. The sandwich of polymer, adhesive, and release layers can all be patterned at the same time and are most practically and inexpensively patterned using a CO_2 laser.

1.12.3.4.1.(ii) SU-8 films SU-8 is a commercially available negative-resist photopolymer that can be spun on to a release substrate and patterned using UV light. The unexposed SU-8 is then typically etched away using an SU-8 developer leaving the patterned SU-8 structure behind. If the SU-8 is spun on to a thin release layer such as polyester (PET), the thin layer of SU-8 can be dried and then attached to another layer of SU-8 using plasma treatment and a slight pressure. The non-cross-linked SU-8 layer can then be cross-linked by UV light passing through the thin PET layer and the uncrosslinked SU-8 will be peeled away along with the PET layer, resulting in a dry release of the SU-8 (Abgrall *et al.* 2006).

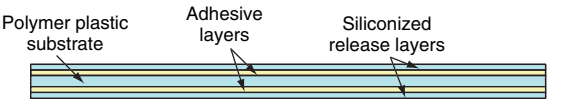


Figure 7 Illustration of how sandwiches of Mylar, adhesive, and release layers are patterned and assembled together.

1.12.3.4.1.(iii) Dry film photoresist Dry film photoresist tape is a prefabricated thin sheet of negative photoresist that is intended to be photopatterned using UV light, quickly bonded to another substrate, and developed using a photoresist developing reagent. Its advantages include reduced process times, reduced UV light intensity exposure, and uniform photoresist thickness. One example of its use is to form photopatterned channels and to sandwich it between two layers of different materials to form microfluidic channels. The use of SU-8 and dry film photoresist layers allows for smooth surfaces and smaller channel thicknesses because of its use in photopatterning, but still requires expensive photomasks, photoresist development, and UV sources and alignment (Tsai *et al.* 2006).

1.12.3.4.1.(iv) Poly(dimethylsiloxane) PDMS is another material that can be used to form 3D laminate microfluidic components and is relatively inexpensive when compared to glass. Some of the advantages of using PDMS include cost, bond strength, softness/deformability, thermal properties, and transparency. PDMS is a soft two-part polymer that is patterned by pouring a thin layer of uncrosslinked PDMS on photopatterned structures, such as SU-8, or by a process of injection molding, then cured using heat or other cross-linking activation processes. Multiple-patterned structures of PDMS can be assembled together using a plasma treatment and pressure, or by applying a thin coating of uncrosslinked PDMS between two partially cured layers, which will create a hermetic cross-linking between each layer when completely cured. This technique has been used to develop a protein array printing method using continuous flow microspotters fabricated from five separately patterned layers of PDMS using injection molding (Chang-Yen and Gale 2005). Some disadvantages associated with PDMS are channel collapsing and poor adhesion effects if channels are too small or too close together. In addition, this material must usually be patterned on the surface of another mold or pattern, which reduces its rapid prototyping abilities and increases costs.

1.12.3.4.2 Laminate patterning methods

Several methods are available for patterning thin laminate layers, including photolithography, as discussed earlier. Examples discussed here include the patterning of polymer laminate cards using a laser, particularly a CO_2 laser. Other polymer plastic patterning methods include techniques using a cutting plotter known as xurography and stamp cutting.

1.12.3.4.2.(i) CO_2 laser cutting Laser ablation and laser cutting first became popular in micromachining because structures could be patterned quickly in a substrate from a simple drawing rather than by using traditional micromachining methods such as photolithography. This process can be taken one step further by using a CO_2 or other laser to completely cut through a polymer sheet rather than only remove a portion of it (Verporte and Rooij 2003). CO_2 lasers can operate anywhere between 10 and 400 W but most materials would typically only require 85 to 120 W lasers to cut through the layers of plastics and adhesives common for MEMS dimensions. Laser cutting is also relatively inexpensive; a typical CO_2 laser can be purchased for less than US$10 000 and is very flexible in its patterning abilities. Though laser cutting is a serial process, the lasers can work very quickly and a significant numbers of parts can be generated rapidly.

The pattern to be etched into the polymer plastic can be drawn in a computer-aided design (CAD) program and input into the translation stage control software of the laser to automate the cutting out and patterning of each polymer layer. This allows for a rapid prototyping setup where fabrication can move from the design stage to the actual patterning and cutting stage within a number of hours. The smallest feature sizes demonstrated using a CO_2 laser in Mylar is about 50 μm; however, 75 μm is the practical limit of the technology and the average dimension of commercial processes are typically 100 μm and higher (Weigel *et al.* 2001). Laser cutting speeds vary depending on the power of the laser, the speed of the equipment used, and the thickness and material properties of the substrates to be cut. One limitation of CO_2 lasers and laminates is that they primarily melt the polymer film and so a thin ridge of material is often left at the edge of the cut, occasionally making bonding or precise manufacturing difficult.

1.12.3.4.2.(ii) Other laminate patterning techniques As mentioned previously, there are other methods used to cut and pattern polymer laminate cards. Xurography is a method that uses a sharp blade controlled by an automated knife plotter to remove structures from a sheet of polymer plastic (Bartholomeusz *et al.* 2005). This technology is equally useful for rapid prototyping; however, it is a slower patterning process, which can possibly deform the material at the edge of cuts and does not have such small feature sizes or tolerances. For more information on this technology, refer to Section 1.12.3.5.

Another method that is highly practical in mass producing laminate structures is stamp cutting. This can be done by forming sharp cookie cutter type structures to cut out the desired features in the laminate by applying a high pressure. Although stamp cutting is a low-cost method of mass producing layers, it is not as useful for creating quick prototypes and for varying patterns. The possible feature sizes and tolerances using stamp cutting techniques are limited by the stamp's fabrication methods, the material thickness, and the complexity of the design to be cut out. Other methods and materials can be used to fabricate layered microfluidic structures such as powder blasting, molding, and others such as is common in patterning PDMS, SU-8, and dry film photoresists mentioned previously.

1.12.3.4.3 Laminate card assembly

Each laminate layer that is to be used to produce a 3D microfluidic device or lab card is patterned one at a time. Once the 2D patterns are designed in each layer, the patterned layers are cleaned and the release layers are removed to expose the adhesive. This layer is then brought into contact with another polymer layer and pressure is applied to bond them together. Each polyester layer can vary as to whether or not it will have adhesive coatings on them. Typically, the assembly is alternated with one layer having adhesive on both sides of the polyester sheet and the next layer having a patterned polyester sheet with no adhesive layers on either side, as illustrated in **Figure 8** with the assembly of a 17-layer lab card (Weigl *et al.* 2003). This helps reduce the thickness of adhesive layers between each polymer layer and simplifies purchased material selections. Layers can also be bonded using a thermal bonding process or UV adhesives. Devices fabricated using this technology typically have depth tolerance of less than 1 μm and a width tolerance of about 10 μm. Typical channel dimensions of these devices are 100–3000 μm in width and 50–400 μm in depth (Weigl *et al.* 2001).

In addition to the design of each layer forming the microfluidic device, numbers and alignment holes are cut out to properly align each card during the assembly stage. The patterned layers are then aligned with each other in their proper sequence, as shown in **Figure 9**. Typical alignment tolerances are in the 1- to 30-μm range depending on how tightly the alignment marks fit with the guide rods. Once the entire 3D device is put together, it is tightly pressed to ensure a sound adhesive bond. In some cases the plain laminate layers can be chemically treated (e.g., with oxygen plasma) to change their wettability, thus modifying the internal surfaces of the channels. Other materials like PDMS or glass can be fabricated and aligned using their typical micromachining processes.

Not only can 3D lab cards be fabricated from multiple layers of the same material, but hybrid circuit structures and varying materials can be used as well. For example, one or two layers of thin Mylar can be sandwiched between glass slides for added structural integrity or optical properties. Polymers, metals, glass, silicon, and dry stored chemical compounds can be incorporated into the devices during assembly. For example, dry stored reagents and dyes can be stored in small cavities inside the lab card where it can be rehydrated and used during one of its operation steps (Weigl *et al.* 2001). Another example of the added benefit of assembling MEMS devices as laminate layers is that necessary interconnects and packaging components can be inserted between the last and second-to-last layers of the card, eliminating the need of manifolds and adhesive bonds for packaging.

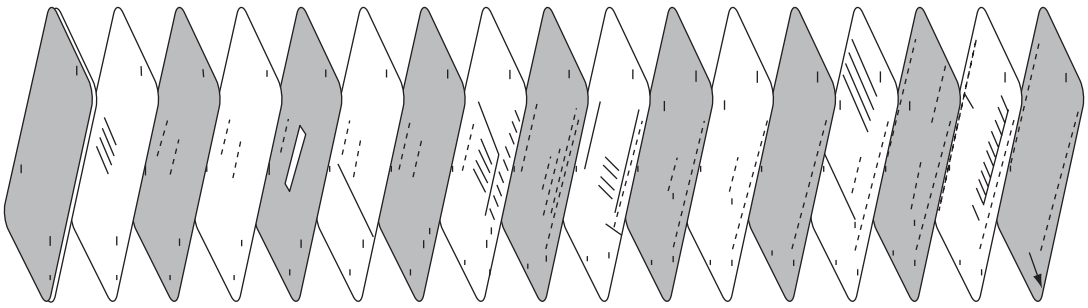


Figure 8 Illustration of how 17 layers of patterned Mylar cards are assembled in alternating layers, the light-colored cards have double-sided adhesive on them and the dark-colored cards have no adhesive on them. (Source: Weigl B H, Bardell R L, Schulte T, Battrell F, Hayenga J 2001 Design and rapid prototyping of thin-film laminate-based microfluidic devices. *J. Biomed. Microdevices* **3(4)**, 267–74. ©2001, reproduced with permission from Elsevier.)

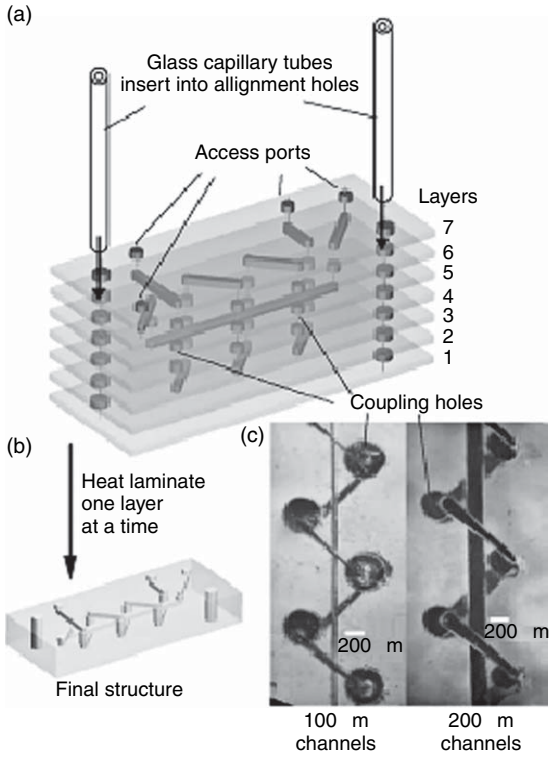


Figure 9 (a) Each polyester layer is aligned in their proper sequence and small glass capillary tubes or other structures are inserted to keep each layer in alignment. (b) The layers are pressed together and the alignment tubes are removed. (c) A picture of the finished card with dye running through 100 and 200-μm channels for visualization. (Source: Bartholomeusz D A, Boutte R W, Andrade J D 2005 Xurography: Rapid prototyping of microstructures using a cutting plotter. *J. Microelectromech. Syst.* **14(6)**, 1364–74. Figure ©2005, reproduced with permission from IEEE.)

Because of the speed and simplicity of patterning and assembling Mylar or other plastic laminate layers using a CO_2 laser, complex 3D lab cards can be very quickly assembled from simple 2D layers. The entire process of moving from a CAD design to a complete 3D finished product can take place in less than 4 h, as opposed to days or weeks using other prototyping and fabrication methods. Micronics, a company heavily involved in this technology, uses a class 10 000 clean room facility and is capable of producing over 100 000 laminate devices annually, with an average production of 40 000 new prototype cards per year.

1.12.3.4.4 Applications and examples of laminate technologies

Laminate methods have been used to create a wide variety of microfluidic devices. Micronics has developed a microcytometer card, a T-sensor card, and an H-filter card (Weigl *et al.* 2001). The 3D laminate approach has been used to develop devices that enable the analysis of whole blood using 3D fluid control and on-chip sample preparation.

The ability to easily incorporate hybrid elements into laminate cards allows a variety of applications. Applications for this technology include being able to develop passively driven point-of-care medical diagnostic devices for use in third world countries and remote areas. This can be achieved by incorporating reagents and analysis tools on chips or in hand-held devices. Another example of how hybrid elements can be used in laminate chips includes components to fabricate ferrofluidic micropumps. A ferrofluidic plug is placed inside the chip and combined with valves and externally moving magnets as seen in **Figure 10**, and a small pump is made for pumping highly controllable flow rates (Yamahata *et al.* 2003).

Owing to the 3D layout of laminate cards, gravity, absorption, and capillary action can all be used as pumping mechanisms to allow for passively driven microfluidic devices. They can also function as actively driven devices by easily interfacing with manifolds or other devices for portable point-of-care diagnostic devices. Because of the inexpensive nature of the materials used and the rapid fabrication processes achievable using CO_2 laser cutting techniques or stamping, the lab cards are well suited as disposable products and are convenient for research and development applications.

1.12.3.4.5 Limitations of laminate technologies

There are several limitations associated with the laminate approach to MEMS and microfluidics. Since a CO_2 laser is being used to cut and pattern each laminate layer, typical laser limitations will still exist. One is that the laser beam can only be focused small, so a limit of about 50 μm to about 75 μm applies to any features. In addition, the use of lasers will cause striations in the sidewalls of the cut material, as noted by Yilbas (1997), as well as bumps next to the cut wall. Striation effects will vary depending on the power of the laser and the speed of the cut; as a general rule, the slower the cutting speed the larger the striation widths. In addition to feature size limits and striation effects, tolerances in the cutting and alignment processes can be problematic. If decent laser patterning equipment and effective alignment techniques are used for patterning and assembly, finished laminate cards can have width tolerances as

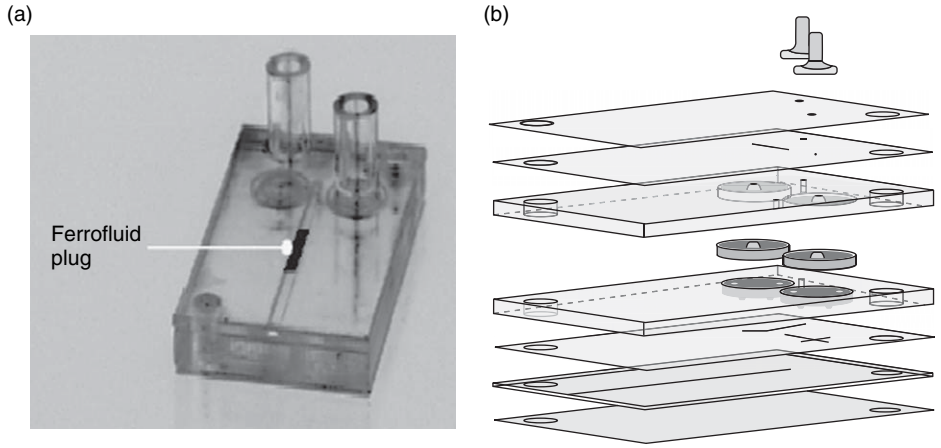


Figure 10 (a) Photograph of an assembled ferrofluidic micropump constituted of seven layers of polymethylmethacrylate (PMMA), two silicone check valves, and a ferrofluidic plug. (b) Burst view of seven PMMA layers with other microcomponents to form the ferrofluidic pump. (Source: Yamahata C, Chastellain M, Parashar K, Petri A, Hofmann H, Gijs M A M 2005 Plastic micropump with ferrofluidic actuation. *J. Microelectromech. Syst.* **14(1)**, 96–102. Figure © 2005, reproduced with permission from IEEE.)

low as 10 μm and depth tolerances of less than 1 μm (Weigl *et al.* 2001). Limitations on striations, feature sizes, and tolerances can be improved using photopatterned techniques such as SU-8, and dry film photopolymers, but these methods are also slower and more expensive.

Another inherent limitation of laminate fabrication is the bonding strength and feature spacing limits when bonding layers together. Although different adhesives and bonding methods have their own advantages and disadvantages, the more the number of layers to be assembled together, the more the chance of a bond failure. Other limitations of laminate MEMS include the material properties of the different materials used. Problems such as thermal stability, optical properties, rigidity, strength, sensitivity to light or long-term storage, varying thermal coefficients of expansion and reactivity of the materials with solvents and solutions are also present. In addition, when combining hybrid elements having different material properties, stresses and strains may occur and fabrication times will likely increase.

1.12.3.4.6 Summary for laminate technologies

Complex 3D lab-on-a-chip cards can be quickly and easily developed by stacking multiple layers of patterned materials. Using laser cutting techniques, each layer can be rapidly patterned without the need of photolithography masks, molds, or other time-consuming techniques. Most polymer plastic materials are suitable for a variety of applications and can be used for fluorescence detection and other optical detection methods due to its transparency. A fabrication setup can be purchased at a relatively low cost and prototyping and fabrication can take place very quickly.

By taking advantage of 3D structures, passively driven devices can be made to operate by themselves or actively driven devices can be made to interface with other equipment. Hybrid elements and multiple materials can be easily incorporated into the lab cards design without extensive design modifications and time-consuming processes. Surface modifications of the internal laminate layers can be made using chemical treatments such as oxygen plasma treatments.

Limitations include striation patterns and deformations in laser cut edges, dimensional tolerance of about 10 μm and feature size limits of about 50 μm, material property limitations, adhesive bond limitations, and minimal feature spacing limitations of about 100 μm. The tolerances and minimal design features given, however, are acceptable for most microfluidic applications. In addition, useful devices can be made at a low cost, making them convenient for disposable lab cards and affordable prototyping cards. Overall, this is a fast and affordable method of creating high-performance microfluidic devices made from inexpensive materials suitable for a variety of applications.

1.12.3.5 Xurography

Cutting plotters have been used in the graphic design industry for some time, generally cutting designs out

of adhesive-backed materials should be used in creating signs or other artwork. The cuts are made by dragging a knife blade through the material to be used. Once the cuts have been made, the undesired material is peeled away and discarded. A transfer tape is then applied to the top surface of the remaining material so that the design can be moved from its material backing to the desired substrate.

Many of these cutting plotters have good mechanical resolution with a step size down to 5 μm with some plotters. At the beginning of 2005, Treise *et al.* used a cutting plotter to fabricate planar microfluidic channels by creating the channel design out of a vinyl film using a cutting plotter and sandwiching the channel between a Plexiglas backing, with inlet and outlet holes, and an adhesive covering layer. Later in 2005, Bartholomeusz *et al.* extensively validated this microfabrication process with multiple materials and multiple fabrication techniques. This group found that microfeatures could be fabricated down to linewidths of approximately 20 μm with some materials. They coined this microfabrication method xurography, using the Greek root words *xuron* and *graphe*, which mean razor and writing, respectively. This prototyping method is intriguing because a microdevice prototype can have a very short turnaround time; microdevices can be designed, fabricated, and tested within a working day.

1.12.3.5.1 Dimensional capabilities of xurography

The feature sizes obtainable in xurography depend on several factors associated with a cutting plotter. The minimum step size of the cutting plotter, blade sharpness, blade angle, material properties of the substrate, cutting speed, cutting force, and cutting mode are all key factors that determine the smallest features possible and will be discussed in greater detail.

The minimum step size for cutting plotters is divided into two categories: mechanical resolution and addressable resolution. The mechanical resolution is defined as the smallest step size of the stepper motors and generally ranges from 5 to 25 μm, depending on the machine used. The resolution of the particular cutting plotter helps to define how small the features are that can be achieved. Most of the high-resolution cutting plotters have a mechanical resolution of 5 μm and can range in cost from about US$4000 to US$5000. The addressable resolution is defined as the programmable step size allowed by the software. In general, the smallest step size range that can be output is 10–25 μm.

The blade size and geometry plays a significant role in the ultimate resolution of xurography. Differing blade thicknesses may be purchased, generally either 0.9 or 1.5 mm. Thicker blades are typically used for thicker materials. The blade sharpness is also a critical factor in deciding final feature sizes. If the blade is dull the material may start to tear, resulting in rougher edges. Also, the thickness of the blade increases at the tip as the blade becomes more dull, which limits the size of features that can be generated (Arcona and Dow 1999). The blade angle also plays a role in feature size. The three most commonly used blade angles are 30°, 45°, and 60°, as measured from the surface being cut to the edge of the blade. These blades range in price from approximately US$15 to US$85 per blade, depending on the manufacturer, angle, and material of the blade. The 45° blade provides a good all-purpose cutting angle. The 30° blade has more of the blade immersed in the material while the 60° blade is minimally in contact with the material being cut, which needs to be taken into consideration when planning for overcut in the design.

Bartholomeusz *et al.* determined the equations necessary to define what qualities determine the minimum feature sizes capable with xurography and found that the material properties of the substrate being cut play a large role in the minimum feature sizes possible. A thinner, softer material, with low Poisson's ratio and a high shear stress at the adhesive interface, yields the smallest features.

The cutting plotter settings (cutting speed, cutting force, and cutting mode) are also important in determining minimum feature size. The cutting speed is usually set to the lowest speed, which is 1 cm s^{-1} for most cutting plotters. If the speed is too fast the features will not form well due to tearing and features may possibly be missed when there is a change in direction. It is also important to have the proper cutting force. It should be set high enough so that the material is completely cut but controlled so that the backing of the material is only slightly scored. If the backing is cut too much the blade will become dull, leading to poor device geometry.

The majority of high-resolution cutting plotters have three cutting modes. The first mode is the drag knife mode, which uses a swivel blade. This mode simply drags the knife through the material and uses the frictional forces to make turns in the cut. This mode can have problems in cutting features that have sharp turns and may break the blade at some turns if the material is hard enough. The second mode is the

true tangential cutting mode, which will cut at corners by lifting the blade completely out of the material, rotating the blade with a motor, then inserting the blade back into the material. This mode can perform overcuts to ensure the material is completely cut and is a good mode for thicker materials. The third mode, the emulated cutting mode, uses a swivel blade as well. However, it will lift the blade just to the top of the material and then pivot to make a turn in the cut and will then be lowered back into the material. This mode helps to protect the blade during turns. Overcuts can be performed with this mode as well, insuring that the material is completely cleaved.

1.12.3.5.2 Xurography applications

Xurography has proven capable of prototyping microfabricated devices by creating either positive or negative features. Prototyping methods have been used in shadow masks, electroplated microchannels, micromolding (MM), and laminated microfluidic structures; another group more recently developed a variation on laminates, producing microchannels using a double-layer adhesive material. These applications will be discussed in greater detail.

Electroplated microchannels take advantage of Rubylith® material as a sacrificial layer. The material is patterned with the cutting plotter and the undesired material is peeled away. This pattern is then transferred to a substrate, which then undergoes sputtering of a titanium seed layer followed by a gold seed layer. The substrate channel ends are covered in Kapton tape, in order to prevent the inlets of the channel from being coated in metal, and then placed in an electroplating bath, depositing nickel onto the gold seed layer. Once the nickel metal is deposited, the Rubylith® is then dissolved with acetone, leaving hollow nickel channels behind.

MM over a xurography-created pattern can be done by creating positive features with a cutting plotter out of a thermal laminate or an adhesive-backed material. This material is then transferred and adhered to a dish that PDMS can be poured into, thus creating a mold. The dish and features are covered with a silane coating to prevent the PDMS from sticking to the mold. PDMS is then poured into the dish, degassed in a vacuum chamber, and cured in an oven. Once cured the PDMS is peeled from the mold, leaving microfeatures in the PDMS (Kim and Gale 2005).

Laminated microfluidic structures can be used to create 3D prototypes. This is done by cutting parts of the channel and access ports on single slices, with alignment holes cut on opposing corners of each layer. Several of these layers can be stacked together, aligned by sticking glass capillary tubes through the alignment holes, and then adhered together by applying heat and pressure (Bartholomeusz *et al.* 2005).

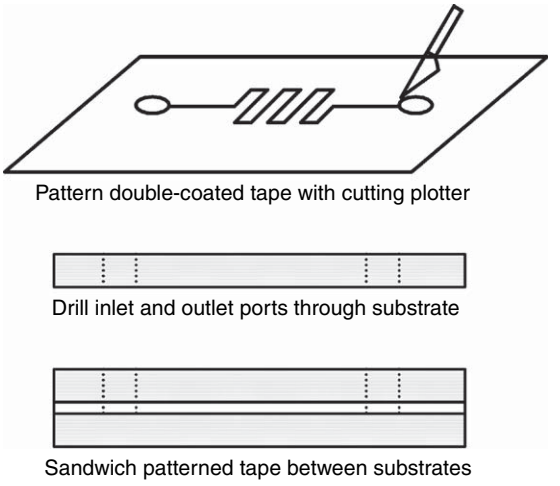


Figure 11 Microchannel fabrication process using a double-coated tape. The double-coated tape is patterned with the cutting plotter and the undesired material is peeled away. Holes are then drilled through a substrate to create inlet/outlet ports for the microchannel assembly. Lastly, the patterned double-coated tape is sandwiched between two substrates, one that has been drilled and one that has not, to seal the microchannel. (Source: Bartholomeusz D A, Boutte R W, Andrade J D 2005 Xurography: Rapid prototyping of microstructures using a cutting plotter. *J. Microelectromech. Syst.* **14(6)**, 1364–74. © 2005, reproduced with permission from IEEE.)

A double-layer adhesive material patterned with a cutting plotter and sandwiched between two subtrates for channel fabrication has been used as a very quick and simple method for microfluidics. This method involves cutting the desired channel dimensions and inlet/outlet ports out of a double-layer adhesive material using the cutting plotter. One of the substrates to be used has holes drilled at the inlet and outlet port locations for access to the channel. The double-layer adhesive material is then sandwiched between the two substrates, thus creating a microchannel (Sundberg *et al.* 2006). **Figure 11** shows the fabrication process for these microchannels.

1.12.4 Low-Cost Methods for Rapid Prototyping

In many instances, low-cost, proof-of-concept devices are the objective and more expensive techniques for further prototyping or mass production can be used

later if concepts can be demonstrated early on. Traditional MEMS methods can be extremely slow and expensive for prototyping due to the high fixed costs associated with the tooling and the difficulties in developing new processes on these tools. Outsourcing using traditional MEMS tools, which may be more cost-effective, usually has lead times of 6 months or more, which can be problematic as new concepts are being explored and there is the risk of failure in early attempts. Thus, a range of rapid prototyping techniques for MEMS devices have been developed to provide proof-of-concept before the more expensive and slow processes are applied. Note that most of these rapid, inexpensive processes are designed to implement mechanical functions, as most electrical processes are best implemented in traditional silicon and best prototyped using computer simulations.

One of the best processes for rapid prototyping in MEMS and microfluidics applications is soft lithography, which is discussed at length in its myriad forms. Other techniques for inexpensive rapid prototyping include powder blasting, photopatternable polymers (Section 1.12.2.2.1), xurography (Section 1.12.3.5), and laminate technologies (Section 1.12.3.3), the latter three have already been discussed.

1.12.4.1 Soft Lithography

Soft lithography involves the use of soft polymers that are molded over molds typically produced using photolithography. The earliest version of soft lithography, μCP (Xia and Whitesides 1998), then allows for that pattern to be replicated many times by stamping materials onto a substrate (thus the term soft lithography, Xia and Whitesides 1998, Xia *et al.* 1999). In more advanced versions of soft lithography, the molded polymers can be used as mechanical devices, most typically in microfluidics. A number of other variations of soft lithography have developed, and each will be discussed. The major advantage of soft lithography is the simplicity of the process that makes it easy for the chemist, biologist, and material scientist who want to apply these advances to their own research. The fabrication processes used to create plastic devices using soft lithography are based on replication (casting, embossing, or injection molding) and are faster and less expensive than those used on glass and silicon. The low cost of both the raw materials and the processes associated with soft lithography allows the production of potentially disposable devices. The soft polymers used with soft lithography tend to produce devices that are robust and easy to handle, which is not the case with glass and silicon microdevices. Soft lithography is also applicable to nonplanar surfaces and 3D microsystems, which is typically not the case with more traditional microfabrication techniques (Gates 2004, Xia and Whitesides 1998).

With polymers, researchers have been able to fabricate high-quality microsystems, even those on the nanoscale. We call these unconventional fabrications softlithography due to the use of a soft material and the ability to generate micropattens.

Generally, the soft lithography procedure starts by making a mold structure based on the initial idea using traditional lithography or other methods. A soft polymer is then poured over the mold and cured to produce a polymer part. The most commonly used polymer in soft lithography is PDMS as it has a number of useful properties such as reasonable strength, optical clarity, and short cure times; moreover, it has the capability of conforming to even nanoscale patterns (Roger and Nuzza 2005). Usually soft lithography is organized into two categories depending on the pattern transfer method: MM and μCP. MM includes replica molding (RM), microtransfer molding (μTM), micromolding in capillaries (MIMIC), and solvent-assisted micromolding (SAMIM). μCP is used for elastomeric stamping in order to transfer the pattern onto a substrate (Gates 2004, Xia and Whitesides 1998, Xia *et al.* 1999). In this section, the main elements of soft lithography and the primary methods of soft lithography are introduced along with a review of recent applications.

1.12.4.1.1 Material for soft lithography

Soft lithography relies on soft polymers that are appropriate for stamping, molding, and embossing. Those polymers that have a low interfacial energy, that are chemically inert, that have a high thermal stability, and that are easily modified appear to be the best candidates. Many research groups have explored a variety of materials for soft lithography including materials such as polyimide, polyurethanes, Novolactm™ resins, and PDMS, which is the most commonly used.

1.12.4.1.1.(i) PDMS material properties PDMS has several useful material properties for stamping and molding. It provides a surface that has a low interfacial free energy, is chemically inert, has good gas permeability and good thermal stability, and is

optically transparent. The surface properties of PDMS are also relatively easy to modify. The commercially available PDMS most commonly used is Sylgard™ 184 produced by Dow Corning. It consists of two parts – a silicon base and a curing agent. For basic soft lithography, the base and the curing agent are mixed well and poured over the master mold. To remove the bubbles in the liquid-state PDMS, the mold with the liquid PDMS is put into a vacuum chamber until the bubbles are completely removed. The PDMS can then be cured at room temperature (~2 days) or at elevated temperatures (a few hours). The physical properties of Sylgard™ 184-PDMS include a high tensile modulus (1.8 MPa) and high physical toughness (4.77 MPa), low shrinkage after curing (1.1%), and high elongation (160%). PDMS also has a low free surface energy ($\gamma = 21.6\,\mathrm{dyn\,cm^{-2}}$) that makes it possible to release the PDMS structures easily from the mold. The low thermal conductivity ($k = 0.2\,\mathrm{W\,m^{-1}\,K^{-1}}$) allows PDMS to be used as a thermal isolation material. PDMS has a thermal expansion coefficient of $310\,\mu\mathrm{m\,m^{-1}\,°C^{-1}}$ and is thermally stable below 150 °C. (Clarson and Semlyen 1993, Dow Corning Co. 2005, McDonald and Whitesides 2002).

The gas permeability of PDMS and the impermeability of PDMS to liquids are also useful characteristics for some chemical applications, especially in microfluidics. For example, a unique gas permeation pump has been developed by using these properties (Eddings and Gale 2006). PDMS also has good biocompatibility properties and it is possible to implant within the body and to make the cell culture chips because of its nontoxic properties (El-Ali *et al.* 2006).

The fabrication of nanoscale features has to be performed carefully with PDMS. One reason is the difference in the thermal expansion coefficient between the PDMS and the mold. This difference can cause distortion of the replicated features. The other reason for distortion is shrinkage during the curing of the prepolymer. To combat these drawbacks during the fabrication of nanoscale features, no solvent evaporation can occur and a low-temperature curing environment needs to be used. During the peel-off step (removal from the mold), distortion of fabricated structures may also occur. Also, for μCP, the stamping feature needs to have high mechanical stability to prevent the collapse of the feature, so it is important to consider the stamping geometry. To overcome these potential problems, different versions of PDMS have been developed including harder PDMS and UV-curable PDMS. With these modified PDMS, features, aspect ratios >4:1 are possible without collapsing (Quake *et al.* 2000, Roger and Nuzzo 2005).

1.12.4.1.1.(ii) Surface modification of PDMS

While the bulk properties of PDMS make it generally useful, the surface properties of PDMS are often difficult to control and pose challenges when working with many liquids and proteins. PDMS is strongly hydrophobic, which can make filling reservoirs difficult and can cause channels holding aqueous solutions to collect bubbles since the solutions dewet from the surface. PDMS structures can also become distorted when they contact organic solvents and hydrophobic analytes. The low surface energy of PDMS also makes it difficult to deposit other materials on PDMS (though PDMS does bond well to itself). For applications that need surface electrodes to produce an electric field, such as in capillary electrophoresis (CE), in electrochromatography, and in electroosmotic pumps, electrode deposition can be problematic. The PDMS surface is also highly active in that the molecules at the surface are continuously being replaced by PDMS in the bulk, making long-term control of PDMS surface properties difficult. To overcome some of these challenges, several techniques for modifying the surface of PDMS structures have been developed.

The surface of PDMS is highly inert; therefore, surface modification can be difficult to achieve. Nevertheless, several surface coating approaches have been developed including plasma, covalent, and dynamic modification methods.

Oxygen plasmas, which are composed of ionized oxygen at high energy, have been used to create hydrophilic PDMS surface through oxidation. Oxygen plasma can be used to form silanol groups (–OH) at the end of methyl groups ($-CH_3$). As the silanol groups are polar, the treated surface is easily wettable with polar solvents. This method is commonly used to treat the surface of PDMS because the process can be completed in only a few minutes, it does not require the use of chemicals, and it has good repeatability and consistency. With this treatment, PDMS can be used to implement the electroosmotic flow and bonding between molded PDMS and flat PDMS/glass. Unfortunately though, oxygen plasma treatments cannot be used for microsystems that require long-term stability. Usually, the plasma treatment is effective for less than 3 h in the

presence of aqueous solutions (Duffy *et al.* 1998, Makamba *et al.* 2003).

To overcome the mobility of the surface in PDMS, static coating, such as by covalently linking ionizable molecules to PDMS monomers on the surface, was developed for permanent PDMS surface modification. To bind these ionic polymers to PDMS, UV graft polymerization, silanization, or CVD is usually used. UV graft polymerization is used to create reactive sites (radicals) on the PDMS surface, which otherwise has no functional group on the surface. UV or ionizing radiation can generate free radicals that can be used as the initiation site for a polymeric chain on the surface of PDMS. UV graft polymerization has been shown to produce stable surface properties and is relatively easy to perform; the surface polymers have low penetration into the bulk PDMS. Silanization has been used to make an immobilized polyethylene glycol (PEG) layer on the surface of PDMS. Before stacking the PEG layer on the PDMS, the PDMS surface is oxidized in oxygen plasma and immersed into Si–PEG–Si solution. This technique is usually used to inhibit protein adsorption on the surface of PDMS. CVD can be used to deposit thin films of various materials and can also be used to modify the surface of PDMS. This approach was used to deposit reactive coatings of poly(*p*-xylylene carboxylic acid pentafluorophenolester-co-*p*-xylylene) (PPX-PPF) and poly(*p*-xylylene-2,3-dicarboxylic acid anhydride). Without the need for further activation, the high chemical reactivity of these functional groups supported conversion with biological ligands or proteins and was used for surface patterning of PDMS using μCP (Barbier *et al.* 2006, Hu *et al.* 2002, 2003, Lahann *et al.* 2003, Slentz *et al.* 2002, Xia *et al.* 2004). Another surface modification method is the use of a surfactant, which has a hydrophobic tail that can adsorb to the hydrophobic PDMS surfaces easily. This hydrophobic tail interacts strongly with the PDMS, while the charged head stretches out of the surface and changes the surface charge density. Neutral poly(oxyethylene)-based surfactants and cationic surfactants are usually used to modify the surface of PDMS (Badal *et al.* 2002).

1.12.4.1.2 Microcontact printing

μCP is the most important patterning technique among the soft lithography methods. In μCP printing, like in conventional printing techniques, the patterned stamp is brought into contact with a substrate to transfer an ink, a solution of an alkanethiol, or other molecules to a surface modified substrate and create patterns on the substrate, as shown in **Figure 12**. The technique was developed to deposit arrays of self-assembled monolayers (SAMs) within a few seconds (Jackman *et al.* 1995). A precisely arranged monolayer builds up rapidly at the area of contact (Gates 2004, Xia and Whitesides 1998, Xia *et al.* 1999). μCP of SAMs on a substrate as a resist layer or protective layer from etchants has been demonstrated as a rapid photolithography replacement technique (Geissler *et al.* 2002). Bernard *et al.* (1998) were able to pattern distinct regions of antibodies and other proteins on polystyrene substrates for further use in binding assays.

Because of the complexity of working with biomolecules, several considerations must be understood when selecting the solution to be patterned. Biomolecule solutions that will not contribute to PDMS swelling should be employed to retain the patterned resolution on the molded PDMS stamp head. Solvents that will diffuse or that have a high affinity for PDMS rather than the substrate should also be avoided to promote uniform, concentrated

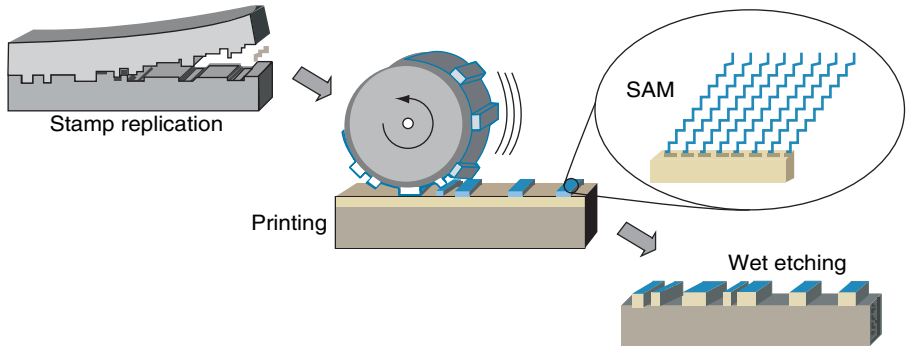


Figure 12 The microcontact printing procedure with an elastomeric stamp. SAM, self-assembled monolayer. (Source: http://www.research.philips.com/technologies; reproduced with permission.)

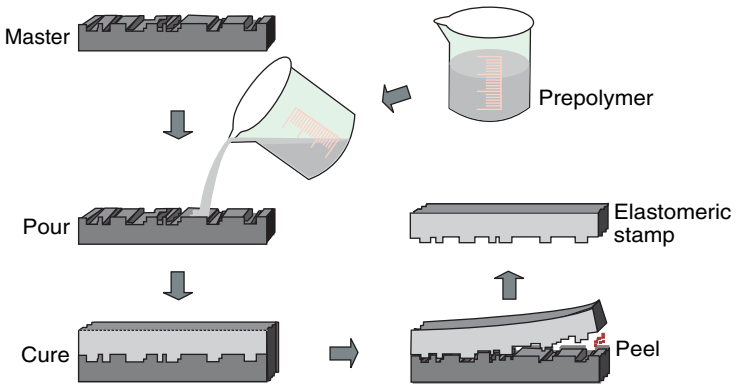


Figure 13 The stamp replication process. After curing the polymer, the elastomeric stamp is peeled off the master and uses the master for microcontact printing. (Source: http://www.research.philips.com/technologies; reproduced with permission.)

printing. Other important criteria include low vapor pressures at ambient conditions and narrow diffusion characteristics, which will assist in printing resolution and biomolecule stability (Delamarche *et al.* 2003).

The stamp for μCP is usually made from an elastomeric polymer PDMS. This material allows for conformal contact with the substrate combined with advantageous chemical and physical properties important for the molecule transfer behavior. Stamps are fabricated by casting prepolymers on a master with a negative of the desired pattern, curing it, and peeling the cured stamp off the master as shown in **Figure 13**. Depending on the stamping method, the technique is referred to as planar stamping or rolling stamping. One particular advantage of μCP is its ability to print on nonplanar surfaces with a modified planar stamp (Jackman *et al.* 1995). In roller stamping, altering the configuration slightly by curving the PDMS stamp can also achieve unique circular patterns (Franssila 2004).

The resolution of μCP is generally as good as the method used to make the mold, since μCP can replicate even nanoscale features. Recently, the resolution of μCP has been improved by solving problems such as surface diffusion, disorder at the edges of the printed SAMs, and the isotropic nature of many of the etching and deposition methods. These problems are eliminated by the use of what is called nanotransfer printing (nTP). In nTP the SAMs are used as covalent glues and release layers for transferring material from relief features on a stamp to a substrate. This approach is a purely additive technique that can generate complex patterns of single or multiple layers of functional materials with nanometer resolution over large areas in a single process step (Loo *et al.* 2002a, b).

An interesting development in μCP is to use PDMS itself as the ink material. This method, which is referred to as decal-transfer printing, utilizes a PDMS stamp to transfer other structures (such as membranes) made of PDMS. Coating the handle PDMS stamp with an organic material aids the release of the carried structures (Childs and Nuzzo 2002, Childs *et al.* 2005).

μCP can be used to pattern other microdevices. μCP has been used to make a charge pattern on thin electrets (e.g., polymethylmethacrylate (PMMA)). This technique can pattern regions of charge with lateral dimensions down to at least 100 nm (Jacobs and Whitesides 2001). μCP can be used to make patterns within microchannels as well. For example, selective adsorption of fibronectin and bovine serum albumin onto the patterned microfluidic channels and the deposition of various proteins within multiple patterns using laminar flow have been accomplished using μCP (Khademhosseini *et al.* 2004). Dozens of other applications of μCP and related techniques have been outlined in other review articles (Gates *et al.* 2004).

1.12.4.1.3 Micromolding with soft lithography

Several variations of MM using soft lithography exist including RM, μTM, MM in capillaries, and solvent-assisted MM.

1.12.4.1.3.(i) Replica micromolding MM is used to create useful structures in the soft polymers associated with soft lithography. The molds are then sometimes used to create structures in other materials. To create the PDMS mold or structures, any micromaching technology can be used to construct the pattern, but traditional silicon etching is often used. After making the original mold out of silicon or any other material, the surfaces of the patterned silicon

must be coated with silane to allow easy separation between PDMS and the patterned silicon. The prepolymer is coated on the patterned silicon and then polymerized. After curing the prepolymer, the polymer is peeled off from the master and can be used for further processing steps or used directly.

SU-8 is also used regularly to make master replica molds for PDMS molding (Mata *et al.* 2006). More information on using SU-8 to make these molds is available in Section 1.12.2.2.1.(ii). After completion of the SU-8 lithography, the SU-8 master mold has to be coated with silane to allow the molded polymer to peel off easily after curing. The prepolymer can then be cast over the SU-8 master, cured, and released. This approach has been used to produce structures under 100 nm in size.

Once a polymer (PDMS) mold has been generated, it can then be used to produce multiple copies of the original mold. The PDMS mold can be used to make multiple PDMS copies or other materials can be used. Generally, the prepolymer is not able to penetrate into PDMS and each PDMS mold can be used about 20 times before the molds begin to fail (Gates 2004). Using the RM method, it is possible to make copy nanostructures with high resolution (even copying a protein structure on the surface of the mold). For example, structures as small as 1.5 nm were replicated with a PDMS mold (Gates and Whiteside 2003). To accomplish this feat in a practical way, the feature is fabricated with a focused electron beam in PMMA and replicated in a PDMS mold. Then, polyurethane (PU) can be used to replicate the original structure using the PDMS mold. Thus, this method can replicate nanopatterns over large areas rapidly and at low cost.

1.12.4.1.3.(ii) Microtransfer molding In μTM, the elastomeric mold is filled with a polymer precursor, and the excess prepolymer is removed (**Figure 14**). The soft mold is pushed up against a substrate, the patterned feature in the soft stamp is cured while in contact with the substrate by appropriate ways, and the stamp is peeled off. By repeating this procedure, a multilayer structure can be created on a single substrate.

Compared with μCP, μTM can produce thick and multiple layers easily. But, μTM requires the use of a viscous prepolymer that consumes more time for curing and for filling the capillaries of the patterns. To speed up the process, hydrophobic dendrimers (G4 aldehyde, CHO end group) can be used to fill the capillaries of the stamp within 30 s. By controlling the concentration of dendrimers in the solution inking the PDMS stamp, minimum feature sizes of 40 nm can be achieved (Thibault *et al.* 2006). The creation of 3D structures can also be completed in a single molding step called membrane-assisted μTM. In this process, a technique such as multiphoton absorption polymerization is used to create a master structure that has closed loops that are interrupted by thin membranes. After creating the loop structure with the MAP technique, the mold is immersed in PDMS to make the mold for μTM. The PDMS is cured and the PDMS mold is peeled off from the substrate. The PDMS mold is filled with molding material and the pattern is transferred to a substrate. While the PDMS mold contacts the substrate, the material being molded is cured and the mold is removed (LaFratta *et al.* 2006).

1.12.4.1.3.(iii) Micromolding in capillaries In MIMIC, a soft mold like those made with PDMS is contacted with the surface of a substrate to form a network of open microchannels (**Figure 14**). The soft stamp is pressed onto the substrate evenly and a prepolymer is applied to access holes in the mold. This process can be vacuum assisted to speed flow into the holes. The prepolymer flows into the cavities formed by the soft mold on the substrate and fills the cavities automatically due to capillary forces. After curing and peeling it from the soft mold, the polymer structure remains on the surface of substrate. For long capillaries, the rate of filling decreases significantly owing to the viscous drag of the fluid in the capillary and the distance over which the fluid has to be transported, which should be considered during the design phase. The rate of filling also decreases as the cross-sectional dimension of the capillary decreases and as the interfacial free energy of the surface decreases (Gates 2004, Xia and Whitesides 1998, Xia *et al.* 2003). Recently, a preceramic polymer that is transparent, solvent-resistant, thermally stable, and biocompatible was fabricated using MIMIC. Transparent poly(silazane) glass-derived microchannels and structures can be fabricated based on this technique. The structures may be useful in the field of micro-total-analysis systems (Asthana *et al.* 2006).

1.12.4.1.3.(iv) Solvent-assisted micromolding SAMIN combines RM and embossing concepts simultaneously. SAMIN uses an elastomeric PDMS mold in combination with an appropriate solvent (**Figure 14**) – instead of a rigid mold and high temperatures and pressures – to emboss the polymer.

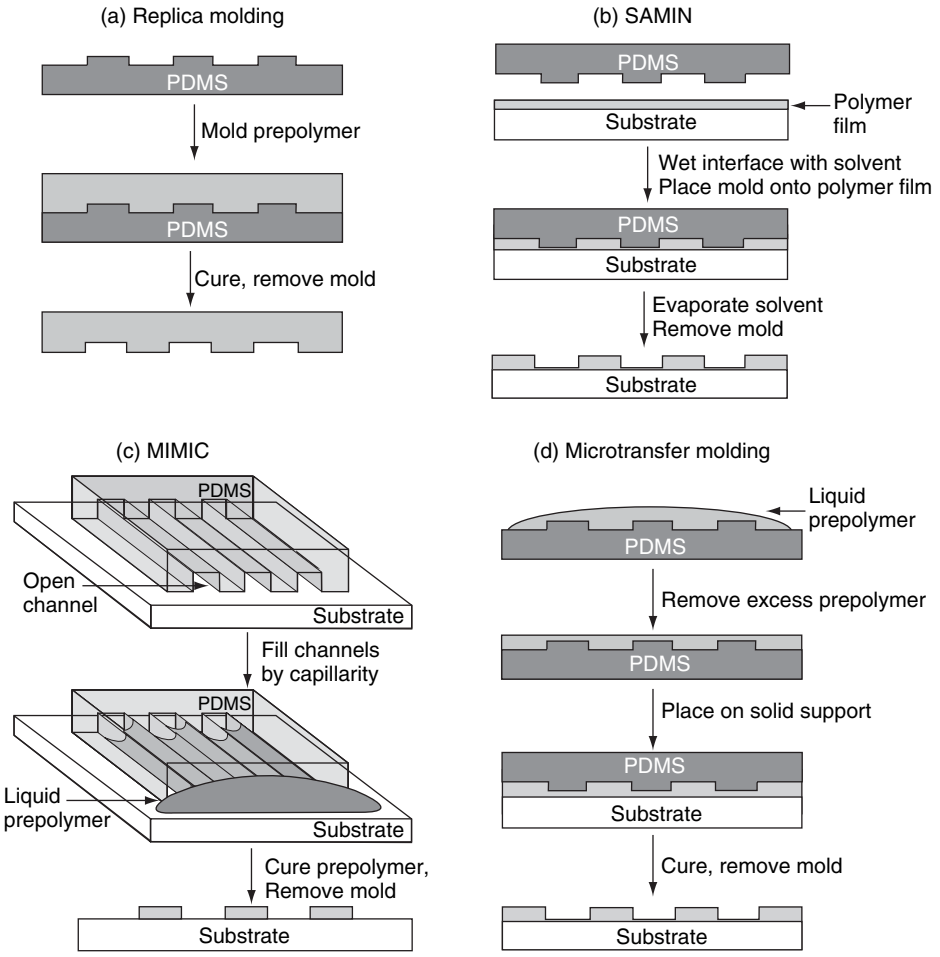


Figure 14 A schematic of four different micromolding techniques. (Source: Gates B D, Xu Q, Love J C, Wolfe D B, Whitesides G W 2004 Unconventional nanofabrication. *Annu. Rev. Mater. Res.* **34**, 339–72. © 2004, reproduced with permission from Annual Reviews, www.annualreviews.org.)

The solvent is either applied to the PDMS mold or retained in the polymer film before placing the two surfaces in contact. A small amount of solvent is dispensed on a patterned PDMS stamp and the stamp is placed on the polymer to be molded. The solvent causes swelling in the polymer and causes expansion of the polymer into the patterns of the stamp. The gas-permeable mold prevents trapping of air pockets and nonuniform solvent evaporation. The mold is then removed to reveal a relief structure complementary to the topography of the mold. For this technique, the wetting of the PDMS mold by the solvent and conformal contact between the PDMS mold and the substrate are the most important factors for successful fabrication (Gates 2004). The application of SAMIN to nanoscale features has been limited primarily by the lack of appropriate masters rather than by the fundamental characteristics of this process. To overcome this limitation, composite elastomeric mold like hard PDMS can be used to make smaller features (~20 nm) (Odom *et al.* 2002).

A limitation of elastomeric PDMS mold is that swelling may occur when PDMS is used with organic solvents. This swelling causes distortion of the pattern in soft lithography and may lead to pattern collapse in the mold. Surface modification of the PDMS mold with poly(urethaneacrylate) has been shown to reduce the rate of solvent absorption into PDMS (Lee *et al.* 2006).

1.12.4.1.4 Packaging of components fabricated using soft lithography

In microfabrication, packaging is defined as all the integration and connection steps after fabrication of the fundamental device is complete. It includes postprocessing release, package/substrate fabrication,

assembly, testing, and reliability assurance. The package provides functional interfaces between microdevices and the environment (Lee *et al.* 2003). These interfaces are directly related to the application. Thus, unique packaging approaches have been developed for use with soft lithography.

1.12.4.1.4.(i) Bonding One of the most basic packaging steps is bonding. Two PDMS substrates can be bonded to each other using a variety of techniques. The most common technique for bonding PDMS substrates is the use of an oxygen plasma to oxidize the surface followed by rapidly placing the two substrates in contact with one another. The seal between the two pieces of PDMS can be sufficiently strong such that the two substrates cannot be peeled apart without failure in the bulk PDMS. The bond strength has been found to be highly dependent on the measured contact angle of the PDMS after exposure to the oxygen plasma. The lower the contact angle, the better the bond (Xie *et al.* 2005). As a result, a contact angle below 5° provides a good bond strength (Bhattacharya *et al.* 2005). An identical process can be used to bond PDMS to glass substrates.

Several other methods for bonding PDMS are available. For example, placing two partially cured substrates (or even if only one is partially cured) in contact before final curing results in a good bond strength between PDMS substrates. These films of uncured PDMS can also be placed between layers of cured PDMS to produce a high-quality bond.

Rigid polymers like PMMA are often used in making rigid molds or stamping structures for soft lithography. To package micro- or nanofabricated structures with PMMA, thermal bonding, lamination, adhesives and solvent-assisted bonding are usually used. Surface modification such as air plasma treatment, hydrolysis and aminolysis are also available to assist in developing strong bonds. A comparison of techniques and a determination of the optimal methods for bonding PMMA without damage to the microstructures is available (Brown *et al.* 2006).

1.12.4.1.4.(ii) Microfluidic packaging In microfluidic systems made by RM, the channel inlets and outlets often need to be connected to pumps or other fluid components. Several methods have been used to accomplish this task. The most basic are glued interfaces, where connecting tubing is attached with an adhesive (often PDMS) to the channels. One method for use with SU-8 master molds requires that an inlet and an outlet post structure be created after fabrication of the microchannel structure on the silicon wafer. Essentially, the packaging channels are molded into the PDMS. With this mold, the PDMS replica can be peeled off and connections with a syringe needle can be made. Usually, oxygen-reactive ion etching and epoxy bonding are used to bond PDMS interconnects to various substrate materials such as glass, silicon, and other polymers (Li 2003).

A second method for creating a packaging channel is to use a coring tool. In this case, a channel is cored from the surface to the buried microchannel and a syringe needle is inserted into the hole and a press fit is used to keep the channel sealed. The use of a needle that has a larger diameter than the coring tool leads to a good fit and a good seal within the microfluidic structure (Christensen *et al.* 2005). Using this concept of an interference fit, seals designed for various pressures were fabricated and tested for leakage and internal wear due to repeated insertion of a tube. The sealing pressure is a function of the contact pressure in each design due to the low stiffness of PDMS. On the basis of this result, guidelines can be set for selecting the appropriate sealing method and for modeling more complex captive connectors (Friedrich *et al.* 2005).

1.12.4.2 Powder Blasting

Abrasive blasting or sand blasting is a widely used material removal process. The microfabrication version, typically called powder blasting, has been demonstrated as a viable, low-cost option for the fabrication of MEMS and microfluidic devices, especially for brittle materials (**Figure 15**). The resolution possible for powder blasting is reduced by the size of the particles, which typically range from 10 to 50 μm and the complexity of the mask patterning undertaken. However, a high etch rate and no clean room requirement make powder blasting an attractive alternative to standard wet and dry etching techniques. These advantages make powder blasting an ideal candidate for rapid prototyping applications.

Powder blasting utilizes a pressurized air beam containing ceramic particles such as Al_2O_3 to create 3D microstructures. By adjusting and controlling parameters such as the beam pressure, beam diameter, beam–substrate separation distance, beam angle of incidence, mask pattern, and mask undercutting, novel 3D features are possible (Belloy *et al.* 2000, 2001, Slikkerveer *et al.* 2000). Masking for the

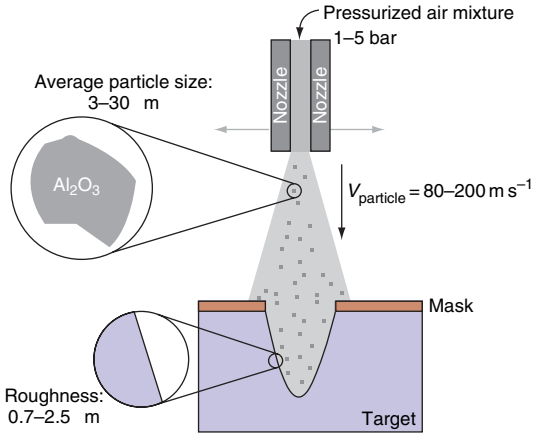


Figure 15 A schematic of the powder blasting process using Al_2O_3 as the blasting particle. (Source: Wensink H, Jansen H V, Berenschot J W, Elwenspoek M C 2000 Mask materials for powder blasting. *J. Micromech. Microeng.* **10**, 175–80; reproduced with permission from the Institute of Physics.)

powder blasting process consists of an etch-resistant coating such as a polymer that is patterned on top of the substrate. Pawlowski *et al.* demonstrated the use of PDMS/SU-8 polymers for an etch mask. PDMS was found to have an etch rate 15 times smaller than glass, allowing for high mask integrity in the course of experiments. With the patterning and etch-resistant capabilities of the PDMS/SU-8 combination, 30-μm structures were reported (Pawlowski *et al.* 2005).

1.12.5 Low-Cost Methods for Volume Manufacturing of MEMS

In many cases, traditional MEMS technologies can be used to produce inexpensive microdevices since the fixed costs associated with the expensive tools can be spread over a large number of devices. For example, precision MEMS accelerometers for use with car airbags can sell for less than US$2 each. For volumes that do not reach millions of parts per year, and for devices that are not easily fabricated using traditional techniques, molding and embossing techniques are the most cost-effective for mass production. Several of these techniques will be reviewed for their application to microdevices.

1.12.5.1 Injection Molding

The utilization of injection molding has made possible the mass production of polymer components for various industries. The extension of injection molding to MEMS and microfluidic components is helping to develop academic and commercial devices without the need of costly MEMS foundry facilities. Injection molding offers solutions such as low cost (for large volumes), high tolerances, and fully automated production. The limitations for injection molding are often high tooling costs (for small volumes), shrinkage, and partial mold filling. These limitations are often exacerbated for microscale components as even tighter tolerances are required for mold inserts, shrinkage, and mold filling. For microscale applications, mold inserts are often fabricated by high-precision laser ablation, lithography, EDM, or other costly micromachining techniques that may cost more than the injection molding equipment. However, recent discoveries at the academic and commercial level have begun to address some of the obstacles that have deterred the use of injection molding for microsystems.

1.12.5.1.1 Process overview

The injection molding process generally encompasses four basic steps: melting, injection, cooling, and ejection of the finished part. The entire cycle time will vary from seconds to minutes, based on the polymer, part size, mold complexity, and equipment used. While there are a considerable number of different injection molding machines, they all consist of three major components: the injection unit, the mold cavity, and the clamping unit.

The typical injection unit consists of three basic components: a hopper, a heated barrel, and reciprocating screw, as shown in **Figure 16**. Polymer in the form of pellets is forced through the hopper into the heated barrel, where the temperature of the polymer is raised above its melting point. A shear force is then applied by the reciprocating screw, which contributes to polymer melting through viscous heating. Polymer flow is also improved by shear thinning that occurs in the melted polymer. At this stage, the reciprocating screw is driven forward by a hydraulic drive, injecting the molten polymer into the mold cavity, which forms the shape of the desired component. A mechanical or hydraulic clamping unit opposes the force of the injection, holding the mold plates together. Backward polymer flow is prevented from returning to the barrel cavity by a nonreturn valve that sits at the tip of the reciprocating screw. The mold is allowed to cool, allowing the polymer to solidify. The mold is then opened and the part is ejected from the mold. After the mold cycle is

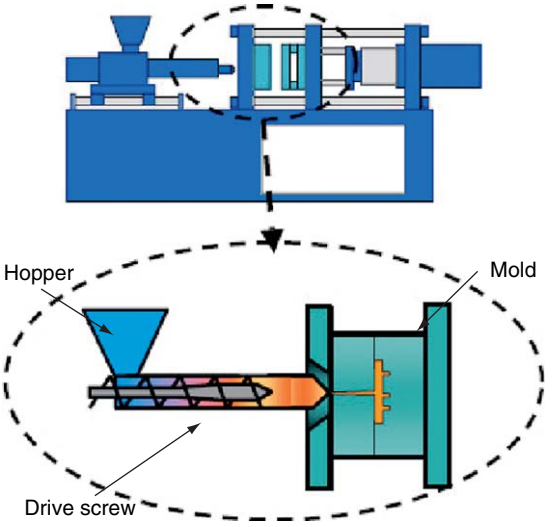


Figure 16 A schematic of a typical injection molding machine showing the hopper, drive screw, and mold. (Source: Becker H, Heim U 2000 Hot embossing as a method for the fabrication of polymer high aspect ratio structures. *Sens. Actuators* **83(1)**, 130–5. © 2000, reproduced with permission from Elsevier.)

completed, the screw retracts and the process is repeated (Bryce 1996).

1.12.5.1.2 Microscale injection molding

Standard injection molding processes require slight modifications when microscale components are fabricated. Microscale injection molding processes focus on the release of delicate microstructures from the mold and proper mold filling in microcavities. Mold release agents are generally used to assist in component release from the mold cavity; however, for certain applications the mold release may adversely affect device biocompatibility. To circumvent this step, the mold cavity walls are given a slight incline, provided the incline is acceptable in the final part. If vertical sidewalls are required, special care must be taken to reduce the effects of friction that may hinder the smooth release of the components from the cavity (Heckele and Schomburg 2004). It has been reported that an average surface roughness value of less than 300 nm will work for most low aspect ratio applications. To ensure proper release for high aspect ratios features (>5) a surface roughness of less than 100 nm is necessary (Pfleging *et al.* 2003).

Complete mold filling is critical for the success of microscale injection molding. A typical strategy for proper filling of the mold is to evacuate and heat the mold cavity. Compression effects (also known as diesel effects) in mold filling, where compressed air is trapped in blind corners and polymer surfaces are burned, provide a unique challenge when optimizing an injection molding process. Evacuation of the mold cavity is an effective strategy designed to remove regions of trapped air that may leave voids in the final polymer component. Another strategy imposed to improve mold filling is heating of the mold cavity. Because of the large surface area-to-volume ratio typical in microscale components such as microchannels, a rapid equilibration temperature occurs between the molded polymer and the temperature of the mold. The equilibrium change in temperature for microscale injection molding has been found to be up to two orders of magnitude faster than conventional molding (Whiteside *et al.* 2005).

The optimization of injection molding parameters and processes is often based on operator experience and trial and error. However, recent modeling tools have been developed by companies and researchers to provide computer simulations of the filling process (Yao and Kim 2002). Moldflow Corporation has developed a commercial tool specifically for microscale flow analysis, referred to as Micro Moldflow Analysis™. The implementation of these simulation tools will help speed up the injection molding parameter selection stage, while simultaneously reducing injection molding cost and operator time.

1.12.5.1.3 Microscale applications of injection molding

As academic research does not generally require high-volume production of polymer components, most injection molding research remains unreported in commercial production settings. However, some recent novel attempts have been reported for microfluidic injection molding applications. Edwards *et al.* used a modified SU-8 mold directly as an insert. Aluminum and titanium were sputtered on the SU-8 to reinforce structural integrity of the structures and facilitate better release of the polymer (Edwards *et al.* 2000). Madou *et al.* combined deep reactive ion etching (DRIE), PDMS, and epoxy to create a mold. DRIE etched the channels, followed by PDMS cast to form the negative pattern of the structures. The epoxy resin was then used to cast the master replica (Madou *et al.* 2001). Kim *et al.* (2002) reported creating micromachined mold inserts by electroplating nickel on PDMS. Svedberg *et al.* used injection molding to produce microelectrospray tips for electrospray ionization for coupling with a time-of-flight mass

spectrometer. Svedberg *et al.* fabricated a mold insert by patterning photoresist on 100-mm silicon wafers and wet etching microchannels into the silicon. Nickel was electroplated to form a mold, which was then placed into a compact disk (CD) injection molding machine using polycarbonate (PC) and PMMA (Svedberg *et al.* 2003). Melin *et al.* used a CD Zeonor injection molding technique that combined e-beam evaporation of silica with a PDMS–Zeonor membrane to create a cell culture device. Zeonor provided quick cycling times and the ability to create multiple devices on an individual CD. The device dramatically increased the volume of devices produced in a given week compared to the same process undertaken with soft lithography (Melin *et al.* 2005). Dang *et al.* (2005) reported a method for mass producing 10-channel CE devices for high-throughput genetic analysis. In an attempt to reduce the costs required for serially producing CE chips in glass, Dang *et al.* explored polymer molding for their CE chip design. The design consisted of 10 square channels (50 μm), with 100 μm separating the channels. The mold was generated using deep X-ray lithography to create a master PMMA chip. The channel walls were inclined (85°) to allow for better release of the molded part. For durability, nickel was electroplated onto the PMMA master to ensure sustainable, defect-free molded parts (Becker and Gartner 2000, Yang *et al.* 2001, Dang *et al.* 2005, McCormick *et al.* 1997, Ouellette 2003). The continued development of novel rapid prototyping techniques for mold inserts will greatly enhance the influence injection molding has on microsystem fabrication.

1.12.5.2 Casting

The use of casting for rapid prototyping overlaps with other mold-based microfabrication techniques. Casting encompasses the various types of mold patterning processes such as standard lithography, etching, or laser ablation to create a master pattern to form a replica part in polymers, ceramics, and metals. For most microsystem applications low-cost biocompatible polymers like PDMS are used to produce the replica parts (see Section 1.12.4.1). The mold patterning technique selected will ultimately determine the cost, time, durability, and resolution of the formed parts.

1.12.5.2.1 Applications

The use of casting has been an important rapid prototyping tool for microfluidics researchers. The Whitesides and Quake Research groups highlighted the advantages of casting PDMS for microfluidic systems. Duffey *et al.* demonstrated the implementation of a CE microfluidic configuration in an oxidized PDMS chip (Duffy *et al.* 1998). Anderson *et al.* showed the possibility of creating complex 3D microfluidic networks using casting techniques. Interweaving channels and serpentine channels were cast using multilayer lithography steps (Anderson *et al.* 2000). Unger *et al.* illustrated the capabilities of multilayer PDMS devices that could be used for the active movement and control of fluids through pneumatic actuation of a thin membrane. Microvalves and peristaltic micropumps have been characterized and reported (Studer *et al.* 2004, Unger *et al.* 2000). Since these initial demonstrations of casting for microfluidics, complex microfluidic devices for biological assays and protein deposition have also been reported. McDonald *et al.* (2001), by utilizing PDMS casting, created a disposable microfluidic network for an enzyme-linked immunosorbent assay (ELISA) detection assay. Bernard *et al.* utilized reactive ion etching to etch channels in a silicon wafer for casting microfluidic biomolecule deposition tools in PDMS. The PDMS deposition tool confined solutions of interest to channels that would pattern a substrate with biomolecules. A second deposition tool will replace the first perpendicular to the original flow lanes, creating intersecting regions where binding events can take place (Bernard *et al.* 2001). Xurography (see Section 1.12.3.5) has also been used to form a casting master for direct molding with PDMS (Bartholomeusz *et al.* 2005). The development of novel, low-cost mold patterning techniques will continue to make casting an integral part of rapid prototyping in research development applications.

1.12.5.3 Hot Embossing

Hot embossing, like injection molding, uses polymers as the primary structural material. Using polymers in conjunction with a simple fabrication process allows hot embossing to be one of the ideal methods for fast, inexpensive replication of microsystems. Additionally, this process is good for creating high aspect ratios and small structures. The feature size and quality depend mainly on the master mold and its fabrication process. Depending on the application, master mold fabrication methods can range from classic photolithography to LIGA (LIGA is a German acronym for X-ray lithography (X-ray lithographie), electroplating (Galvanoformung)). Some limitations of hot embossing are a restriction to planar

features, difficulty in creating multidepth features, and a requirement that all polymers used have a glass transition temperature (T_g). Overall, with the wide range of polymers and master mold fabrication methods, hot embossing has the potential to fit the requirements of a large variety of applications.

1.12.5.3.1 Fabrication process

The first step in hot embossing is to choose a polymer that is compatible with both hot embossing and the application desired. Some common polymers used in hot embossing are PC, PMMA, polyetheretherketone (PEEK), and polystyrene (Becker and Heim 2000, Truckenmuller *et al.* 2002). PC, PMMA, and polystyrene are all transparent, which allows the use of these polymers in optical applications (Fleger *et al.* 2004). Other polymers, like PEEK, may be tougher and useful for applications with higher stresses.

The second step in developing a hot embossing process is choosing which fabrication method to use to create the master mold, which is a relief structure of the design. Photolithography is the classic method for mold fabrication with hot embossing. Photoresist on a silicon substrate can be electroplated with a metal, like nickel, to create metal master molds (Becker and Heim 2000, Fleger *et al.* 2004). Unfortunately, this method has limitations with regard to aspect ratio. An alternative method is DRIE, which can allow for high aspect ratios and vertical walls (Becker and Heim 2000). Here, a silicon substrate can be used as the embossing tool because the silicon can be directly etched instead of a resist layer. This method, though, can yield high surface roughness and is relatively expensive (Narasimham and Papautsky 2004). Another expensive method that also yields vertical walls and high aspect ratios but has very low surface roughness is LIGA (Becker and Heim 2000). Mechanical machining like computer numerical controlled machining has been used to machine steel or nickel molds. These types of masters have very long lifetimes, but are only good for relatively large featured designs (Narasimham and Papautsky 2004), though the ability of these tools is improving (Hupert *et al.* 2006).

The material used to develop the master mold is critical to determining the success of the hot embossing operation. The material generally needs to be stable at the glass transition temperature of the polymer and should be strong to withstand multiple uses; also, it should be thermally conductive and smooth on all mating surfaces. The surface roughness usually depends on what type of fabrication method is used to make the mold. Nickel masters are usually silicon substrates with nickel electroplated onto them. Electroplating, unfortunately, is associated with problems such as nonuniformity, bubbles, poor adhesion to the substrate, and difficulty in plating high aspect ratio structures (Narasimham and Papautsky 2004). Silicon is also used, but as mentioned before, the techniques used to etch silicon can lead to high surface roughness. A unique material used by Narasimhan and Papautsky (2004) is PDMS. Intuitively, PDMS would not be ideal material for hot embossing because of its soft malleable nature, but through parameter optimization quality devices were fabricated and the PDMS mold withstood multiple uses. Thus, master mold material is not restricted to typical stiff and hard materials.

To begin replication, the master mold is placed within a hot embossing press as seen in **Figure 17** (Becker and Heim 2000), which consists of two hot plates and a force mechanism. The master is placed on one hot plate and the polymer substrate is placed on the other. The two plates are heated independently while pulling a vacuum on the entire chamber. Once the hot plates have reached the desired temperature, a certain temperature above T_g, the two plates are pressed together for a specific force and time. After the allotted time, the plates are cooled while keeping the force constant. After cooling, the de-embossing step pulls the two plates apart (Heckele 2003, Simdikova *et al.* 2002). Another type of process is the continuous process in which the master mold is replicated on a heated belt. A polymer sheet is pushed through the heated belt and a base, thereby replicating the mold multiple times on a single sheet (Madou 2002).

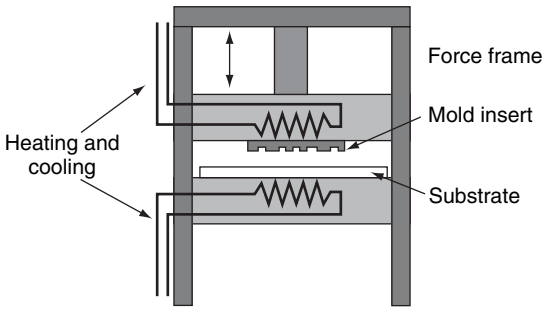


Figure 17 Basic setup for a hot embossing press (Source: Becker H, Heim U 2000 Hot embossing as a method for the fabrication of polymer high aspect ratio structures. *Sens. Actuators* 83, 130–5. © 2000, reproduced with permission from Elsevier.)

1.12.5.3.2 Process problems and limitations

Although the hot embossing process seems simple, there are many parameters that must be considered. Depending on the dimensions of the design, time, temperature, and pressure must be optimized to ensure full filling of the master mold channels. With small channel dimensions, the flow of the polymer into these spaces can be difficult. This concept is the same for deep or high aspect ratio structures. Optimization of the parameters was studied by Shen *et al.* (2002). When the temperature is raised, the height of the polymer reaches into channels increases. Although an increase in temperature leads to an increase in polymer flow, too large an increase will cause the polymer to no longer hold the shape of the mold. Also, the polymer and master materials have different thermal expansion coefficients, and if the operation temperature range is too large high stresses will be placed on the polymer material. An increase in pressure has the same effect as temperature, but increasing pressure can cause high stresses that may break the polymer substrate. By increasing the molding time, the polymer height can also be extended. The height, though, was found to increase to a maximum point and lengthening the time after this point has no effect.

A second challenge is the de-embossing step. With high temperatures and pressures a certain amount of adhesion occurs between the master and the polymer substrate; therefore, during de-embossing, when the two plates are pulled apart, damage can occur to the fabricated structures. To alleviate distortion, releasing agents or antiadhesive films may be used by coating the mold with a thin layer before pressing it into the polymer (Becker and Heim 2000, Madou 2002). Unfortunately, if hot embossing is used for a life science application, releasing agents are not preferred because they can contaminate the surfaces and reduce biocompatibility. A second method to reduce de-embossing damage is reducing surface roughness because high surface roughness can cause large amounts of friction. For structures having high aspect ratios above 0.5 to achieve good de-embossing, surface roughness must be less than 80-nm rms. Fortunately, some fabrication processes like LIGA can achieve surface roughness of 10-nm rms and ASE-processed silicon can achieve 8-nm rms (Becker and Heim 2000). A final method to alleviate high stresses is to create wall angles that are less than 90° (Becker and Heim 2000). By doing so, vertical forces are reduced. Some other problems that may occur in hot embossing are high stresses due to pressure on the substrate and bonding problems during packaging.

General limitations of hot embossing are largely based on the limitations of the mold fabrication process. For example, if typical wet silicon etching is used as the master mold fabrication, the shape of the channels within the polymer substrate will not be rectangular. The overall maximum aspect ratio is around 50 and the minimum feature size is in the nanometer range, which leads us to nanoimprint lithography. Hot embossing tooling can be relatively expensive, which is why it is preferable for low and midrange production. A hot embossing press can cost around US$150 000. Another cost consideration is choosing a polymer because they have a wide range of prices depending on the type of material. PC, for example, can cost about US$24 for a 12″ × 12½″ sheet. In contrast, a same-size sheet of PEEK will cost about US$600 (www.usplastics.com).

1.12.5.3.3 Advantages and disadvantages of hot embossing

Hot embossing has many advantages and disadvantages compared to other microfabrication methods. Hot embossing is good for small structures and high aspect ratios (Becker and Heim 2000). Another advantage is the simple fabrication process used in hot embossing, which is because of the small number of steps required for fabrication (Heckele 2003, Simdikova *et al.* 2002, Truckenmuller *et al.* 2002). After the master mold is created, fabrication of the fluidic channels can be done in minutes and repeated several times. The simplicity of fabrication inherently leads to another advantage, a strong potential for mass production. Mass replication also provides a possibility to design disposable parts (Truckenmuller *et al.* 2002). Other fabrication methods that are timely make disposability a costly function. The inexpensive cost of the polymers also makes it possible to manufacture disposable parts; low long-term costs is another advantage. Finally, many polymers are also biocompatible and are currently used in many medical applications.

Hot embossing also has many disadvantages. Because the channels are created by pressing a mold into a planar sheet, the features that can be created must also be planar. This is a huge disadvantage compared to traditional methods like UV lithography, where multiple layers can be manufactured easily. Hot embossing's cousin, injection molding, which also can be inexpensive and quick, can also create nonplanar features. Another disadvantage is the high residual stresses (Becker and Heim 2000). The use of temperature in conjunction with pressure on polymers, which are not the most durable materials, can

cause cracks and broken edges unlike in silicon wet etching. Corresponding to high residual stresses, multiple-depth features are difficult to make because of the required higher temperatures and pressures (Madou 2002). Hot embossing also has high initial costs. For example, if the master mold is made using LIGA, the cost of paying for a LIGA mold will be extremely expensive, along with having to buy the LIGA mold materials and a hot embossing press.

1.12.5.3.4 Hot embossing applications

Hot embossing is a relatively new method for microfabrication. Currently, the most common explorations are in the fields of optics, motors, bioMEMS, and IC packaging industry. In the IC industry this method is being explored as a packaging solution called molded interconnect devices (MIDs) (Kunststoffe 1998). The concept of these MIDs is to integrate an electronic circuit in a polymer housing. Through this packaging system the electronic parts can be arranged three-dimensionally and they can consequently reduce the materials needed and the total volume.

A common and simple use of hot embossing in optics is fabrication of waveguides. Fleger *et al.* (2004) accomplished this task by using a nickel master mold made using UV lithography with nickel electroplating and embossing into a PMMA substrate. The polymer waveguides were characterized by performing propagation loss and transmission measurements. Flegler also demonstrated the simultaneous manufacturing of microfluidics and optics on a chip using hot embossing.

Another unique example of using hot embossing is electrostatic comb drives. Zhao and Cui (2003) successfully designed a comb drive out of PMMA. In this case, the master mold was fabricated out of silicon using DRIE. Multiple embossing layers as well as metallization steps were employed, showing the flexibility and compatibility of the hot embossing process. The pull-in voltage, stroke length, and natural frequency were all reasonable and the process could easily compete with traditional silicon methods.

A final application example of hot embossing is a microfluidic analysis chip, which is the traditional use of hot embossing. A good example is from Truckenmuller *et al.* (2002). To create their mold, micromachined trenches were milled in brass plates. Next, the mold was placed into a hot embossing press with a two-layered substrate of polystyrene consisting of a thin top layer of high-impact polystyrene and a thick bottom layer of polystyrene. The substrate was coated with a heat-sealing layer and then typically pressed together with the mold in the embossing press. By placing an electrode in between the two layers of the polystyrene, Truckenmiller was able to create a fluidic device for CE.

Various other applications of hot embossing have been explored, with the majority of them being related to microfluidics. Overall, hot embossing can be a low-cost method for low-range to midrange fabrication of molded polymer microdevices.

1.12.5.4 Roller Embossing

The process of roller embossing is a rapid fabrication technique generally using a roller covered with a thin mold that patterns a polymer or some other moldable material. The embossed material is typically formed through a heat-forming or photocuring configuration. Roller embossing is a useful tool for creating large arrays of continuous microstructures. Chang *et al.* reported the use of roller embossing of a UV-curable photopolymer in a custom UV exposure facility to form polymer microlens arrays. Before fabrication, a flat mold is generated from electroplating nickel or some other durable metal onto a polymer microlens master substrate. The resulting flat mold consisting of microlens indentations is wrapped around a cylindrical device to form the roller. The UV-curable polymer is embossed onto a glass substrate as the roller presses the microlens arrays. The roller embossing is implemented within the UV exposure zone, allowing the polymerization and formation of the microstructures (Chang *et al.* 2006). Gallagher demonstrated roller embossing for fuel cell applications using curable polymer materials. The technique is used to form fluid flow cells on separator plates that sandwich the membrane electrolyte assembly (MEA). The flow cells are used to circulate coolant or supply reactants to the MEA (Gallagher 2004). A similar method has been used to fabricate antireflective microstructures for LCD displays (Parrika and Saarinen 2005).

1.12.6 Conclusions

Though MEMS and microfabrication have been dominated by processes that require expensive tooling, a large number of techniques have been developed that can help eliminate the high costs associated with traditional MEMS processes. One of the easiest ways to reduce costs is to use tools available in the mass consumer market, such as laser or

ink-jet printers for making masks. Techniques such as xurography can use commercially available double-sided tapes that can not only produce high-resolution devices, but also eliminate most of the processing steps such as lithography, etching, and bonding. The recent use of plastics has also been a major development that has the capability to drive down costs. Directly patterning plastics using photosensitive polymers may be the simplest, but laser cutting and molding of polymers is becoming increasingly commonplace. The adoption of rugged soft polymers for microfluidics research (soft lithography) has rapidly driven the field forward, since nearly any researcher can easily set up the facilities needed for prototyping. The precision with which plastics can be molded has led to interesting applications of techniques such as soft lithography, which allows nanoscale features to be generated repeatably and reliably. Thus, for some applications, the inexpensive techniques are replacing the more traditional MEMS techniques as the primary processing method as the function of plastics is more amenable to biomedical and microfluidic applications.

In many cases, the question is no longer, "Can an inexpensive microfabrication technique do the job?" but "Which inexpensive process is best for this application?" The answer to this question is not always simple, but a few guidelines can make the decision process simpler. For example, some of the techniques discussed in this chapter are limited in their ultimate resolution, such as screen printing, xurography, and laser printing, but for larger microscale devices they are the simplest and least expensive. Combining some inexpensive methods with some more expensive tools can lead to high precision at a relatively low cost. For example, using e-beam lithography to make a soft lithography mold can lead to inexpensive nanoscale structures, even if the e-beam step is rather expensive. Some of the techniques are better for prototyping than for mass production, such as soft lithography and xurography, but new efforts to commercialize these techniques may nullify this current limitation. The pure molding processes are typically more expensive unless high volumes of devices are fabricated, so they are typically precluded from prototyping. The one limitation of most of these techniques is that they do not necessarily work well with silicon, the traditional MEMS material. Thus, converting to a less expensive process often requires a change of material, but if circuits are not the focus, this conversion can often be painless, not to mention less expensive.

References

Abgrall P, Chuda K, Coqueret X, Gue A M 2005 Characterization of a new generic 3D polymer technology for microTAS. In: Jensen K F, Han J, Harrison D J, Voldman J (eds.) *Proc. 9th Int. Conf. Miniaturized Systems for Chemistry and Life Sciences.* Transducers Research Foundation, San Diego, CA, USA, pp. 687–9

Abgrall P, Lattes C, Conedera V, Dollat X, Colin S, Gue A M 2006 A novel fabrication method of flexible and monolithic 3D microfluidic structures using lamination of SU-8 films. *J. Micromech. Microeng.* **16**, 113–21

Aden J S, Bohorquez J H, Collins D M, Crook M D, Garcia A, Hess U E 1994 Third generation HP thermal inkjet printhead. *Hewlett-Packard J.* **45(1)**, 41–5

Anderson J R, Chiu D T, Jackman R J, Cherniavskaya O, McDonald J C, Wu H, Whitesides S H, Whitesides G M 2000 Fabrication of topologically complex three-dimensional microfluidic systems in PDMS by rapid prototyping. *Anal. Chem.* **72**, 3158–64

Arcona C, Dow T A 1999 The role of knife sharpness in slitting plastic films. *J. Mater. Sci.* **18**, 93–5

Asthana A, Asthana Y, Sung I K, Kim D P 2006 Novel transparent poly(silazane) derived solvent-resistant, biocompatible microchannels and substrates: Application in microsystem technology. *Lab Chip* **6**, 1200–4

Badal M Y, Wong M C, Hiem N, Salimi-Moosavi H, Harrison D J 2002 Protein separation and surfactant control of electroosmotic flow in poly(dimethylsiloxane)-coated capillaries and microchips. *J.Chromatogr. A* **947**, 277–86

Barbier V T M, Li H, Arefi-Khonsari F, Ajdari A, Tabeling P 2006 Stable modification of PDMS surface properties by plasma polymerization: Application to the formation of double emulsions in microfluidic systems. *Langmuir* **22**, 5230–2

Bartholomeusz D A, Boutte R W, Andrade J D 2005 Xurography: Rapid prototyping of microstructures using a cutting plotter. *J. Microelectromech. Syst.* **14(6)**, 1364–74

Becker H, Gartner C 2000 Polymer microfabrication methods for microfluidic analytical applications. *Electrophoresis* **21**, 12–16

Becker H, Heim U 2000 Hot embossing as a method for the fabrication of polymer high aspect ratio structures. *Sens. Actuators* **83**, 130–5

Beebe D J, Moore J S, Bauer J M, Yu Q, Liu R H, Devadoss C, Jo B H 2000 Functional hydrogel structures for autonomous flow control inside microfluidic channels. *Nature* **404**, 588–90

Belloy E, Sayah A, Gijs M A M 2000 Powder blasting for three-dimensional microstructuring of glass. *Sens. Actuators A Phys.* **86**, 231–7

Belloy E, Sayah A, Gijs M A M 2001 Oblique powder blasting for three-dimensional micromachining of brittle materials. *Sens. Actuators A Phys.* **92**, 358–63

Bernard A, Delamarche E, Schmid H, Michel B, Bosshard H R, Biebuyck H 1998 Printing patterns of proteins. *Langmuir* **14**, 2225–9

Bernard A, Michel B, Delamarche E 2001 Micromosaic immunoassays. *Anal. Chem.* **73**, 8–12

Bhattacharya S, Datta A, Berg J M, Gangopadhyay S 2005 Studies on surface wettability of poly(dimethyl)siloxane (PDMS) and glass under oxygen-plasma treatment and correlation with bond strength. *J. Microelectromech. Syst.* **14**, 590–7

Brown L, Koerner T, Horton J H, Oleschuk R D 2006 Fabrication and characterization of poly(methymethacrylate) microfluidic devices bonded using surface modifications and solvents. *Lab Chip* **6**, 66–73

Bryce D M 1996 *Plastic Injection Molding: Manufacturing Process Fundamentals*. Society of Manufacturing Engineers, Dearborn, MI

Chang C Y, Yang S Y, Sheh J L 2006 A roller embossing process for rapid fabrication of microlens arrays on glass substrates. *Microsyst. Technol.* **12**, 754–9

Chang-Yen D A, Gale B K 2003 An integrated optical oxygen sensor fabricated using rapid-prototyping techniques. *Lab Chip* **3**, 297–301

Chang-Yen D A, Gale B K 2005 PDMS microfluidic spotter for fabrication of protein chips and micro-arrays. *Proc. SPIE* **5718**, 110–20

Chen C, Hirdes D, Folch A 2003 Gray-scale photolithography using microfluidic photomasks. *Proc. Natl. Acad. Sci. USA* **100(4)**, 1499–504

Childs W R, Nuzzo R G 2002 Decal transfer microlithography: A new soft lithographic patterning method. *J. Am. Chem. Soc.* **124**, 13583–96

Childs W R, Motala M J, Lee K J, Nuzzo R G 2005 Masterless soft lithography: Patterning UV/ozone-induced adhesion on poly(dimethylsiloxane) surface. *Langmuir* **21**, 10096–105

Chow Y M, Lau W M, Schetty R E, Karim Z S 2000 Feasibility and reliability study on the electroless nickel bumping and stencil solder printing for low-cost flip chip electronic packaging. *Int. Symp. Electronic Materials & Packaging,* Hong Kong, China, pp. 79–85

Christensen A M, Chang-Yen D A, Gale B K 2005 Characterization of interconnects used in PDMS microfluidic systems. *J. Micromech. Microeng.* **15**, 928–34

Clarson S J, Semlyen J A 1993 *Siloxane Polymers*. Prentice-Hall, Englewood Cliffs, NJ

Cooley P, Wallace D, Antohe B 2001 Applications of ink-jet printing technology to bioMEMS and microfluidic systems. *Proc. SPIE* **4560**, 177–88

Conradie E H, Moore D F 2002 SU-8 thick photoresist processing as a functional material for MEMS applications. *J. Micromech. Microeng.* **12**, 368–74

Dang F, Tabata O, Kurokawa M, Ewis A A, Zhang L, Yamaoka Y, Shinohara S, Shinohara Y, Ishikawa M, Baba Y 2005 High-performance genetic analysis on microfabricated capillary array electrophoresis plastic chips fabricated by injection molding. *Anal. Chem.* **77**, 2140–6

Delamarche E, Donzel C, Kamounah F S, Wolf H, Geissler M, Stutz R, Schmidt-Winkel P, Michel B, Mathieu H J, Schaumburg K 2003 Microcontact printing using poly (dimethylsiloxane) stamps hydrophilized by poly(ethylene oxide) silanes. *Langmuir* **19**, 8749–58

Deng T, Tien J, Xu B, Whitesides G M 1999 Using patterns in microfiche as photomasks in 10-μm-scale microfabrication. *Langmuir* **15**, 6575–81

Deng T, Wu H, Brittain S T, Whitesides G M 2000 Prototyping of masks, masters, and stamps/molds for soft lithography using an office printer and photographic reduction. *Anal. Chem.* **72**, 3176–80

Despont M, Lorenz H, Fahrni N, Brugger J, Renaud P, Vettiger P 1997 High aspect ratio ultrathick, negative-tone near-UV photoresist for MEMS applications. *MEMS'97, IEEE*, Nagoya, Japan, pp. 518–22

Doshi P, Mejia J, Tate K, Rohatgi A 1996 Integration of screen-printing and rapid thermal processing technologies for silicon cell fabrication. *IEEE Electron Device Lett.* **17(8)**, 404–6

Dotson N A, Kim P T, Mason A 2004 Low cost MEMS processing techniques. *Proc. 2004 ASEE/NCS Spring Conference*. Kalamazoo, MI, USA

Dow Corning Co. 2005 Product data sheet for silicone encapsulants. www.dowcorning.com

Duffy D C, McDonald J C, Schueller O J A, Whitesides G M 1998 Rapid prototyping of microfluidic systems in polydimethylsiloxane (PDMS). *Anal. Chem.* **70**, 4974–84

Eddings M A, Gale B K 2006 A PDMS-based gas permeation pump for on-chip fluid handling in microfluidic devices. *J. Micromech. Microeng.,* **16**, 2396–2402

Edwards T L, Mohanty S K, Edwards R K, Thomas C, Frazier A B 2000 Rapid tooling using SU-8 for injection molding microfluidic components. *Proc. SPIE Int. Soc. Opt. Eng.* **4177**, 82–9

El-Ali J, Sorger P K, Jenson K F 2006 Cells on chips. *Nature* **442**, 403–11

Fleger M., Siepe D, Neyer A 2004 Microfabricated polymer analysis chip for optical detection. *Nanobiotechnology* **151**, 159–61

Friedrich C R, Avula R R K, Guglae S 2005 A fluid microconnector seal for packaging applications. *J. Micromech. Microeng.* **15**, 1115–24

Fuller S B, Wilhelm E J, Jacobson J M 2002 Ink-jet printed nanoparticle microelectromechanical systems. *J. Microelectromech. Syst.* **11(1)**, 54–60

Franssila S 2004 *Introduction to Microfabrication*. Wiley & Sons, UK

Gale B K, Caldwell K D, Frazier A B 2002 Geometric scaling effects in electrical field-flow fractionation. 2. Experimental verification. *Anal. Chem.* **74**, 1024–30

Gallagher E R 2004 *US Pat.* 6 818 165

Gates B D 2004 Nanofabrication with molds and stamps. *Mater. Today* **8**, 44–9

Gates B D, Whiteside G W 2003 Replication of vertical features smaller than 2 nm by soft lithography. *J. Am. Chem. Soc.* **125**, 14986–7

Gates B D, Xu Q, Love J C, Wolfe D B, Whitesides G W 2004 Unconventional nanofabrication. *Annu. Rev. Mater. Res.* **34**, 339–72

Geissler M, Schmid H, Bietsch A, Michel B, Delamarche E 2002 Defect-tolerant and directional wet-etch systems for using monolayers as resists. *Langmuir* **18**, 2374–7

Gomez-Morilla I, Abraham M H, de Kerckhove D G, Grime G W 2005 Micropatterning of foturan photosensitive glass following exposure to MeV proton beams. *J. Micromech. Microeng.* **15**, 706–9

Gothait H 2003 *US Pat.* 6 644 763

Gracias A, Xu B, Castracane J 2005 Fabrication of three dimensional microchannels in SU8. In: Jensen K F, Han J, Harrison D J, and Voldman J (eds.) *Proc. 9th Int. Conf. Miniaturized Systems for Chemistry and Life Sciences.* Transducers Research Foundation, San Diego, CA, USA, pp. 663–5

Hayes D J, Grove M E, Cox W R 1999a Development and application by ink-jet printing of advanced packaging materials. *Proc. Int. Symp. Advanced Materials Process, Properties and Interfaces*, Braselton, GA, USA, pp. 88–92

Hayes D J, Wallace D B, Cox W R 1999b Microjet printing of solder and polymers for multi-chip modules and chip-scale packages. *Proc. IMAPS Int. Conf. High Density Packaging and MCMs,* Denver, CO, USA, pp. 242–7

Heckele M 2003 Hot embossing – A flexible and successful replication technology for polymer MEMS. *Proc. SPIE* **5354**, 108–17

Heckele M, Schomburg W K 2004 Review on micromolding of thermoplastic polymers. *J. Micromech. Microeng.* **14**, R1–14

Hongo T, Sugioka K, Niino H, Cheng Y, Masuda M, Miyamoto I, Takai H, Midorikawa K 2005 Investigation of photoreaction mechanism of photosensitive glass by femtosecond laser. *J. Appl. Phys.* **97**, 063517: 1–4

Hu S R X, Bachman M, Sims C E, Li G P, Allritton N 2002 Surface modification of poly(dimethylsiloxane) microfluidic devices by ultraviolet polymer grafting. *Anal. Chem.* **74**, 4117–23

Hu S R X, Bachman M, Sims C E, Li G P, Allritton N 2003 Cross-linked coatings for electrophoretic separations in poly(dimethylsiloxane) microchannels. *Electrophoresis* **24(21)**, 3679–88

Hughes D C, Ernster S E 2003 Screen printed feature size capabilities. *Proc. IMAPS Conf. Exhibition on Ceramic Interconnect Technology: The Next Generation*, Denver, CO, USA, pp. 58–62

Hupert M L, Guy J W, Llopis S D, Situma C, Rani S, Nikitopoulos D E, Soper S A 2006 High-precision micromilling for low-cost fabrication of metal mold masters. *Proc. SPIE* **6112**, 61120B: 1–12

Ito T, Uchiyama K, Ohya S, Kitamori T 2001 Application of microchip fabricated of photosensitive glass for thermal lens microscopy. *Jpn. J. Appl. Phys.* **40**, 5469–73

Jabbour G E, Radspinner R, Peyghambarian 2001 Screen printing for the fabrication of organic light-emitting devices. *IEEE J. Sel. Top. Quant. Electron.* **7(5)**, 769–73

Jackman R L, Wilbur J L, Whitesides G M 1995 Fabrication of submicrometer features on curved substrates by microcontact printing. *Science* **269**, 664–6

Jacobs H O, Whitesides G M 2001 Submicrometer patterning of charge in thin-film electrets. *Science* **291**, 1763–6

Jeong J W, Rohatgi A, Yelundur V, Ebong A, Rosenblum M D, Kalejs 2001 Enhanced silicon solar cell performance by rapid thermal firing of screen-printed metals. *IEEE Trans. Electron Devices* **48(12)**, 2836–41

Jeong W J, Kim J Y, Kim S J, Lee S H, Mensing G, Beebe D J 2004 Hydrodynamic microfabrication via 'on the fly' photopolymerization of microscale fibers and tubes. *Lab Chip* **4**, 576–80

Jeong W J, Kim J Y, Choo J, Lee E K, Han C S, Beebe D J, Seong G H, Lee S H 2005 Continuous fabrication of biocatalyst immobilized microparticles using photopolymerization and immiscible liquids in microfluidic systems. *Langmuir* **21**, 3738–41

Khademhosseini A S K Y, Jon S, Eng G, Yeh J, Chen G, Langer R 2004 A soft lithographic approach to fabricate patterned microfluidic channels. *Anal. Chem.* **76**, 3675–81

Kim J, Gale B K 2005 Multi-DNA extraction chip based on an aluminum oxide membrane integrated into a PDMS microfluidic structure. *Proc. IEEE-MMB*, Oahu, HI, USA, pp. 5–7

Kim K, Park S, Lee J B, Manohara H, Desta Y, Murphy M and Ahn C H 2002 Rapid replication of polymeric and metallic high aspect ratio microstructures using PDMS and LIGA technology. *Microsyst. Technol.* **9**, 5–10

Khoury C, Mensing G A, Beebe D J 2002 Ultra rapid prototyping of microfluidic systems using liquid phase photopolymerization. *Lab Chip* **2**, 50–5

LaBianca N, Gelorme J 1995 High aspect ratio resist for thick film applications. *Proc. SPIE Proc. SPIE* **2438**, 846–52

LaBianca N, Gelorme J, Lee K, Sullivan E, Shaw J 1995 High aspect ratio optical resist chemistry for MEMS applications. *4th Int. Symp. Magnetic Materials, Processes, and Devices*, Vol. 95-18, pp. 386–96

LaFratta C N, Li L, Fourkas J T 2006 Soft-lithographic replication of 3D microstructures with closed loops. *Proc. Natl. Acad. Sci. USA* **103**, 8589–94

Lahann J, Balcells M, Lu H, Rodon T, Jensen K F, Langer R 2003 Reactive polymer coatings: A first step toward surface engineering of microfluidic devices. *Anal. Chem.* **75**, 2117–22

Lee K, LaBianca N, Rishton S, Zohlgharnain S 1995 Micromachining applications for a high resolution ultra-thick photoresist. *J. Vac. Sci. Technol. B* **13**, 3012–6

Lee Y C, Parviz B A, Chiou J A, Chen S 2003 Packaging for microelectromechanical and nanoelectromechanical system. *IEEE Trans. Adv. Packag.* **26**, 217–6

Lee H-H, Chou K-S, Huang K-C 2005 Inkjet printing of nanosized silver colloids. *Nanotechnology* **16**, 2436–41

Lee J, Kim M J, Lee H H 2006 Surface modification of poly(dimethylsiloxane) for retarding swelling in organic solvents. *Langmuir* **22**, 2090–5

Li S, Chen S 2003 Polydimethylsiloxane fluidic interconnects for microfluidic systems. *IEEE Trans. Adv. Packag.* **26**, 242–7

Loo Y L, Willett R L, Baldwin K W, Roger J A 2002a Interfacial chemistries for nanaoscale transfer printing. *J. Am. Chem. Soc.* **124**, 7654–5

Loo Y L, Willett R L, Baldwin K W, Roger J A 2002b Additive, nanoscale patterning of metal films with a stap and a surface chemistry mediated transfer process: Application in plastic electronics. *Appl. Phys. Lett.* **81**, 562–4

Lorenz R M, Kuyper C L, Allen P B, Lee L P, Chiu D T 2004 Direct laser writing on electrolessly deposited thin metal films for applications in micro- and nanofluidics. *Langmuir* **20**, 1833–7

Love J C, Wolfe D B, Jacobs H O, Whitesides G M 2001 Microscope projection photolithography for rapid prototyping of masters with micron-scale features for use in soft lithography. *Langmuir* **17**, 6005–12

Madou M J 2002 *Fundamentals of Microfabrication* 2nd edn. CRC Press, Boca Raton, FL, USA, p. 365

Madou M J, Lee L J, Koelling K W, Daunert S, Lai S, Koh C G, Juang Y J, Yu L, Lu Y 2001 Design and fabrication of polymer microfluidic platforms for biomedical applications. *ANTEC-SPE 59th,* **3**, 2534–8

Makamba H, Kim J H, Lim K, Park N, Hahn J H 2003 Surface modification of poly(dimethylsiloxane) microchannels. *Electrophoresis* **24**, 3607–19

Manz A, Graber N, Widmer H M 1990 Miniaturized total chemical analysis systems. A novel concept for chemical sensing. *Actuators B Chem.* **B1**, 244–8

Masuda M, Sugioka K, Cheng Y, Aoki N, Kawachi M, Shihoyama K, Toyoda K, Helvajian H, Midorikawa K 2003 3-D microstructuring inside photosensitive glass by femtosecond laser excitation. *Appl. Phys. A Mater. Sci. Process.* **76**, 857–60

Mata A, Fleischman A J, Roy S 2006 Fabrication of multi-layer SU-8 microstructures. *J. Micromech. Microeng.* **16**, 276–84

McCormick R M, Nelson R J, Alonso-Amigo M G, Benvegnu D J, Hooper H H 1997 Microchannel electrophoretic separations of DNA in injection-molded plastic substrates. *Anal. Chem.* **69**, 2626–30

McDonald J C, Whitesides G W 2002 Poly(dimethylsiloxane) as a material for fabricating microfluidic devices. *Acc. Chem. Res.* **35**, 491–9

McDonald J C, Metallo S J, Whitesides G M 2001 Fabrication of a configurable, single-use microfluidic device. *Anal. Chem.* **73**, 5645–50

Mei J, Lovell M R, Mickle M H 2005 Formulation and processing of novel conductive solution inks in continuous inkjet printing of 3-D electric circuits. *IEEE Trans. Electron. Packag. Manuf.* **28(3)**, 265–73

Melin J, Johansson H, Solderberg H, Nikolajeff F, Landegren U, Nilsson M, Jarvius J 2005 Thermoplastic microfluidic platform for single-molecule detection, cell culture, and actuation. *Anal. Chem.* **77**, 7122–30

Mensing G, Pearce T, Beebe D J 2005 An ultrarapid method of creating 3D channels and microstructures. *JALA* **10**, 24–8

Meyer W 2001 Micro dispensing of adhesives and other polymers. *First Int. IEEE Conf. Polymers and Adhesives in Microelectronics and Photonics*, Potsdam, Germany, pp. 35–9

Moorthy J, Beebe D J 2003 *In situ* fabricated porous filters for microsystems. *Lab Chip* **3**, 62–6

Narasimhan J, Papautsky I 2004 Polymer embossing tools for rapid prototyping of plastic microfluidic devices. *J. Micromech. Microeng.* **14**, 96–103

Odom T W, Love J C, Wolfe D B, Paul K E, Whitesides G M 2002 Improved pattern transfer in soft lithography using composite stamps. *Langmuir* **18**, 5314–20

Olsen CE, Serpa L J 2004 *US Pat.* 4 096 626

Oullett J 2003 A new wave of microfluidic devices. *Ind. Physicist* **August/September,** Volume 9,14–17

Parikh M R, Quilty W F Jr., Gardiner K M 1991 SPC and setup analysis for screen printed thick films. *IEEE Trans. Components Hybrids Manuf. Technol.* **14(3)**, 493–8

Parrika M, Saarinen K 2005 *US Pat.* 6 888 676

Pawlowski A, Sayah A, Gijs M A M 2005 Precision poly-(dimethyl siloxane) masking technology for high-resolution powder blasting. *J. Microelectromech. Syst.* **14**, 619–24

Pfleging W, Hanemann T, Torge M, Bernauer W 2003 Rapid fabrication and replication of metal, ceramic and plastic mould inserts for application in microsystem technologies. *J. Mech. Eng. Sci. Proc. Inst. Mech. Eng. C* **217**, 53–63

Qin D, Xia Y, Whitesides G M 1996 Rapid prototyping of complex structures with feature sizes larger than 20 μm. *Adv. Mater.* **8**, 917–19

Quake S R, Scherer A 2000 From micro- to nanofabrication with soft materials. *Science* **290**, 1536–40

Roger J A, Nuzzo R G 2005 Recent progress in soft lithography. *Mater. Today* **8**, 50–6

Rohatgi A, Yelundur V, Jeong J-W, Kim D S, Gabor A M 2003 Implementation of rapid thermal processing to achieve greater than 15% efficient screen-printed ribbon silicon solar cells. *3rd World Conf. Photovoltaic Energy Convention*, Osaka, Japan, pp. 1352–5

Sakai J, Fujinaka E, Nishimori T, Ito N, Adachi J, Nagano S, Murakami K 2005 High efficiency organic solar cells by screen printing method. *Conf. Rec. 31st IEEE Photovoltaic Specialists Conference*, Lake Buena Vista, FL, USA, pp. 125–8

Sasserath J, Fries D 2002 Rapid prototyping and development of microfluidic and bioMEMS devices. *IVD Technol.* **June**, 1–11

Saunders R, Gough J, Derby B 2005 Ink jet printing of mammalian primary cells for tissue engineering applications. *Mater. Res. Soc. Symp. Proc.* **845**, 57–62

Shaw J M, Gelorme J D, N. LaBianca N C, Conley W E, Holmes S J 1997 Negative photoresists for optical lithography. *IBM J. Res. Dev.* **41**, 81–94

Shen X J, Pan L, Lin L 2002 Microplastic embossing process: Experimental and theoretical characterizations. *Sens. Actuators* **97–8**, 428–33

Shimizugawa Y, Handa K, Qiu J R 2003 X-ray and UV irradiation effects on Ce^{3+} ion doped in UV sensitive glass. *J. Mater. Sci. Lett.* **22**, 15–16

Simdikova I, Kueper A, Sbarski I, Harvey E, Hayes J P 2002 A study of hot embossed microchannels using confocal microscopy. *Proc. SPIE* **4936**, 82–92

Sippola C B, Ahn C H 2005 A ceramic capacitive pressure microsensor with screen-printed diaphragm. *The 4th IEEE Conf. Sensors (IEEE Sensors 2005)*, Irvine, CA, USA, pp. 1271–4

Sirringhaus H, Kawase T, Friend R H, Shimoda T, Inbasekaran M, Wu W, Woo E P 2000 High-resolution inkjet printing of all-polymer transistor circuits. *Science* **290**, 2123–6

Slentz B E, Penner N A, Regnier F E 2002 Capillary electrochromatography of peptides on microfabricated poly(dimethylsiloxane) chips modified by cerium(IV)-catalyzed polymerization. *J. Chromatogr. A* **948**, 225–33

Slikkerveer P J, Bouten P C P, de Haas F C M 2000 High quality mechanical etching of brittle materials by powder blasting. *Sens. Actuator A Phys.* **85**, 296–303

Stampfer S, Ehrenstein G W 1998 Hot-embossing of electronic circuits. *Kunstoffe Plast Europe* **88**, 47–48

Stookey S D, Beall G H, Pierson J E 1978 Full-color photosensitive glass. *J. Appl. Phys.* **49**, 5114–23

Studer V, Hang G, Pandolfi A, Ortiz M, Anderson W F, Quake S 2004 Scaling properties of a low-actuation pressure microfluidic valve. *J. Appl. Phys.* **95**, 393–8

Suleski T J, O'Shea D C 1995 Fidelity of POSTSCRIPT-generated masks for diffractive optics fabrication. *Appl. Opt.* **34(4)**, 627–35

Sundberg S O, Greer J, Wittwer C T, Pryor R J, Elenitoba-Johnson O, Gale B K 2006 Homogeneous DNA melting analysis for mutation scanning using nanoliter volumes. *10th Int. Conf. Miniaturized Systems for Chemistry and Life Sciences (μTAS2006)*. Tokyo, Japan

Svedberg M, Pettersson A, Nilsson S, Bergquist J, Nyholm L, Nikolajeff F, Markides K 2003 Sheathless electrospray from polymer microchips. *Anal. Chem.* **75**, 3934–40

Tekin E, Gans B-J, Schubert U S 2004 Ink-jet printing of polymers – From single dots to thin film libraries. *J. Mater. Chem.* **14**, 2627–32

Thibault C, Sverac C, Trevisiol E, Vieu C 2006 Microtransfer molding of hydrophobic dendrimer. *Microelectron. Eng.* **83**, 1513–16

Treise I, Fortner N, Shapiro B, Hightower A 2005 Efficient energy based modeling and experimental validation of liquid filling in planar micro-fluidic components and networks. *Lab Chip* **5**, 285–97

Truckenmuller R, Rummler Z, Schaller T H, Schomburh W K 2002 Low-cost thermoforming of micro fluidic analysis chips. *J. Micromech. Microeng.* **12**, 375–9

Tsai Y C, Jen H P, Lin K W, Hsieh Y Z 2006 Fabrication of microfluidic devices using dry film photoresist for microchip capillary electrophoresis. *J. Chromatogr. A* **1111(2)**, 267–71

Unger M A, Chou H, Thorsen T, Scherer A, Quake S 2000 Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113–16

www.usplastics.com (accessed February 1, 2006).

Voit W, Zapka W, Belova L, Rao K V 2003 Application of inkjet technology for the deposition of magnetic nanoparticles to form micron-scale structures. *IEEE Proc. Sci. Meas. Technol.* **150**, 252–6

Verpoorte E, Rooij N 2003 Microfluidics meets MEMS. *Proc. IEEE* **91(6)**, 930–53

Wallace D B 1989 A method of characteristics model of a drop-on-demand ink-jet device using an integral method drop formation model. *The American Society of Mechanical Engineers, Winter Annual Meeting*, San Francisco, CA, USA, pp. 1–9

Weigl B H, Bardell R L, Schulte T, Battrell F, Hayenga J 2001 Design and rapid prototyping of thin-film laminate-based microfluidic devices. *J. Biomed. Microdevices* **3(4)**, 267–74

Weigl B H, Bardell R L, Cabrera C R 2003 Lab-on-a-chip for drug development. *Adv. Drug Delivery Rev.* **55**, 349–77

Wensink H, Jansen H V, Berenschot J W, Elwenspoek M C 2000 Mask materials for powder blasting. *J. Micromech. Microeng.* **10**, 175–80

Whiteside B R, Brown E C, Ono Y, Jen C K, Coates P D 2005 Real-time ultrasonic diagnosis of polymer degradation and filling incompleteness in micromoulding. *Plastics Rubber Composites* **34**, 387–92

Whitesides G M, Ostuni E, Takayama S, Jiang X, Ingber D E 2001 Soft lithography in biology and biochemistry. *Annu. Rev. Biomed. Eng.* **3**, 335–73

Xia Y, Whitesides G M 1998 Soft lithography. *Angew. Chem. Int. Ed.* **37**, 550–75

Xia Y, Rogers J A, Paul K E, Whitesides G M 1999 Unconventional methods for fabricating and patterning nanostructures. *Chem. Rev.* **99**, 1823–48

Xia Y, Kim E, Whitesides G M 2003 Micromolding in capillaries: Applications in material science. *J. Am. Chem. Soc.* **118**, 5722–31

Xia D, Le T V, Wirth M J 2004 Surface modification of the channels of poly(dimethylsiloxane) microfluidic chips with polyacrylamide for fast electrophoretic separations of proteins. *Anal. Chem.* **76**, 2055–61

Xie L, Chong S C, Premachandran C S, Pinjala D, Iyer M K 2005 Disposable bio-microfluidic package with passive fluidic control. *Proc.7th Electronics Packaging Technology Conference,* Singapore, pp. 93–7

Xu X, Liu S, Ju H 2004 Disposable biosensor based on a hemoglobin colloidal gold-modified screen-printed electrode for determination of hydrogen peroxide. *IEEE Sens. J.* **4(4)**, 390–4

Yahaya M, Salleh M M, Hoe T K 1998 Fabrication of photodiode by screen printing technique. *ICSE'98 Proc.*, Bangi, Malaysia, pp. 254–9

Yamahata C, Chastellain M, Hofmann H, Gijs M A M 2003 A ferrofluid micropump for lab-on-a-chip applications. *Proc. 17th Eurosensors Conf.,* Guimaraes, Philippines, September 21–24, pp. 26–7

Yang H, Pan C T, Chou M C 2001 Ultra-fine machining tool/molds by LIGA technology. *J. Micromech. Microeng.* **11**, 94–9

Yao D, Kim B 2002 Simulation of the filling process in microchannels for polymeric materials. *J. Micromech. Microeng.* **12**, 604–10

Yilbas B S 1997 The analysis of CO_2 laser cutting. *Proc. Inst. Mech. Eng. B* **211**, 223–32

Zhao Y, Cui T 2003 Fabrication of high-aspect-ratio polymer-based electrostatic comb drives using the hot embossing technique. *J. Micromech. Microeng.* **13**, 430–5

Biographies

Bruce K. Gale, currently Director of the Utah State Center of Excellence for Biomedical Microfluidics and an Assistant Professor of Mechanical Engineering at the University of Utah since 2001, has been working in the area of microfluidics and micro-total-analysis systems (μ-TAS) for the past decade. His interests include lab-on-a chip devices that require a variety of microfluidic components for the completion of complex and challenging medical and biological assays. These components fall into three broad categories: sample preparation, sample separation or analysis, and detection. His work has recently involved micromachined particle separation systems and detectors, microarray manufacturing methods, and sensors related to these applications. Specifically, he is working to develop a microfluidic toolbox for the rapid design, simulation, and fabrication of devices with medical and biological applications. The ultimate goal is to develop platforms for personalized medicine, which should allow medical treatments to be customized to the needs of individual patients.

Mark A. Eddings is a Doctorate student in Bioengineering at the University of Utah. He also received his undergraduate degree in Mechanical Engineering from the University of Utah in 2004. While working on his undergraduate degree, he worked as an Engineering Intern for Medron, a medical device manufacturing company, specializing in percutaneous catheters and other related devices. He has been an NSF IGERT fellow and is currently working in the Utah State Center of Excellence for Biomedical Microfluidics as a research assistant. His research interests focus on clinical applications of microfluidic immunoassays for therapeutic monitoring and point-of-care diagnosis. His previous research consisted of micropump development for lab-on-a-chip applications.

Scott O. Sundberg received the BS degree in mechanical engineering from the University of Utah in 2004. In 2005, he received the NSF IGERT fellowship for his current Ph.D. dissertation research in solution-phase DNA melting analysis for SNP genotyping and mutation scanning using nanoliter volumes. During undergraduate studies, he was an intern at Becton, Dickinson and Company, Salt Lake City, Utah. While there, he worked on a less painful injection device for transdermal drug delivery. He acknowledges his wife, Stacy, for the support she continually gives him.

Andrew Carter Hatch was born in Provo, Utah, on April 10, 1981. He graduated from the University of Utah with a BS in Biomedical Engineering having a focus on MEMS technology applications. In 2005, he performed a 14-week internship at Sandia National Laboratories, Livermore, CA, USA, where he researched the integration of on-chip protein concentration using a polymer membrane, with liquid gel protein separations in glass microchips. From 2005 to 2006, he performed research at the University of Utah, Center of Excellence for biomedical microfluidics, generating multilayered PDMS microspotters and testing their function and performance in generating protein arrays. He is currently working toward his Ph.D. in Biomedical Engineering at the Henry Samueli School of Engineering at the University of California, Irvine. After acquiring a Ph.D. degree he intends to pursue his research goals in MEMS technology and continue his career in academics as a university professor.

Jungkyu Kim was born in Busan, South Korea, in 1974. He received the BS in Biomedical engineering from Inje University in 2000 and the MS in Mechanical Engineering from Sogang University, South Korea, in 2002. In 2000, he joined the biomedical research center at the Korea Institute of Science and Technology (KIST) as a research scientist. He was involved in the development of an artificial bone and joint using tissue engineering techniques for 4 years in KIST. Between 2003 and 2005, he received the Korea Science and Engineering Foundation (KOSEF) fellowship. Currently, he is pursuing a Ph.D. degree in Bioengineering at University of Utah. He is working on a project that involves the implementation of a microscale system for DNA extraction and amplification in Utah State Center of Excellence for Biomedical Microfluidics. This project uses a nanoporous membrane that will be implemented in a microscale system to extract the DNA, amplify the DNA using PCR, then monitor the reaction in real time using optics.

1.13 Micromachining

Holger Reinecke and Claas Müller, Department of Microsystems Engineering - IMTEK, University of Freiburg, Freiburg, Germany

1.13.1 Introduction

Mechanical machining processes have been used for decades in the production of components from different materials. In particular, the method used for the processing of metals and alloys such as copper, brass, bronze, and steel is used for the fabrication of single products, such as tools, as well as standard products.

In the field of precision engineering, components requiring micrometer precision have been produced for several decades. Most often, these are components in the watch and the clock as well as the optical industry, or tools that are needed for the manufacture of plastic components that are used within these industries.

Machining processes such as precision milling and turning and electroerosive processes such as wire and spark erosion are used in manufacturing processes. In processes that make use of machinable materials that cannot be subjected to any thermal change during processing, electrochemical machining (ECM) is used.

In microsystem technology, because of the increased use of plastic injection molding and hot embossing there is a great need for tools made from metal. This need can be satisfied by a modified LIGA (LIGA is a German acronym for X-ray lithography (X-ray lithographie), electroplating (Galvanoformung)) process, where a cast made by employing lithographic or etching techniques is transferred to a tool, mostly from nickel, by means of electrochemical processes. These molds have quite a few disadvantages such as higher corrosiveness, lower hardness, and lower Young's modulus when compared to common steel molds. Furthermore, due to the process being used, their fabrication is quite time-consuming. Mechanical fabrication of the molds is a way out.

Because of the increasing need of and the progress in machine technology, the existing methods of mechanical precision manufacturing were adapted specifically to the demands of microtechnology, or were specifically developed for these demands.

1.13.2 Ultra-Precision Machining

Ultra-precision milling (UPM) is one way to fill the gap between conventional machining technologies and the semiconductor industry. UPM makes it possible to conform to the stringent requirements for optical surfaces. When manufacturing mold inserts for the mass production of micro-optical components, most conventional processing technologies reach their limits. On the other hand, technologies of the semiconductor industry offer no profitable way to generate free-form geometries (Brinksmeier and Gläbe 2002, Brinksmeier *et al.* 2002, Fleischer *et al.* 2004, Grimm *et al.* 2004, Ritzhaupt-Kleissl *et al.* 2005, Weck 2003).

1.13.2.1 Requirements of a UPM Machine

In order to realize surfaces with roughness less than 10 nm and structures with a precision higher than 100 nm, both the machine itself and its environment have to conform to high standards. The machine has to be set up in an air-conditioned room with a temperature of 20°C and constancy better than 0.5 K. Furthermore, care has to be taken that no environmental vibrations are transmitted to the machine. To meet this requirement, the machine should have a mass of about 6 tons and so-called self-leveling machine legs. In order to avoid vibrations during the machining process, such machines are set up using air-bearing milling spindles that can be fine balanced and that have engine speeds up to 150 000 rpm. All axes too use air bearings and are positioned using laser interferometers with nanometer precision. **Figure 1** shows a UPM machine manufactured by Fraunhofer Gesellschaft, IPT Aachen.

With the aid of a camera system, the whole machining process can be observed even from a remote distance. The UPM that is present at the chair has a large diameter and a 400-mm machining spindle, with a maximum engine speed of 3500 rpm.

Figure 1 Ultra-precision milling (UPM) machine.

It enables fixation of the workpiece as well as of the tool, depending on the machining technique. It therefore offers the possibility of turning and facing on the same machine. The milling process can be realized using high-frequency spindles that have an engine speed ranging from 1000 to 150 000 rpm (Table 1).

By means of this ultra–precision (UP) machine, highly precise 3D structures, free-form surfaces, and optical structures from nonferrous metals can be fabricated. Because UPM employs diamond tools, machining metals containing iron cannot be machined. The carbon from the diamond tool diffuses into the steel during processing, causing excessive wear of the tool. Figure 2 shows a diamond milling tool with a diameter of 250 μm. Since the clamping accuracy of such a milling cutter is about 5 μm, each tool has to be rotationally calibrated using the integrated tool gauging system in order to realize structures with the desired precision.

During UP machining, the surface quality of the tool is transferred directly to the workpiece. It is therefore necessary that the tool have a surface quality in the range of 10 nm. Cooperation with important tool suppliers guarantees the highest standard.

During machining of surfaces, standard cooling agents cannot be used. These fluids would flood the machine and damage the air bearings of the axles and the spindles. Fluids with high vapor pressures have to be vaporized onto the surface of the material that is to be milled. At the same time, excessive cooling of the workpiece and/or the tool has to be avoided since this could lead to dimensional deviations of up to several micrometers. Before machining, the whole setup as well as the workpiece has to acclimatize, as even handling of the workpiece induces thermal energy in the complete system.

During processing, the following machining techniques can be used in principle:

- Turning and facing on the large machining spindle
- UPM and fly cutting on the high-frequency spindle

The different machining techniques are described in greater detail below.

Table 1 Important process parameters for ultra-precision machining equipment

Application	Turning and milling, four-axis machining
Manufacturer	FhG-IPT
Resolution	10 nm
Rotational speed	Up to 150 000 rpm
Tools	>250 μm (diamond), >50 μm (carbide)
Material	Nonferrous metals, polymers
Surface quality	$R_a = 10$ nm
Precision	0.5 μm

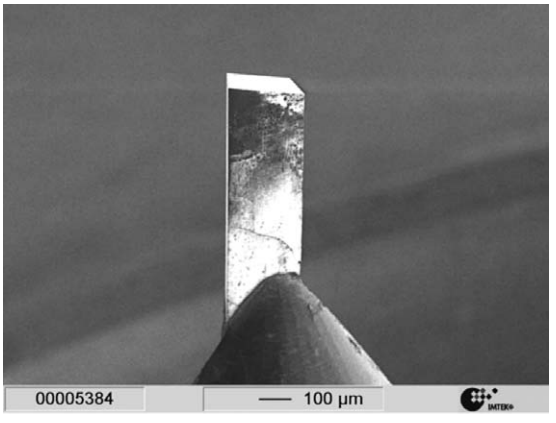


Figure 2 A single crystalline diamond tool.

1.13.2.1.1 Turning

During UP turning, the workpiece is fixed to the spindle. For smaller structures, this can be realized using vacuum as well as through mechanical clamping. Figure 3 shows a machining spindle with a tool holder placed in front. The holder can be adjusted in

Figure 3 Detailed view of the machining spindle.

height with micrometer precision so that for each tool, an optimized machining position can be arranged.

During turning, care has to be taken that the spindle, including the workpiece, is always properly balanced, since a slight unbalance can damage the spindle. By using this UP turning processing technique, rotationally symmetric structures with a precision in the range of 10 nm and having a diameter of up to 400 mm can be fabricated. Depending on the tool used, a wide variety of structures can be realized. The scope ranges from simple spherical and aspherical lenses to diffractive lenses such as a Fresnel lens. **Figures 4** and **5** show a UPM polymer lens with a diameter of 20 mm and a Fresnel lens with a diameter of 40 mm. Both structures have a surface quality less than 10 nm.

Table 2 lists typical process parameters for different materials with regard to turning. For UP machining as well, there is a difference between premachining and finishing. During premachining, a maximum feed depth of 50 µm can be achieved. For finishing, this is reduced to feed depths of 2 µm. In case the chip thickness falls below a material-dependent value, a continuous chip is no longer produced. This directly influences the surface quality. The machining parameters must be developed according to each material combination.

Table 2 Typical process parameters for different materials with regard to turning

Material	*Premachining*	*Finishing*
OFH-Copper	20–40 mm min^{-1}, 1000 rpm, 50 µm	3–5 mm min^{-1}, 1500 rpm, 5 µm
Brass	30–50 mm min^{-1}, 1000 rpm, 50 µm	3–5 mm min^{-1}, 1500 rpm, 5 µm
Red brass	30–50 mm min^{-1}, 1000 rpm, 50 µm	3–5 mm min^{-1}, 1500 rpm, 5 µm
Germanium	10–20 mm min^{-1}, 1000 rpm, 20 µm	1–2 mm min^{-1}, 1000 rpm, 2 µm
Silicon	10–20 mm min^{-1}, 1000 rpm, 20 µm	1–2 mm min^{-1}, 1000 rpm, 2 µm

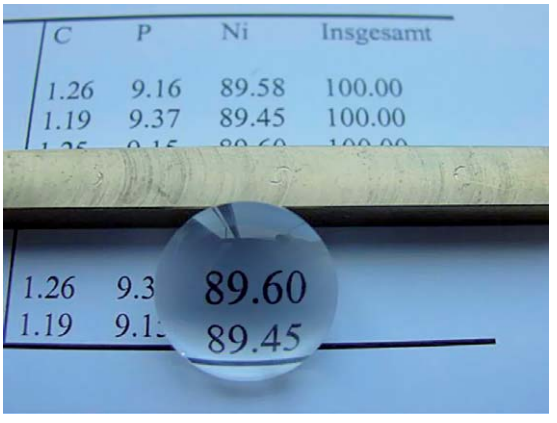


Figure 4 Example sphere in polycarbonate.

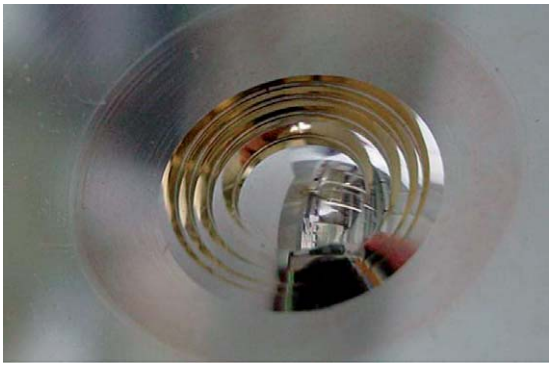


Figure 5 Example of Fresnel lens in brass.

1.13.2.1.2 Facing

During facing, a diamond tool with a relatively large radius of several millimeters is mounted on the spindle. For processing, the workpiece is moved in front of the spindle. This technique achieves parallelisms of less than 1 µm over a distance of 400 mm, with surface qualities less than 10 nm (**Figures 6** and **7**).

1.13.2.1.3 Ultra-precision milling

During UPM, standard tools made entirely from carbide steel and with a minimal diameter of 50 µm as well as diamond milling cutters with a minimal diameter of 250 mm can be used. By using diamond

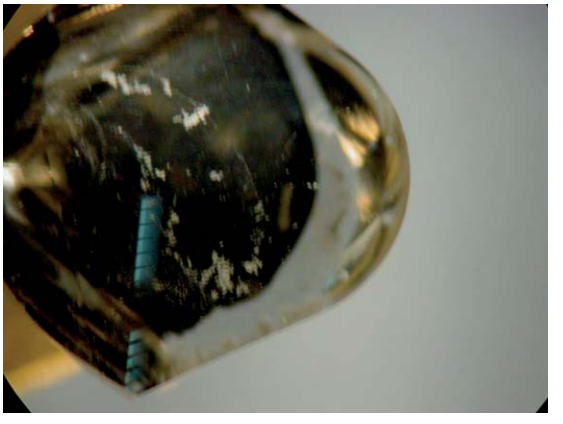


Figure 6 Diamond milling cutter with a radius for facing.

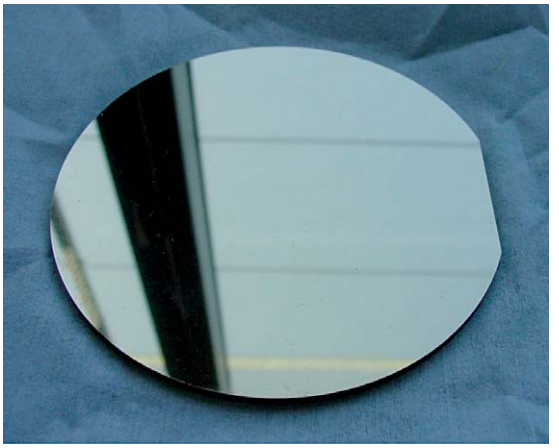


Figure 7 Surface quality after facing of a brass wafer.

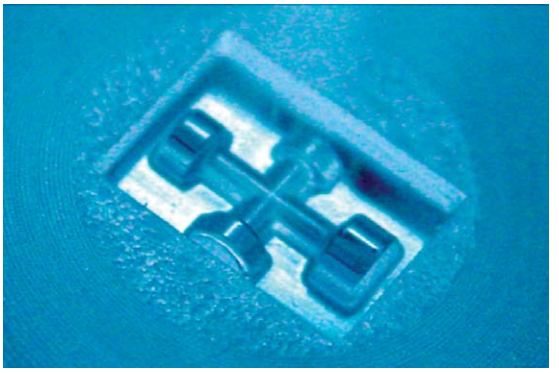


Figure 8 Ultra-precision (UP) machined free-form surface in polycarbonate.

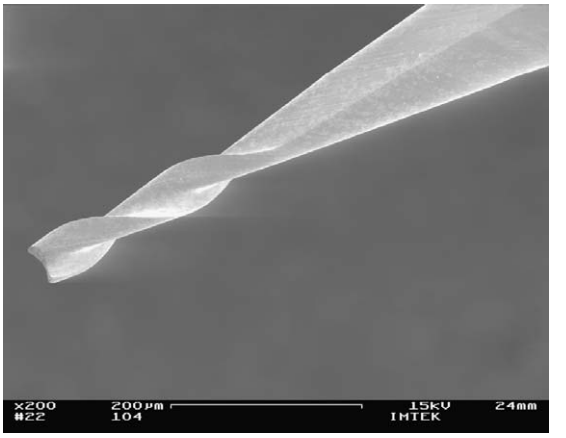


Figure 9 Carbide steel milling cutter with a diameter of 50 µm.

Table 3 Typical process parameters for ultra-precision milling

Material	*Premachining*	*Finishing*
OFH-Copper	50–100 mm min^{-1}, 50 000 rpm, 30–40 µm	20–50 mm min^{-1}, 50 000 rpm, 5 µm
Brass	50–100 mm min^{-1}, 50 000 rpm, 20 µm	10–50 mm min^{-1}, 50 000 rpm, 5 µm
Polymer, PMMA, PC	50–100 mm min^{-1}, 50 000 rpm, 20 µm	10–50 mm min^{-1}, 50 000 rpm, 5 µm

PC, polycarbonate; PMMA, polymethylmethacrylate.

milling cutters, free-form surfaces with optical surface quality can be fabricated. An example is given in Figure 8. The structure was worked in polycarbonate using a 250-µm-diameter diamond milling cutter.

With the aid of carbide steel milling cutters that are currently available with a diameter of 50 µm, microfluidic systems with minimal channel width of 50 µm can be fabricated directly in plastics (Becker *et al.* 1989). Direct machining of the polymer material offers a fast and economical way to transfer fluidic designs into functional structures for their efficient characterization. Figure 9 shows a 50-µm-diameter carbide steel milling cutter seen through a scanning electron microscope (SEM). By using a fourth axis, even complex 3D structures can be fabricated with a precision of 0.5 µm.

Table 3 lists typical processing parameters of UPM using a 500-µm-diameter diamond tool. As can be seen from the table, for such processing, engine speeds of 50 000 rpm are needed. If the tool diameter is reduced further, engine speeds of more than 100 000 rpm might become necessary in order to obtain the best possible machining results.

1.13.2.1.4 Fly cutting

When fly cutting, the tool does not rotate around its own axis but is eccentrically mounted, so that the tool blade follows a circular path. Figure 10 shows a typical setup used for fly cutting.

Fly cutting offers the possibility of fabricating parallel running structures with an extremely high sidewall quality. This technology also offers surface qualities up to 100 nm. In Figures 11 and 12, V-grooves with an angle of 90° machined into a brass foil are shown. This structure serves as the basis for a retro reflector that has been made keeping in mind the great precision required for the complete structure. By a smart arrangement of the single foils, the complete

geometry can be fabricated. By nickel electroplating, the mold insert for injection molding can be made from the brass master. In this application, the replicated parts have an angular deviation of less than 0.01°.

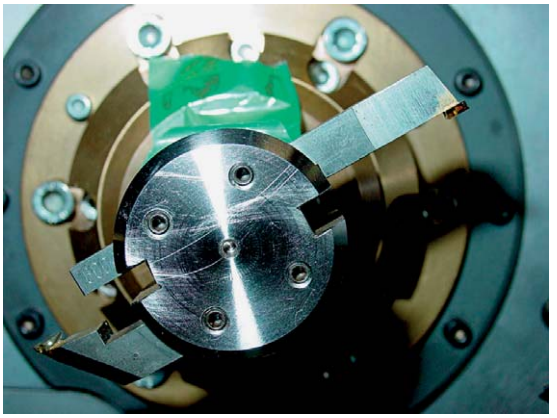


Figure 10 Spindle with tool for fly cutting.

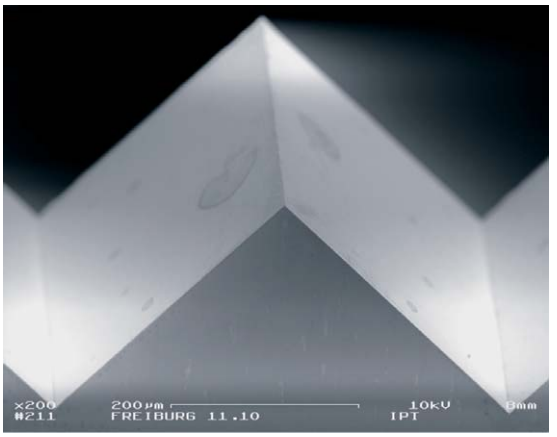


Figure 11 V-grooves, fabricated by fly cutting.

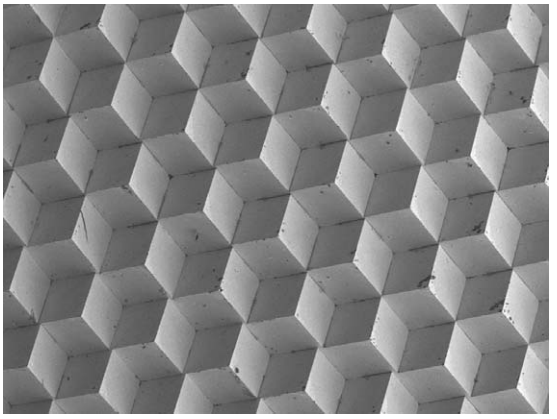


Figure 12 Mounted molding tool for the retro reflector.

Table 4 Typical process parameters for fly cutting

Material	*Premachining*	*Finishing*
Brass	25–30 mm min^{-1}, 1400 rpm, 15 µm	5 mm min^{-1}, 1400 rpm, 5 µm
Polymer, PMMA, PC	30–45 mm min^{-1}, 1400 rpm, 30 µm	5–10 mm min^{-1}, 1400 rpm, 5 µm

PC, polycarbonate; PMMA, polymethylmethacrylate.

Table 4 lists typical processing parameters for fly cutting. For this machining technique as well, a minimal chip size is mandatory, since otherwise no optimal machining is possible.

1.13.3 Electrodischarge Machining

1.13.3.1 Physical Principle

Electrodischarge machining (EDM or spark erosion machining) processes aid in material removal of electrically conductive materials through electric discharges (spark) in an electrically nonconducting liquid (dielectric). Spark erosion processing is characterized by successive, temporally separated, nonstationary electric discharges (sparks) taking place in a dielectric fluid (König 1990, N.N. 1975). Spark erosion is a shaping technology where the contour of the tool electrode is transferred into the workpiece by making use of an additional process gap (König 1990). Shaping of the workpiece electrode occurs through spatially and temporally separated, highly dynamic discharges of high frequency (Spur and Stöferle 1984). In a dielectric liquid, the workpiece and the tool electrode approach each other, so that a very small gap remains between them. When a voltage is applied, depending on the disruptive strength of the dielectric used, an energy-rich plasma channel can develop if the process gap is small enough. Because of the highly complex physical processes occurring during a spark discharge and the high discharge rate, these are hard to understand and describe. In order to understand the ablation mechanism of spark erosion and to enhance understanding of the process, experiments are often carried out using single discharges (König 1990, Nöthe 2001, Spur and Stöferle 1984, Wertheim 1975). The macroscopically visible ablation consists of the sum of the single microvolumes that are removed during every single discharge. The single discharge describes the processes that take place between the application of the voltage to the electrodes and the breakdown of the dielectric.

Generalizing the results given above is unfortunately not possible given that processing conditions in the process gap are very complex and several factors interfere with each other (Schönbeck 1992, Wertheim 1975). A great number of scientists came up with various theories about the mechanism of the ablation process for spark erosion processing. However, these often lead to inconsistent explanations of the processes running (Schoepf 2001, Spur and Stöferle 1984). The electrothermal erosion theory, developed by the Russian scientists Lazarenko and Zolotych, is the most commonly accepted (Lazarenko and Lazarenko 1944, Zolotych 1955). This theory proposes that because of the high temperatures in the discharge channel, melting and vaporization processes take place at the material surface of the electrodes in the proximity of the channel base points. For temperatures inside the discharge channel, values in the range of 2000–100 000 K are given by different authors. Material ablation at the workpiece is achieved by ejection of the molten metal. According to electrothermal erosion theory, a single discharge can be subdivided into three main phases: the buildup, the discharge, and the decomposition phase (Dauw 1985, van Dijck 1973, Mirnoff 1968).

Figure 13 shows the processes taking place during a single discharge between the anode and the cathode, along with the corresponding voltage and current curves.

1.13.3.1.1 Buildup phase

The processes necessary for the ionization of the discharge path and the development of the plasma channel are called the buildup or the ignition phase. In this phase, the ionization of the discharge path and the channel buildup take place up to the moment when a current breakthrough inside the discharge channel occurs (Spur and Stöferle 1984). The dielectric electrically insulates the two electrodes from each other. When a DC voltage is applied between the workpiece and the tool electrode, the highest electric field occurs at the smallest gap separating both. When the voltage inside the process gap exceeds the breakdown strength of the dielectric, a discharge channel forms (Förster 1979). The electrons expelled from the cathode are accelerated toward the anode under the

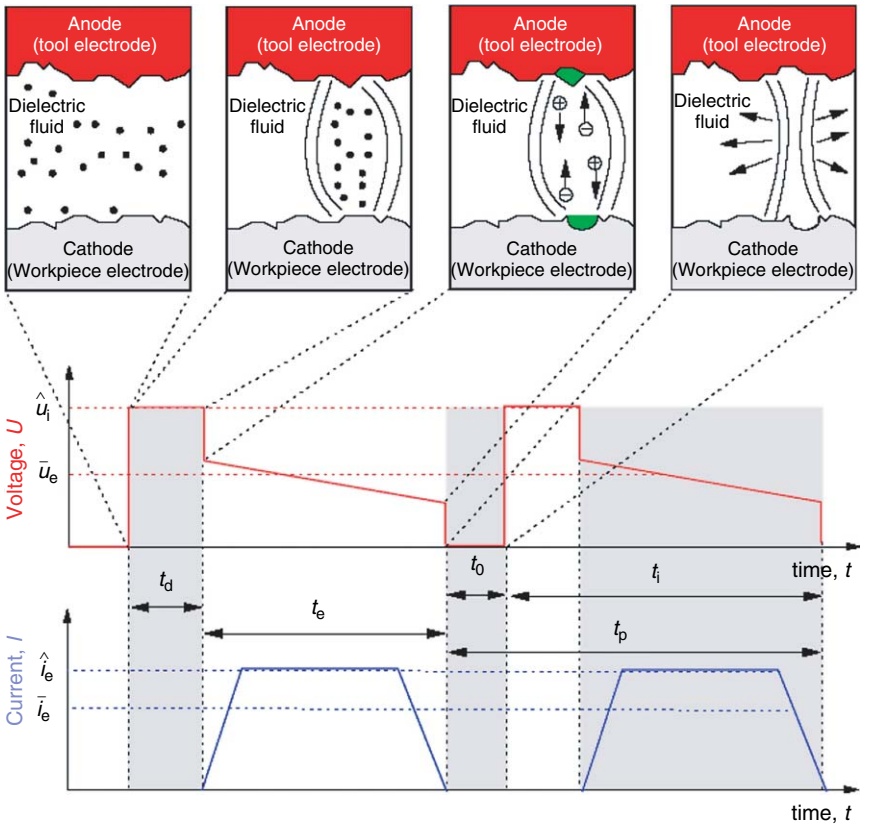


Figure 13 Electrodischarge machining (EDM) process.

influence of the electric field. Because of the ionization resulting from the collision of the molecules in the dielectric with the electrodes, the number of charge carriers increases greatly. Through formation of ions, an anode- as well as a cathode-directed plasma channel branch develops.

The electrons hit the anode considerably earlier than the positive ions hit the cathode (Finkenberg *et al.* 1956, Green 1987, Siebers 1993, Weitzel and Pompe 1949). The anode vaporizes partly due to the impact of the electrons just before the actual development of the plasma channel. However, the cathode mostly remains unaffected. In this phase, a large time-dependent current and voltage variation takes place (König 1990).

1.13.3.1.2 Discharge phase

During the subsequent phase – the discharge phase – current and voltage remain nearly constant. In this phase, according to the thermal erosion theory, material ablation takes place. In the discharge zone, electrical energy is converted into thermal energy.

Energy conversion is mainly affected by the physical characteristics (electrical and thermal conductivity, specific heat, melting and boiling temperature, and specific melting and vaporizing energy) of the tool electrode used (Jutzler 1982). The energy conversion in the discharge zone results in very high temperatures at the boundary of the plasma channel and the tool electrode. Consequently, these boundaries start to melt and the material partly vaporizes. Part evaporation of the material volume leads to a continuously growing gas bubble, which expands with an initial velocity of approximately $50\,\mathrm{m\,s^{-1}}$ (König 1990). When temperatures in the discharge channel are in a range far greater than 10 000 K, evaporation, dissociation, and ionization of all the components involved in the discharge process takes place. At very short discharge durations, up to 80% of the crater volume can vaporize. At longer discharge durations, melting processes use the greater energy part (Förster 1979, Jutzler 1982, Zolotych 1971). According to Förster (1979), local pressure can increase to 15 MPa. Hensgen (1984) found minimal discharge pressures of 50 MPa by experimenting on thin disks of lead. Using a self-developed model, Eubank *et al.* (1993) calculated pressures between 35 and 610 MPa and temperatures of 20 000–53 000 K for short discharge durations. Tsuchiya and Inoue (1982) determined by experiments and calculations a maximum pressure of 280 MPa and a maximum temperature of 87 000 K. For pressures of gaseous discharges inside the plasma channel, literature gives values of 100 MPa (Finkenberg *et al.* 1956).

1.13.3.1.3 Decomposition phase

During the third phase – the decomposition phase induced by the current shutdown – the plasma channel and the gas bubble collapse, and the melted and partly vaporized material is thrown out of the discharge zone (König 1990). After the ablated particles are washed out from the processing gap, the dielectric resolidifies. In order to enable the next discharge process, the working volume is deionized to a great extent by the fluid medium. The spark extinguishes after shutdown of the voltage, and the process can be repeated in another region (König 1990, Kurr 1972, Mirnoff 1965).

Material removal occurs mainly at both electrodes, although it should occur mainly at the workpiece electrode. An economical usage of spark erosive machining processing is possible only if the material ablation rates at the anode and at the cathode are different (König 1990, Lazarenko 1974, Lazarenko and Lazarenko 1944, Spur and Stöferle 1984, Zolotych 1955, 1957). At very short discharge durations, temperatures at the positive electrode are higher than at the negative electrode. Because of the high discharge rate of the electrons from the material's surface, there is a surplus of electrons at the beginning of the process. As ablation is stronger at the positive electrode, the tool is polarized negatively and the workpiece positively. At long discharge durations, approximately the same number of positive and negative particles are removed from the electrodes. Since a positively charged particle has a significantly larger mass than does a negatively charged one, the impact on the negative electrode results in higher temperatures and therefore in stronger ablation. Thus, long discharge durations result in positive polarization of the tool electrode and in negative polarization of the workpiece electrode (Schoepf 2001, N.N. 1985). By appropriate selection of the tool electrode's material and by varying the adjustable parameters, it is possible to achieve a strong asymmetry between ablation of workpiece and tool (König 1990, Spur and Stöferle 1984).

1.13.3.2 Process Parameters

The parameters for the description of the erosion process can be divided into electrical and technological ones.

1.13.3.2.1 Electric process parameters

The electrical processes during spark discharge at the discharge path are characterized by characteristic voltage and current progressions. The ones important for characterization and valuation of the process are defined as follows (König 1990, N.N. 1976):

The discharge duration t_e is the time interval of the current through the ignited discharge path. The ignition delay time t_d is the time between the switching on of the voltage pulse and the breakthrough of the discharge path, i.e., up to the increase in current. The pulse duration, or pulse width, t_i is the time when the voltage pulse is switched on. It is equal to the sum of discharge duration and the ignition delay time:

$$t_i = t_e + t_d \quad [1]$$

The pause duration t_0 is the time interval between two voltage pulses. The period duration t_p is the time between the switching on of a voltage pulse and the switching on of the following pulse. It corresponds to the sum of the pulse and the pause duration:

$$t_p = t_i + t_0 \quad [2]$$

The pulse frequency f_p is the number of voltage pulses per unit time. It is the reciprocal of the period duration:

$$f_p = \frac{1}{t_p} \quad [3]$$

The discharge frequency f_e is the number of actual ignited discharges inside the discharge path per unit time. The open-circuit voltage u acts as a peak value at the discharge path when no current flows. The discharge voltage u_e occurs at the discharge path, when the discharge has been ignited and current flows. The discharge current i_e flows during a discharge through the discharge path. The discharge energy W_e is the energy converted inside the discharge path during a discharge:

$$W_e = \int u_e(t) i_e(t) \mathrm{d}t \approx \bar{u}_e \bar{i}_e t_e \quad [4]$$

The duty cycle τ is the pulse width divided by the period duration:

$$\tau = \frac{t_i}{t_p} \quad [5]$$

The frequency λ is the discharge frequency divided by the pulse frequency:

$$\lambda = \frac{f_e}{f_p} \quad [6]$$

The electric parameters might indirectly comprise also the conductance and/or the specific resistance coefficient of the dielectric. By using water as a dielectric, its resistance coefficient can be adjusted to the processing task by continuous deionization. When hydrocarbons are used, a significant change in the processing task also makes a change of the dielectric necessary, as hydrocarbons have a constant coefficient of resistance that is not adjustable.

1.13.3.2.2 Technological process parameters

With the aid of the following technological process parameters, the quality of the erosion process can be described. The most important parameters for rating the efficiency of the erosion process, according to the VDI 3400 scale and König (König 1990, N.N. 1976, Stempel 1977), are as given below:

The ablation rate V_W is the workpiece volume removed per unit time. The wear rate V_E is the tool electrode volume removed per unit time. The relative wear λ is the wear rate divided by the ablation rate:

$$\lambda = \frac{V_E}{V_W} \quad [7]$$

The use of this parameter is recommended for gathering information about material-related behavior and for comparison with values from literature (Stempel 1977). Other technological parameters are the reproduction accuracy, the average of the cutting line, and various roughness parameters. The average cut track s_m is the average distance of the cut surfaces. They limit the minimum inner radius.

1.13.3.3 Procedure Variants

Spark erosion processing can be categorized as follows:

- Spark erosion by sinking with a shaped tooling electrode (spark erosion)
- Spark erosive cutting by running wire (wire erosion)
- Spark erosion by single tools (erosive milling and drilling)

Table 5 lists typical parameters and limits for EDM processes.

1.13.3.3.1 Spark erosion by sinking

Spark erosion by sinking is characterized by the simultaneous relative movement between tool

Table 5 Typical process parameters and limits for electrodischarge machining

Application	*Spark erosion*	*Wire erosion*
Applied voltage (V)	80–160	80–160
Precision (μm)	$\leq \pm 1$	$\leq \pm 1$
Roughness, R_a (nm)	200	100
Minimum structures (μm)	20 (lines), 20 (spaces)	20 (lines), 40 (spaces)
Maximum aspect ratio	15 (lines), 25 (spaces)	30 (lines), 80 (spaces)
Minimum inner radius (μm)	10	20
Minimum tool dimensions (μm)	10	30

electrode, workpiece, and feed movement. Movement of the tool electrode or of the workpiece can generate this relative movement. Furthermore, an additional movement can superimpose this relative movement (König 1990, N.N. 1976). The tool electrode represents the negative image of the body volume minus the process gap (N.N. 1976, Stempel 1977). Engraving can be seen as the simplest form of spark erosion by sinking, since the tool electrode is fed only in a single direction. By using this procedure, volumetric shapes are generated with the tool electrode representing the negative image of the volume to be removed, which copies itself into the workpiece with the additional process gap. According to Stempel (1977), matrices can be fabricated more economically by wire erosion. A significant enlargement of the field of application of spark erosion is the so-called planetary erosion. The conventional rectilinear sinking movement is overlapped by a spatial translation movement of the tool electrode. Further derivations of the sinking process are multiple-channel machining that is used for structuring large areas and the simultaneous processing of several components, by processing with rotating tool electrode and by processing the line of erosion (Behmer 1988, Kips 1960, König 1991, Kracht 1970, Staelens 1990).

All procedures described above can be used in principle for structuring tasks specific to microsystem engineering. The main areas of application of microspark erosion of sinking are the fabrication of air-cooling ducts in turbine blades, microreactors, injectors, ink-jet nozzles, tools for plastic molding, and punching tools for leadframes (Allen and Lecheleb 1996, Allen *et al.* 1999, Gruber 1999, Sato *et al.* 1985, Wolf 1997, Wolf *et al.* 1999).

1.13.3.3.2 Spark erosive cutting – wire EDM

The term spark erosive cutting comprises all processes where workpieces are separated and cut in or out (N.N. 1974). Normally, care has to be taken to work with as small a process gap as possible, in order to enable production of very small geometries with high accuracy. As an independent procedure variant, spark erosive cutting with wire electrode has been developed. Besides separating workpieces, this technique enables the fabrication of highly complex recesses and complex components and is used in prototyping, in tool and mold construction, as well as in series production (König 1990, Schumacher 1975, Stempel 1977).

By cutting with an uncoiling wire electrode, the desired geometry is generated by a numerically controlled relative movement between the tool electrode (uncoiling wire) and the workpiece electrode. This shaping process can realize only those geometries that correspond to the movement of a straight line in space, i.e., only regular forms (Bronstein 1999, Schönbeck 1992). **Figure 14** shows the work area of the five-axis wire erosion machine Charmilles Robofil 2020SI, from the company Charmilles Technologies SA, Geneva, Switzerland, with its most important components. By moving the upper wire duct in the direction of the axes v and u, conical shapes can be produced. With this process variant it is possible to fabricate stamps and matrices for tool construction without division of molds. Tool wear

Figure 14 Work area of a spark erosive wire erosion machine.

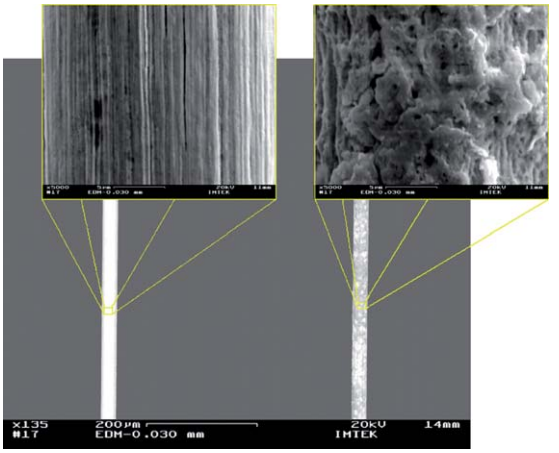


Figure 15 Comparison of unused (*left*) and used wire (*right*) (30 µm wire diameter).

during wire erosion is relatively large; however, the effect on the process is almost negligible, since the wire is renewed permanently (König 1990, Spur and Stöferle 1984). **Figure 15** (*left*) shows a tungsten wire with a diameter of 30 µm in its delivery status, and **Figure 15** (*right*) shows the wire in the state after processing. The discharge craters can clearly be identified on the surface of the wire. Because of the relatively great wear, the wire can be used only once.

1.13.3.3.3 Wire EDM with thin wires – microwire erosion

In the early 1980s a multitude of products and components from several industries required wires with diameters that are significantly thinner than the diameters of standard wires (100–250 µm). **Figure 16** shows the common standard wires and, in comparison, recently developed and used microwires. It can be seen that by using these wires a great number of new problems arise, particularly in wire handling. This is particularly the case in tool and mold construction (drawing profiles, cutting tools), in the watch and the clock industry, and in the textile industry (manufacturing of spinnerets) (Christmann 1999, Gruber 1999).

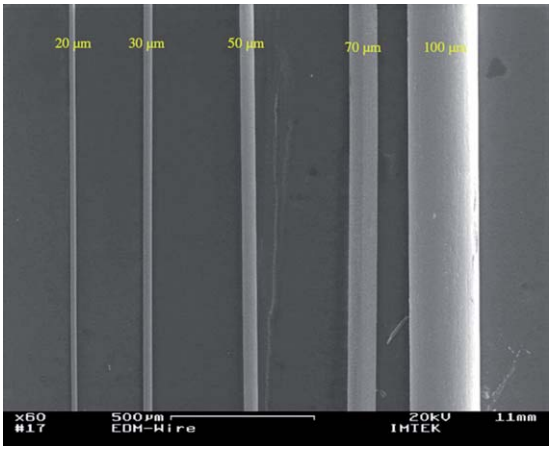


Figure 16 Comparison of different wires (20–100 µm).

By developing process energy sources and controls further and by acquiring greater expertise with regard to the EDM process itself, surface roughness in the range of $R_a = 0.1$ µm can be achieved. The best reproduction accuracies are in the range of a few micrometers (Dauw and Coppennolle 1995, Gruber 1999, Wolf *et al.* 1999).

Another very successful variation of EDM methods was done in the 1990s by Masuzawa (1990, 1998). He described a technology for manufacturing a rotationally symmetric electrode by electrowire grinding. The electrode is used as a sinking electrode in an additional manufacturing step. This technology is called wire electrodischarge grinding (WEDG) and by using this technology 3D microstructures made by EDM became possible.

1.13.3.3.4 Electrodischarge milling and drilling

Micro-EDM milling is an improvement on EDM; like numerically controlled milling, it uses a universal tool electrode of simple shape (e.g., spherical or cylindrical) that moves along a complex path to machine hard materials like tools. EDM is a machining technique that removes a small amount of workpiece material continuously by controlling the discharge energy applied between the tool electrode and the workpiece. Both electrodes are immersed through a side flushing in a dielectric fluid (e.g., deionized water or oil) that electrically insulates and cools the electrodes as well as removes the particles out of the gap. For this work, a micro-EDM milling machine (e.g., SX-100 (Sarix SA, Switzerland)) that integrates a 3D micro-EDM milling CAM concept, enabling computer-aided manufacturing, and a multiaxis motion that allows for contour-controlled machining according to a preprogrammed path in directions x, y, and z simultaneously can be used. It thereby achieves a high-precision positioning accuracy of up to ±2 µm and a resolution of 0.1 µm. By using standard tool electrodes with a simple shape, the tool manufacturing costs are reduced to a minimum; additionally, the degree

of automation is increased and simplified process planning is enhanced. The machining speed and surface finish can be adjusted by properly controlling the micromilling and discharge strategies (shape pulse, energy).

The overall fabrication sequence for creating a mold insert with microfluidic structures is illustrated schematically in **Figure 17**. Initially, a semifinished part is fabricated by CNC (computer numerical control) milling and a subsequent polishing step in order to get a surface quality of about 15 nm. The micro-EDM milling process is divided into two steps. First, a carbide electrode with a diameter of 340 µm for machining large areas is used. Next, the details are structured by fine machining with an electrode having a diameter of 45 µm. This electrode size is prepared automatically by using a built-in microwire EDM device (WEDG unit). The fine machining process operates using a new additional microfine pulse shape generator (SX-MFPS) in order to make small, deep, and precise round-shaped holes as well as channels. **Figure 18** is a SEM micrograph of the microstructure as part of a mold insert. It shows the structured geometry that results in reservoirs and microchannels after the replication process. The SEM micrograph in **Figure 18** shows a detailed view of the bend region of the microstructure.

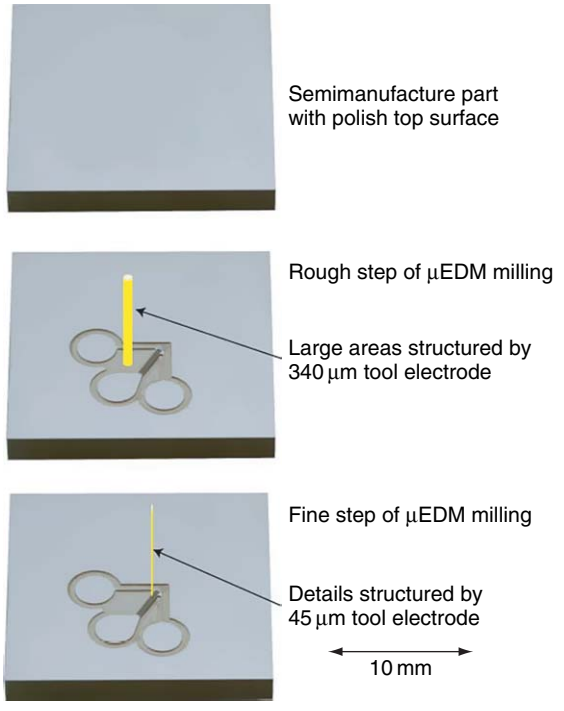


Figure 17 Schematic process of tool fabrication.

1.13.4 Electrochemical Machining

1.13.4.1 Principles

In 1930, Gusseff patented electrochemical metal dissolution for the production of boreholes and profile recesses. Further development and the increasing use of high-strength and high-temperature-resistant materials, initially in the aerospace industry, eliminated a number of problems in commonly used processes that mainly used machining. Because of the extremely high tool wear, the efficiency of the machining processes is restricted by the mechanical properties of the material to be machined (Kubeth 1965, Heitmann 1966). From the 1960s and 1970s up to the mid-1980s, a great number of investigations on and articles about the application of the ECM technique in almost all fields of the metal-working industry were published. The strongest driver for the rapid development was the increasing demand for the materials being used, since they could not be machined or proved expensive to machine with the commonly used material cutting techniques (Stanek 1970). Because of the further development of other manufacturing processes and the increased focus on ecological subjects, the ECM technique lost importance within industrial practice and did not reach the high level of usage once predicted in industrial production. Today, ECM processing techniques are still being used to a relatively large extent in the aircraft industry and in the defense industry and in stationary gas turbine construction. Because ECM makes it possible to structure highly loaded materials with no regard to their mechanical properties, ECM sinking provides very interesting possibilities in microsystem engineering: the company PEM Technologiegesellschaft für elektrochemische Fertigung mbH, Dillingen/Saar, utilized this feature in an ECM sinking process that makes use of an oscillating tool electrode. This technique became known under the brand name PEM (precise electrochemical machining). This electrochemical process can be described as a sinking process with an oscillating tool electrode. Publications about ECM processing mainly deal with investigations on the process of anodic metal dissolution, the workability of different materials, the processing rates, the achievable removal rates, the electrode correction factors, and the different electrolytes. Against the many advantages that ECM manufacturing processes offer are a number of very complex problems that make a reliable control of the process and predictions

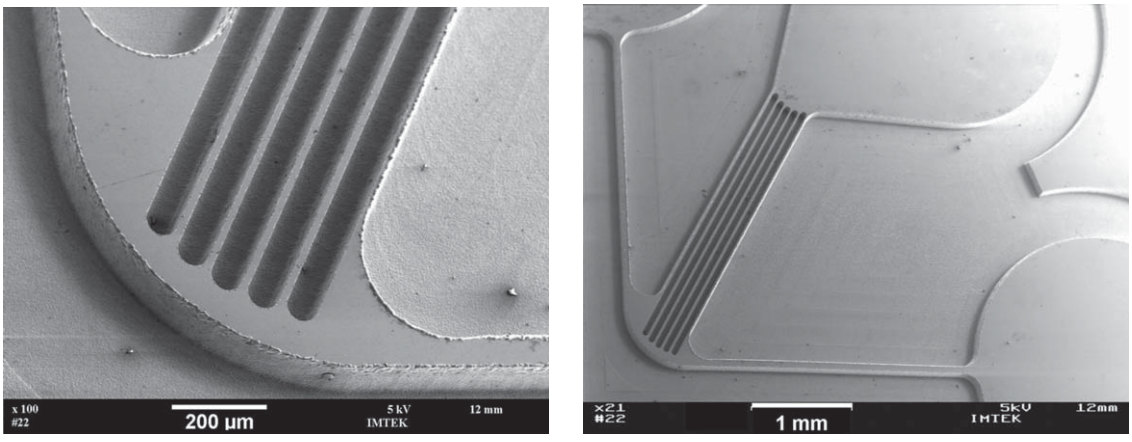


Figure 18 Detailed view of the bend region of the microstructure (*left*) and top view of the microstructure (*right*).

of the work results difficult. The great number of parameters do require the user of this technique to carry out extensive empirical investigations in order to find optimal process parameters for the desired work results.

ECM processes are a type of separation process. A primary characteristic of these manufacturing processes is nonmechanical material removal. ECM processes (**Figure 19**), i.e., the processes occurring in the working gap and the basics of the electrochemical removal process, are explained below:

The ECM process is based on the anodic dissolution of a material in an electrically conductive medium. In an electrochemical cell, a DC voltage of 5–20 V is applied. The positive pole (anode) of the power source is connected to the workpiece. The negative pole (cathode) is connected to the tool electrode. An electrically conductive medium – the electrolyte – is present between both the electrodes. The applied voltage drops across the electrolyte solution in the process gap between the tool and the workpiece electrode. The applied voltage U, according to Ohm's law, is:

$$U = U_A(j) + U_C(j) + U_{Ely}(j) \quad [8]$$

The voltage drop across the electrolyte solution is:

$$U_{Ely} = AR(j)j \quad [9]$$

By measuring the change in the galvanic voltage across the local barrier layer, the voltage drop across the anode (U_A) and the cathode (U_C) can be calculated (Forker 1966). The electrical resistance of the electrolyte solution can be determined from the specific electric conductivity χ of the electrolyte, the electrode distance s, and the electrode surface A, using the following equation (Forker 1968, Schwabe 1974):

$$R = \frac{s}{\chi}A \quad [10]$$

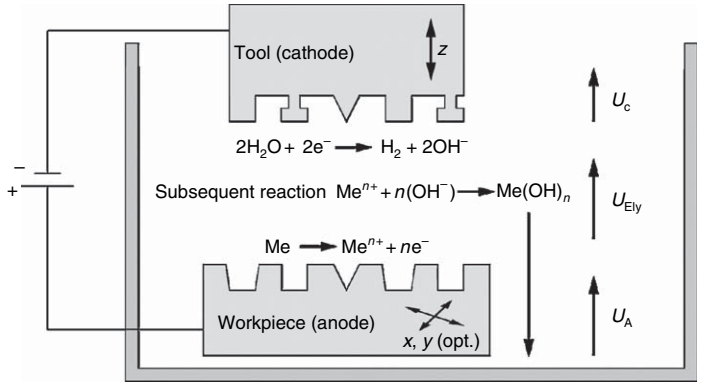


Figure 19 Electrochemical machining processes in the working gap.

If $U_D = U_A + U_C$ is substituted in eqn [8], from eqns [9] and [10] we get:

$$U = U_D + U_{Ely} = U_D + \frac{s}{\chi} j \qquad [11]$$

The current density j is given by:

$$j = (U - U_D)\frac{\chi}{s} \qquad [12]$$

The term U_D depends on the material–electrolyte combination of the anode and also on the material of the cathode (Degner 1984).

1.13.4.1.1 Anodic metal dissolution

In an electrochemical cell, two electrodes are connected to a power source and immersed in an aqueous solution. In the electrolyte, the salt disassociates completely and is present as separated cations K^{in+} and anions Sr^{in-}. Furthermore, the water is, to a small extent, dissociated into H^+ and OH^- ions. During electrolysis, the cations (K^{in+}, H^+) migrate to the cathode and the anions (Sr^{in-}, OH^-) to the anode. The following phenomena can be observed:

- At the tool electrode (anode), metal ablation takes place, which can be determined by weighing the electrode before and after the processing.
- At the workpiece electrode (cathode), depending on the electrolyte used, hydrogen might be produced.
- At the bottom of the cell, a brown coating is deposited, which on chemical analysis reveals metal hydroxide.

These observations can be described by the following chemical reactions:

At the anode, metal ions dissolute by the emission of electrons:

$$Me \rightarrow Me^{n+} + ne^- \qquad [I]$$

At the cathode, hydrogen and hydroxide are formed by the acceptance of electrons:

$$2H_2O + 2e^- \rightarrow H_2 + 2OH^- \qquad [II]$$

In a subsequent reaction, metal hydroxide is formed, which can be precipitated depending on the electrolyte used:

$$Me^{n+} + nOH^- \rightarrow Me(OH)_n \qquad [III]$$

Eqns [I]–[III] depend on the electrolyte used and on the material that has to be machined. If, for example, sodium chloride is used as an electrolyte and only pure iron is machined, eqn [I] becomes:

$$Fe \rightarrow Fe^{2+} + 2e^- \qquad [IV]$$

From eqns [II] and [IV] we get:

$$2H_2O + Fe \rightarrow Fe(OH)_2 + H_2 \qquad [V]$$

In this case, only water, but no electrolyte, is consumed. This may not be the case for other electrolytes. For example, for so-called STEM (shaped tube electrochemical machining) drilling of cooling holes in turbine blades, sulfuric acid is used. For the reactions at the anode and the cathode, eqns [II] and [IV] are valid; however, the equation for the reaction becomes:

$$H_2SO_4 + Fe \rightarrow FeSO_4 + H_2 \qquad [VI]$$

So the formation of insoluble ferrous hydroxide is suppressed because of the formation of soluble iron sulfate. Thus, the electrolyte wears itself out and must therefore be continuously monitored and replenished if necessary. For sodium nitrate, the electrolyte most often used in practice, further peculiarities exist: At the anode, the rate of oxygen production depends on the current density, as given by the following equation:

$$2H_2O \rightarrow 4H^+ + 4e^- + O_2 \qquad [VII]$$

The anodic metal dissolution, as given in eqn [I], starts only for higher current densities. The Fe^{2+} ions reduce; however, the NO_3^- oxidize Fe^{3+} ions that then precipitate as hardly soluble iron (III) hydroxide.

$$4Fe^{2+} + 2NO_3^- + 4H_2O + 2e^- \rightarrow 2NO_2^- + 4Fe^{3+} + 6OH^- + H_2 \qquad [VIII]$$

At the cathode, no or insignificant hydrogen is produced. The main reaction that takes place is the reduction of the nitrate ion to the ammonium ion.

$$NO_3^- + 7H_2O + 8e^- \rightarrow NH_4^+ + 10OH^- \qquad [IX]$$

If oxygen and hydrogen production is ignored, the complete reaction when sodium nitrate is used as an electrolyte is as given:

$$4Fe + 3NO_3^- + 11H_2O \rightarrow NH_4^+ + 4Fe(OH)_3 + 2NO_2^- + 4OH^- + H_2 \qquad [X]$$

Because of the hydroxide ions formed, the electrolyte becomes more alkaline during the process; in

addition, ammonia is formed (Degner 1984). These basic equations represent machining only for a pure material, in this case iron. For materials used in industrial practice – mostly alloys from several elements, i.e., material mixtures – the chemical reaction is very complex; sometimes the reaction is not even possible, as is evident from the attempts that have been made to directly draw a material sample from the working gap to observe the reactions, which have not succeeded. In order to learn about the reactions taking place during the machining process, one can refer to many preliminary works on analytical electrochemistry. However, only quasi-stationary processes are considered there; furthermore, mostly pure materials are examined. Thus, knowledge of process behavior of alloys and the optimal process parameters for the corresponding materials can only be gained by investigations.

1.13.4.1.2 Quantitative description of anodic metal dissolution

For the calculation of the theoretically possible specific ablation volume V_{spez}, Faraday's law

$$m = \frac{M}{zF} I t_B \qquad [13]$$

is used under the condition that at the tool electrode, only electrochemical dissolution of the anode material takes place. In eqn [13], m represents the ablated mass (g), M the molar mass (g mol^{-1}), I the current (A), t_B the machining time (s), z the electrochemical valence, and F the Faraday constant.

If reactions that are not conducive for ablation take place, as described in eqn [VII], the energy input is not completely converted for metal dissolution. If the density of the workpiece, $\rho = m/V$, is included, the equation for the ablation volume becomes:

$$V_{spez} = \frac{A_r}{z} F \rho_{Wst} \qquad [14]$$

For the calculation of the ablation volume of an alloy, eqn [14] can also be used. If the n alloying elements ($i = 1, 2, \ldots, n$) are present in the alloy in the proportion p_i, the following equation is obtained:

$$V_{spezLeg} = \frac{100}{F \rho_{Leg} \sum (p_i (z_i / A_{ri}))} \qquad [15]$$

It can be seen from eqns [14] and [15] that the valency change z greatly influences the result. Since elements with different valencies can be in a dissolved state, the determination of z for an alloy often involves extremely high cost, or is even impossible. Because determination of the electrochemical valency of alloys is often difficult, and because it cannot always be guaranteed that the energy introduced into the process is transformed completely into removal–effective reactions, the real value of the material ablation V_{Abreal} must be determined experimentally and, furthermore, must be compared to the theoretical value of the specific ablation volume, in accordance with eqns [14] and/or [15]. This quotient is referred to as the current efficiency η_j, and can be calculated using the following equation:

$$\eta_j = \frac{V_{Abreal}}{V_{spez}} \qquad [16]$$

1.13.4.1.3 Passivation

According to the principle of anodic metal dissolution, all metallic materials except for some noble metals can be dissolved, provided that suitable electrolyte solutions are used and a sufficient current density is available (Degner and Boettger 1979, König and Klocke 1997). However, for specific electrode potentials, surface (anode) oxides can develop on the workpiece, which hinders the transition of the metal ions in the electrolyte (Vetter 1961). The movement of the coreactants is also impeded or comes to a complete stop. This phenomenon is known as passivity. The development of these coatings, which are also referred to as passive layers, is influenced by the type of electrolyte, the process conditions, and the anode material. A passive layer can cause removal–ineffective reactions and therefore reduce the efficiency of the process, which is expressed through a poor current efficiency. By assuming a current density potential, the passivity can be measured by various techniques (Hoar 1967, 1968, König and Klocke 1997, Schwabe *et al.* 1958, Vetter 1958). The current density increases with increasing anode potential until it reaches a potential barrier. This potential barrier is called Flade potential after its discoverer Flade (Hensler *et al.* 1958, König and Klocke 1997). When the potential barrier is reached, the current density decreases to the level of the corrosion current density. With higher anode potentials, the dissolution of the anode starts and the current density increases once again.

For the quantitative determination of passivity, current density potential characteristics are recorded. The characteristics depend on the material

electrolyte combination used. In the lower potential range, the current density increases with increasing voltage, then declines steeply when the passivation current density is reached because of the formation of nonporous oxide layers. Because of acid electrolytes, the current flowing in the passive field is rather small and is probably used for renewal of the oxide film (Evans 1965). It is virtually independent of the potential since dissolution is a chemical process.

In the transpassive field, the development of gaseous oxygen becomes thermodynamically possible because of the discharge of hydroxide ions. However, an electron-conducting coating layer should be available (Degenhardt 1972). The passive layer gets dissolved locally, which is characterized by a pitting mechanism at the cover layer, and the current increases again (Datta and Landolt 1980). According to the conductivity of the cover layer, the metals can be categorized as electron-conducting, cover layer building, and ion-conducing cover layer-building materials (Hamann and Vielstich 1998, Hoar 1968, Schwabe 1966). For iron, nickel, chrome, and, apparently, cobalt, the effect of electron conductivity is stronger than the effect of the passive layers that are built up during the stationary passive state, such that, for a further voltage increase, oxygen is generated from a higher-valence metal oxide. For other materials (aluminum, tantalum, titanium, and zirconium), metal ions are transported through the passive layer to the dissolution side due to the poor electron conductivity of their oxides. The cover layer grows there subsequent to hydrolysis. The current does not increase if the voltage rises until disruptive discharge occurs. Because these materials are capable of charge storage, they are also referred to as condenser metals. For metals with an ion-conducting passive layer, a high overvoltage, together with the dissolution of the passive layer, is therefore necessary in order to induce an oxidation–reduction reaction (Degenhardt 1972). Degenhardt (1972) gives an overview of the properties of coatings of different metals. For the metals of this group, the dissolution of the passive layer begins at about 20 V. For the development of an oxide layer, oxygen is necessary; during electrochemical processing this must be bound in a sufficient amount to the electrolyte. This is the case only for basic solutions of neutral salts. Development of passive layers can be counteracted by lowering the pH value (Enke 1982, McGeough 1974). Generally, passive layers are more stable in alkaline solutions. From the Pourbaix diagrams, the corrosion resistance of a material for specific combinations of pH value and potential can be determined (Bargel and Schulze 1978). Furthermore, such diagrams provide information on whether a material can be passivated. Pourbaix diagrams have been drawn for many elements. However, they describe only reactions between the metal and its oxides, its hydroxides, and water. For alloys, Pourbaix diagrams can be drawn only occasionally. The passivation behavior of single material–electrolyte systems can furthermore be influenced by temperature and concentration of the electrolyte. When using passivating electrolyte systems, for example, sodium nitrate, the stability of the coatings increases with temperature and the decreasing concentration of the electrolyte (Enke 1982, Pahl 1969).

1.13.4.1.4 Gap

From eqn [10] and according to Ohm's law, we get the following equation:

$$I = U\chi\frac{A}{s} \qquad [17]$$

By rearranging and substituting the current density j, the following equation is obtained for the gap width:

$$s = U\frac{\chi}{j} \qquad [18]$$

The gap width can be decreased by decreasing the specific electric conductivity χ of the electrolyte solution, by decreasing the operating voltage U, and by increasing the current density j, resulting in an increase in the achievable reproduction accuracy.

1.13.4.1.5 Electrolyte solutions

During electrochemical processing the following processes should be possible in electrolyte solutions (Degner 1984, König and Klocke 1997):

- Conduction of electric current (transportation of charge carriers)
- Anodic metal dissolution
- Removal of particles and reaction products
- Conduction of the process heat

Furthermore, for selecting electrolytes the following are next in importance to process-relevant properties:

- Physiological harmlessness
- Ecological neutrality (wastewater)
- Corrosive effect on the processing installation and its environment

- Economical requests (price, availability)
- Chemical stability
- Universal applicability

Aqueous solutions of neutral salts such as sodium chloride, sodium chlorate, sodium nitrate, potassium nitrate are mainly used, partly by adding complexing agents such as citric acid, in order to keep the ablation products in a dissolved state. For special applications (e.g., electrochemical drilling), sulfuric acid and caustic soda are used. Furthermore, a number of electrolyte mixtures are used, partly under organic additions, in order to reduce corrosion or to improve surface quality. For the machining of multiphase materials that show a different solution behavior, electrolyte mixtures are used. Applications of such mixtures have been described by Enke (1982) for nickel base alloys and by DeBarr (1968) for tungsten carbide. The characteristics of the electrolyte that must be taken into account are specific electric conductivity, concentration, temperature, pressure, and pH value. In literature, electrolytes are categorized as passivating and nonpassivating electrolytes. However, the distinction between passivating (e.g., sodium nitrate) and nonpassivating (sodium chloride) electrolytes is only valid for iron and nickel, since other metals such as aluminum, copper, and titanium show similar polarization behavior in both electrolyte types (Datta 1993, Walsch 1979). For most anode materials, the choice of the electrolyte determines whether the dissolution of the metal takes place in the active or in the passive region (Walsch 1979). For dissolution of passive layers on ferrous metals, halogen ions can be used since they have a corroding effect (Schwabe 1966). Here, the dissolution of the passive layers occurs above a critical potential by the entry of anions in the passive layer. In most cases, adsorbed water dipoles or sorbed oxygen are exchanged against the anions in question (Schwabe 1966). These processes mostly occur where the crystal structure is disordered, as, for example, at grain boundaries and dislocations (Hoar 1968). The ions with the smallest radius penetrate the passive layer fastest. Hoar arranged these ions according to their effectiveness: Cl^-, Br^-, I^-, CN^-, NO_3^-, and ClO_4^-.

The halogen anions Cl^-, Br^-, and I^- are absorbed by the surface atoms of the metal at a much more negative potential than the oxidation potential of the metal lets one expect. The passivation therefore does not occur when halogen-containing electrolytes are used (Evans 1965). Highest ablation rates are achieved by the use of Cl^- ions due to their high chemical aggressivity when penetrating passive layers. DeBarr (1968) considers sodium chloride as a common and widespread electrolyte for electrochemical processing of steels and nickel base alloys because of its relatively low acquisition costs. Chloride-containing electrolytes lead, however, to pitting corrosion and development of grooves and dimples at the surface. These defects have a negative impact on the mechanical and corrosion properties of the components (Effertz 1977, La Boda *et al.* 1971, Powers and Wilfore 1970). Effertz observed that pitting corrosion occurs at very small sodium chloride concentrations on heat-treated steel X20Cr13. For the processing with sodium chloride, Mao (1971) found a current efficiency of 100% when the solution is saturated with $FeOH_2$. In industrial practice, sodium chloride-containing electrolyte is used solely for the processing of titanium.

During processing with sodium nitrate, metal ablation occurs only at higher current densities. When the minimal necessary current density $\mathcal{J}_{min}$ is reached, a passive layer can dissolve locally. When sodium nitrate is used as an electrolyte, oxygen is produced at the anode, small or insignificant hydrogen is produced at the cathode, and the current efficiency is poorer (Mao 1971). Below a current density of 12 A cm^{-2}, almost the entire current is used for producing oxygen at the anode. According to Mao (1971) the current efficiency increases with increase in the concentration of the electrolytes and the current density. This observation is confirmed by Yu (1981). Mixed electrolytes are described by Enke and DeBarr for materials that show an opposite solution behavior for sodium chloride and nitrate solutions, as, for example, for nickel base alloys and tungsten carbide (DeBarr 1968, Enke 1982). A high number of experiments were conducted by Thornton (1992) for mixed electrolyte titanium processing. Bemelmans (1970) carried out basic investigations on titanium using different electrolyte mixtures that contained sodium chloride and/or sodium chlorate. Experiments with chloride/nitrate combinations did not increase the ablation rate when compared to processing with pure sodium nitrate (Mao 1973). Other authors too have provided information on different complexing agents. Typical complexing agents are oxalate, tartrate, citrate, and ethylenediaminetetraacetate anions (Bemelmans 1970). Complexing agents prevent precipitation of metallic salts and therefore the development of passive layers in the transpassive region. Since the solubility of these materials is limited, dissolved metals will be deposited onto the cathode in the case of saturation. Such electrolyte mixtures must be monitored continuously and

replenished frequently (Bannard 1997, Kubeth 1965, McGeough 1974). Yu (1981) tested different electrolyte combinations for the ECM of the titanium alloy TC4, for which he could not achieve any increase in the ablation rate.

1.13.4.1.6 Characteristics of electrolyte solutions

For describing the properties of electrolytes, a number of parameters, which are explained in detail in the following sections, are used.

1.13.4.1.7 Specific electric conductivity of the electrolyte solution

As mentioned above, the main task of the electrolyte is conduction of charge and/or ions. The ability to fulfill this task is expressed through the specific electric conductivity χ of the electrolytes. The specific electric conductivity is the reciprocal of the Ohm's resistance:

$$\chi = \frac{1}{R} \qquad [19]$$

For achieving high ablation rates, the conductivity of the electrolyte must be high. The specific electric conductivity of an electrolyte solution depends on its concentration and its temperature. According to Bemelmans (1970), the typical specific electric conductivity for electrolyte solutions is between 0.05 and 0.3 $S\,cm^{-1}$. Warming of the electrolyte can cause gas bubbles to develop, resulting in a decrease in the electrical conductivity. Increased contamination by reaction products and a resulting increase in viscosity cause a minor decrease in the electrical conductivity (Bemelmans 1970, Kubeth 1965, McGeough 1974). When the electrical conductivity decreases, the working gap decreases as well, and the accuracy of the process increases. That is why nitrogen is added to the electrolyte (Degenhardt 1972). The gas concentration and the conductivity are inversely proportional (Heitmann 1966, Hopenfeld and Cole 1966). However, this admixture might also lead to short circuits and nonuniform dissolution of the working materials (Landolt 1970).

1.13.4.1.8 pH value of the electrolyte solution

The pH value is a measure of the H^+ ion concentration in a solution and is calculated according to the following equation:

$$\mathrm{pH} = -\log_{10} C_{\mathrm{H}} \qquad [20]$$

An electrolyte with a pH value of 1 therefore has a high H^+ ion concentration and is strongly acidic. A strong base has a pH value of 14 at a small H^+ ion concentration. For ECM processes, neutral salts with pH values of around 7 are mainly used. When compared to strongly acidic or strongly basic electrolytes, the disadvantage of neutral electrolytes is that they have poor conductivity (DeBarr 1968, Kleiner 1963). The formation of hydrogen at the cathode has negative effects on the machining process. On the one hand, the formation of hydrogen through H^+ ion revocation reduces conductivity of the electrolyte solution; on the other hand, an increasing pH value promotes unwanted deposition of metal on the cathode surface (DeBarr 1968). According to Faraday's law, a dynamic balance is established during the machining process. Eventually, the pH value levels out in a range where the insoluble metal hydroxides precipitate and can be filtered out in a relatively easy manner (Walsch 1979).

1.13.4.1.9 Concentration of the electrolyte solution

The specification of the concentration of an electrolyte solution is mostly given in mass percent and is calculated using the following equation:

$$C_{\mathrm{Ely}} = \frac{m_{\mathrm{salt}} \times 100\%}{m_{\mathrm{Salz}} + m_{\mathrm{Wasser}}} \qquad [21]$$

An increase in electrolyte concentration during electrochemical processing generally results in an increase in the number of charge carriers within the electrolyte solutions, and therefore in a higher specific electric conductivity (McGeough 1974). Pahl and Degenhardt found a concentration dependency of the ablation rate only by usage of sodium nitrate when processing 41Cr4 and 56NiCrMoV7, but not for the usage of sodium chloride. For C45, Kubeth showed an increase in the ablation rate when sodium chloride was used due to the increase in the specific electric conductivity (Degenhardt 1972, Kubeth 1965, Pahl 1969). Klocke's viewpoint, however, contradicts that of Pahl and Degenhardt. He machined 56NiCrMoV7 with sodium chloride and determined the concentration dependence for this process (Klocke and Sparrer 1998).

1.13.4.1.10 Temperature of the electrolyte solution

Ions are mass-bound corpuscles that move under the influence of the working voltage in an electrolyte solution. Because of frictional heat, the temperature

of the solution increases (DeBarr 1968, Heitmann 1966). The conductivity of ionic conductors increases because of the increase in temperature since the mobility of the ions increases. On the other hand, gas bubbles form because of warming and subsequent overheating of the electrolyte. The occurrence of gas bubbles, in turn, leads to a decrease in the electrical conductivity (Bemelmans 1970, Kubeth 1965, McGeough 1974). Heitmann showed that with increasing temperature the increase in conductivity outweighs the decrease in conductivity (Heitmann 1966, Landolt 1970). However, a further temperature rise can lead to the boiling of water-soluble electrolytes (Heitmann 1966, Hopenfeld and Cole 1966). In this case, water vaporizes and the electrolyte crystallizes, which results in a short circuit between the anode and the cathode (Heitmann 1966, Pahl 1969). The large quantity of heat produced at high current densities must be removed from the processing volume (DeBarr 1968). Next to good thermal conductivity, a high specific heat capacity of the electrolyte is mandatory. A strong warming of the tool electrode can lead to deformations and therefore a decrease in the mapping accuracy.

1.13.4.1.11 Pressure of the electrolyte solution

For a higher pressure, the volume flow of the electrolyte and therefore the electrical conductivity increases. The temperature of the electrolyte solution decreases, as a larger amount of heat can be removed. Kawafune (1967) showed that for the processing of unalloyed steel using a 20% sodium chloride electrolyte solution, the current density increases with increasing electrolyte velocity. This statement was confirmed by Neubauer (1984). However, for the material X15CrMo13 and when using a sodium chloride solution, Kubeth (1965) found only a minor dependence of the specific ablation volume on the flow velocity. Clark reduced the flow velocity of sodium chloride electrolytes and detected a temperature rise up to the boiling point (Clark and McGeough 1977).

1.13.4.2 Materials

1.13.4.2.1 Irons, iron alloys, and steels

The idea for the ECM process was based on the use of a homogeneous working material consisting of a single metal. From that emanates a uniform metal dissolution in the feed direction. However, industrially used materials consist of several elements with different potentials. If several elements are present in a working material, the element that is less noble dissolves more quickly than the more noble one does. For example, iron (standard potential of –0.44 V) dissolves faster than nickel (standard potential of –0.25 V) (Lindenlauf 1977, Neubauer 1984). Neighboring structural domains with differing electrochemical properties lead to inhomogeneous current density distributions and therefore to locally differing ablation rates.

If the current density is increased further, potential differences that arise from structural inhomogeneities, together with the equidirectionally increasing electrode potential, loose their influence on the current density distribution, and therefore on the ablation behavior of the different material regions (Lindenlauf 1977, Neubauer 1984). Concentration of the current paths occurs in the region of the less noble structure phase. These phases dissolve faster than the more noble structure phases (Wilson 1971). Lindenlauf (1977) proved that the ferrite phase (Fe + C) dissolves faster than does the perlite phase (Fe + Fe_3C). When coarsely grained perlite phase exists in a ferrite matrix, whole grains can be rinsed out and eroded. These processes can then result in current efficiencies of more than 100%.

Neubauer (1984) observed that during the processing with sodium nitrate there was an increase in the specific ablation volume with increasing ferrite content in the structure. Fukunaga (1981) and McGeough (1974) obtained the same results. It is assumed that by heat treatment, the structure phases are distributed more uniformly. Pramanik examined the influence of an initial heat treatment on the ablation rate and the surface quality. He found that annealing after hardening increases the ablation rate; in the same way, the ablation rate should increase with increasing grain size (Pramanik 1975). Kops obtained different results from his investigations. He observed an increase in the current density and the metal ablation rate for reduced grain size. This observation can be explained by viewing the grain boundaries as surface defects. The atoms at the grain boundaries dissolve more easily from the surrounding material structure due to the different orientation of the bordering grains. The atoms of a grain, therefore, represent areas of increased chemical activity. The more the grain boundaries that are available at the surface, the better are the conditions for dissolution. By reduction of the grain size, the absolute grain border length and also the material ablation increases (Förster 2004, Kops 1976).

1.13.4.2.2 Titanium and titanium alloys

Titanium has a very high affinity for oxygen. Because of its position in the electrochemical series, it is a very basic metal. In oxidizing environments, oxide layers develop, which stick very firmly to the titanium surface and are chemically very inert. These oxide layers are the reason for the excellent corrosion resistance of titanium and titanium alloys. The properties of the oxide layers complicate the electrochemical workability of titanium. In order to carry out the processing, the oxide layer must be removed. This is often done using chemically aggressive electrolytes, which results in frequent surface defects (Thornton 1992). Dimples form in the immediate vicinity of the electrochemically ablated faces (Powers and Wilfore 1970). For high current densities, these dimples are so close to each other that they are not always recognizable as such. Bemelmans (1970) carried out basic investigations on titanium with different electrolyte mixtures that contained sodium chloride and/or sodium chlorate. Degenhardt (1972) concluded that for processing of titanium, there is no difference between a sodium nitrate and a sodium chloride electrolyte. This is contrary to the investigations of Bemelmans (1970) and many others. Thornton examined sodium bromide solutions for the electrochemical processing of titanium. He also used electrolyte mixtures of sodium fluoride, sodium bromide, and sodium nitrate (Thornton 1992). There are doubts regarding the toxic properties of these electrolytes; these electrolytes are hard to handle in industrial practice since they must be monitored and replaced where appropriate. Sauschkin *et al.* (1995) suggests the use of watery organic electrolytes for electrochemical processing of titanium, in order to avoid hydrogen saturation of the titanium surface and to avoid the resulting deterioration in the mechanical properties of the material. The disadvantages of these electrolyte solutions are the high costs, the poor conductivity, a low boiling point, and the danger of explosion of these solutions.

1.13.4.2.3 Hard metals

Hard metals represent a challenge for electrochemical processing as they consist of very different structural components that, electrochemically, behave differently; for example, by processing tungsten carbide, insoluble tungsten trioxide is produced, which prevents metallic dissolution (Bemelmans 1970, Kunert and Schwieger 1977). By short-term conversion of the polarity, where the workpiece is poled as cathode, coatings can be dissolved specifically. The hydrogen that develops on the workpiece surface reduces the passive layer, thereby lowering the valency of oxides or hydroxides, and as such oxides and hydroxides becomes more soluble (Bemelmans 1970).

However, material ablation takes place on the now anodically poled tool electrode. This tool wear leads to a deterioration in the accuracy. It is dependent on the current density, the duration of the polarity reversion, the type of hard metal that is to be depassivated, the electrode material, and the type of the electrochemical reaction that takes place in the electrolyte at the anode (Kunert and Schwieger 1977).

1.13.4.3 Procedure Variants for Microsystem Engineering

1.13.4.3.1 Electrochemical microdrilling

An activity that is increasingly gaining attention in the field of microsystem engineering is the manufacturing of small boreholes with the highest possible aspect ratios. Next to applications in turbine construction (air-cooling ducts), small boreholes in high-temperature materials have a high potential particularly in the automotive industry (injection systems) (Adam 1998). Also, in tool and mold construction, there is a need for small drillings, which are used for the evacuation and ventilation of molds during injection molding. A spindle can be mounted on the z-axis of the electrochemical processing machine. Besides external rinsing, the rotation spindle enables rinsing through the electrode. The rotational speed of the spindle can vary from 0 to 20 000 rpm. However, the construction has not yielded satisfactory results. **Figure 20** shows some boreholes produced during the pretests. The large sloping angle and the relatively large feed opening radius are clearly visible. When compared to boreholes that are fabricated by traditional electrochemical drilling procedures, such as STEM drilling, electrostream triplet drilling (ESD), electrochemical precision drilling (ECF), and electrojet (EJ) drilling, the boreholes fabricated by electrochemical microdrilling are still of limited application. There is an urgent need for further investigations, especially with regard to the effects of the rotational speed, the circumferential speed of the tool, and rinsing.

Another possibility, in particular for the parallel fabrication of a great number of drillings, consists in the sinking of thin pen or needle arrays. Another interesting application of the electrochemical sinking

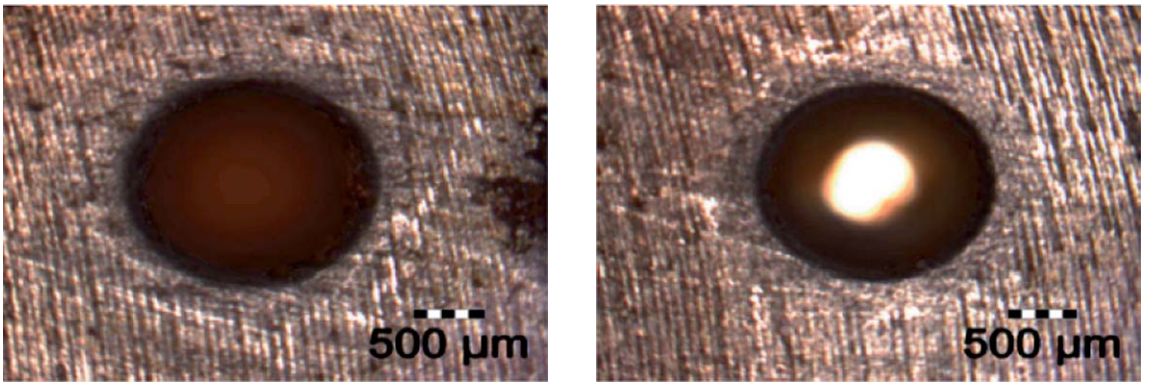


Figure 20 Holes produced by electrochemical microdrilling.

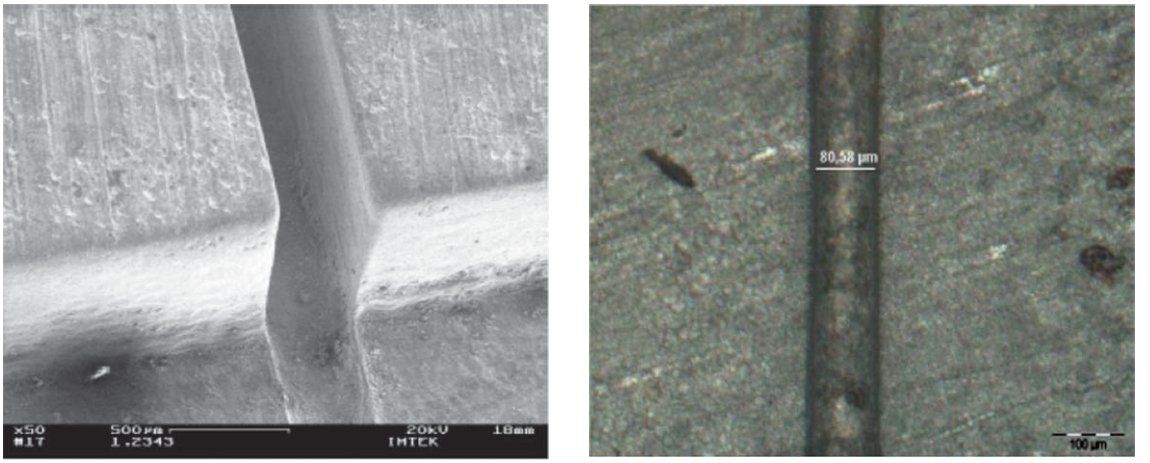


Figure 21 Structures made by electrochemical (EC) wire cutting.

process that uses an oscillating tool electrode is in the production of tools for micropunching with very filigree structures. Advantages result particularly because of the absence of process forces and because of the possibility to structure materials independently of their mechanical properties.

1.13.4.3.2 Electrochemical microwire cutting

The height of structures that can be manufactured with microwire EDM cutting is still limited. To overcome these limits, ECM was used to produce such structures.

Figure 21 shows such structures. The SEM picture clearly shows the very good sidewall angle of the structures produced. The smallest channels that have been manufactured with this process variant have a width of about 80 µm. These structures feature a length of about 15 mm and were manufactured with a 20-µm diameter tungsten wire.

1.13.4.3.3 Electrochemical micromilling

For the fabrication of micropatterns by mapping a negative of the structure to be produced, the fabrication of the negative by using electrochemical sinking of the tool electrode is also a technical challenge. One way of minimizing the relatively high manufacturing costs in the production of tool electrodes is by using a simple rotationally symmetric tool electrode. This tool electrode is fed along the desired shape of the workpiece, as in a traditional milling process. For this application, several controlled axes are necessary in the equipment.

References

Adam P 1998 *Fertigungsverfahren von Turboflugwerken*. Birkhäuser Verlag, Basel, Boston, Berlin

Allen D, Lecheheb A 1996 Micro electro-discharge machining of ink jet nozzles: Optimum selection of material and machining parameters. *J. Mater. Process. Technol.* **58**, 53–66

Allen D, Almond H J A, Bhogal J S, Green A E, Logan P M, Huang X X 1999 Typical metrology of micro-hole arrays made in stainless steel foils by two-stage micro-EDM. *Ann. CIRP* **48**, 27–30

Bannard J 1977 Electrochemical machining. *J. Appl. Electrochem.* **7**, 1–29

Bargel H-J, Schulze G 1978 *Werkstoffkunde*. Schroedel-Verlag, Hannover

Becker H-J, Fuchs K-D, Heberling E 1989 Martensitaushärtbare Werkzeugstähle. *Thyssen Edelst. Techn. Ber.* **15(2)**

Behmer U 1988 Untersuchung kinematischer Einflüsse beim funkenerosiven Senken mit überlagerter Planetärbewegung. Dissertation, TH, Aachen

Bemelmans N J 1970 Untersuchung zur elektrochemischen Bearbeitung von Metallen und Metallcarbiden durch anodische Auösung bei hohen Stromdichten (Elektrochemisches Senken). Dissertation, TH, Aachen

Brinksmeier E, Gläbe R 2002 Precision machining of steel with ultrasonically driven chilled diamond tools. *Proc. 17th Annual ASPE Meeting*, St. Louis, MO, USA

Brinksmeier E, Dong J, Gläbe R 2004 Diamond turning of steel molds for optical applications. *Proc. Euspen,* Glasgow, UK, p. 155

Bronstein I N, Semendjajew K A, Musiol G 1999 *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Frankfurt am Main, Thun

Christmann H 1999 Mikroschneidenerosion in der Anwendung. *Fachtagung: Funkenerosion- Zukunftstechnologieim Werkzeug- und Formenbau*. ADI-TEC gGmbH, Aachen (Hrsg.)

Clark W G, McGeough J A 1977 Temperature distribution along the gap in the electrochemical machining. *J. Appl. Electrochem.* **7**, 277–86

Datta M 1993 Anodic dissolution of metals at high rates. *IBM J. Res. Dev.* **37**, 207

Datta M, Landolt D 1980 Influence of anodic reactions on electrochemical machining performance. *Proc. Int. Symp. Industrial and Oriented Basic E*, Vol. 2, pp. 4.3.1–11

Dauw D F 1985 On-line identification and optimization of electrodischarge machining. Dissertation, KU, Leuven

Dauw D F, Coppennolle B 1995 On the evolution of EDM research, part 2: From fundamental research to applied research. *Proc. ISEM XI,* Lausanne, Switzerland, pp. 133–42

DeBarr A E 1968 *Electrochemical Machining*. MacDonald, London

Degenhardt H 1972 Elektrochemische Senkbarkeit metallischer Werkstoffe. Dissertation, RWTH, Aachen

Degner W 1984 *Elektrochemische Metallbearbeitung*. VEB Verlag Technik, Berlin

Degner W, Boettger H-C 1979 *Handbuch der Feinbearbeitung*. VEB Verlag der Technik, Berlin

van Dijck F 1973 Physico-mathematical analysis of the EDM process. Dissertation, KU, Leuven

Effertz P 1977 Die Lochkorrosionsanfälligkeit des Vergütungsstahles X20Cr13 in NaCl-Lösungen. *Werkstoe und Korrosion* **28**, 809–16

Enke C G 1982 Elektrochemische Metallbearbeitung. *Technica* **19**, 1645–49

Eubank P T, Eubank P T, Patel M R, Barrufet M A, Bozhurt B 1993 Theoretical models of the electrical discharge machining process. III. The variable mass cylindrical plasma model. *J. Appl. Phys.* **73(11)**, 7900–9

Evans U R 1965 *Einführung in die Korrosion der Metalle*. Verlag Chemie, Weinheim/Bergstr

Finkenburg W, Maecker H 1956 *Elektische Bögen und thermische Plasma*, *Handbuch der Physik*. Springer-Verlag, Berlin, Göttingen, Heidelberg, Vol. 22

Fleischer J, Schmidt J, Knoll M, Haupt S, Gehringer A, Förster R 2004 Mikrobearbeitung durch Abtragen. wt-online, December 2004

Forker W 1966 *Elektrochemische Kinetik*. Akademie Verlag, Berlin

Forker L 1968 Über einige verfahrenstechnische Grundlagen der EC Metallverarbeitung. *Die Technik* **23(1)**, 1924

Förster K 1979 Untersuchung der technologischen und physikalischen Zusammenhänge beim funkenerosiven Drahtschneiden. Dissertation, TU, München

Förster R, Schoth A, Menz W 2004 Micromachining of steel using ECM with vibrating tool-electrode. *Proc. Euspen International Conference*, Glasgow, UK

Fukunaga M 1981 Effects of carbon content and microstructures on the current efficiency in case of ECM of Fe-C alloys. *Ann. CIRP* **30(1)**, 117–21

Green L F 1987 High voltage pulsed EDM with special reference to silicon carbide. M.Sc. thesis, UMSIT, UK

Grimm U, Müller C, Menz W, Wölfle M 2004 Fabrication of surfaces in optical quality on pretentious tool steels by ultra precision machining. *Proc. Euspen,* Glasgow, UK, p. 193

Gruber H P 1999 Elektro-Erosion in der Mikrobearbeitung. *Fachtagung: Funkenerosion – Zukunftstechnologie im Werkzeug- und Formenbau*. ADITEC gGmbH, Aachen (Hrsg.)

Gusev V N 1930 *Br. Pat.* 335 003

Hamann C H, Vielstich W 1998 *Elektrochemie*. Wiley-VCH, Weinheim

Heitmann H W 1966 Werkzeugkorrektur und Anwendungsbereiche der Elektrolyte beim EC Senken. Dissertation, RWTH, Aachen

Hensgen G 1984 Werkzeugspezische Einüsse beim funkenerosiven Schneiden mit Ablaufender Drahtelektrode. Dissertation, TH, Aachen

Hensler K E, Weil K G, Bonhoeffer K F 1958 Die Bedeutung des Flade-Potentials für die Passivität des Eisens in alkalischen Lösungen. *Z. Elektrochem.* **15**, 149–61

Hoar T P 1967 The production and breakdown of the passivity of metals. *Corros. Sci*. **(7)**, 341–55

Hoar T P 1968 Passivierende Schichten. *Oberfläche* 9**(1)**, 1824

Hopenfeld J, Cole R R 1966 Electrochemical machining – Prediction and correlation of process variables. *J. Eng. Ind.* **88**, 455–61

Jutzler W-I 1982 Funkenerosives Senken- Verfahrenseinüsse auf die Oberflächenbeschaenheit und die Festigkeit des Werkstücks. Dissertation, RWTH, Aachen

Kawafune K 1967 Accuracy in cavity sinking by ECM. *Ann. CIRP* **15**, 443–55

Kips P 1960 Die funkenerosive Metallbearbeitung mit rotierender Werkzeugelektrode. Dissertation, TH, Aachen

Kleiner W B 1963 Which cutting fluid for ECM? *Metalworking Prod.* **8**, 61–4

Klocke F, Sparrer M 1998 EC finishing-the faster way to finished dies and moulds. *Int. J. Manuf. Sci. Prod.* **(4)**, 247–56

König W 1990 *Fertigungsverfahren, Vol. 3: Abtragen. Vol. 2*. VDI-Verlag, Düsseldorf

König W, Klocke F 1997 *Fertigungsverfahren Vol. 3 Abtragen und Generieren*. Springer-Verlag, Berlin, Heidelberg, New York

König W, Wassenhoven K 1991 Bahnerosion als Alternative – Mehrachsiges Erodieren mit einfachen Elektrodenformen. Industrie-Anzeiger **90(113)**, 22–6

Kops L 1976 Effect of pattern of grain boundary network on metal removal rate in ECM. *Ann. CIRP* **25(1)**, 125–30

Kracht E W 1970 Grundlagen der funkenerosiven Mehrkanalbearbeitung. Dissertation, TH, Aachen

Kubeth H 1965 Der Abbildungsvorgang zwischen Werkzeugelektrode und Werkstück beim Elektrochemischen Senken. Dissertation, RWTH, Aachen

Kunert G, Schwieger E 1977 Probleme und Möglichkeiten der Bearbeitung von Sinterhartmetallen. *Wissenschaftliche Zeitschrift der Technischen Hochschule Otto von Guericke Magdeburg* **21(1)**, 141–7

Kurr R 1972 Grundlagen zur selbsttätigen Optimierung des funkenerosiven Senkens. Dissertation, TH, Aachen
LaBoda M A, Hoare J P, Beacom S E 1971 The importance of the electrolyte in ECM. *Collection Czechoslov. Chem. Commun.* **36**, 680–8
Landolt D 1970 Crystallographic factors in high-rate anodic dissolution of copper. *Fundamentals of Electrochemical Machining*, The Electrochemical Society, Pennington, NJ, pp. 316–37
Lazarenko B R 1974 Die Elektrofunkenbearbeitung von Metallen. *Vestuik Maschinostroia*
Lazarenko B R, Lazarenko N I 1944 *Elektrische Erosion von Metallen*. Cosenergoidat, Moskau
Lindenlauf P 1977 Werksto- und elektrolytspezische Einüsse auf die elektrochemische Senkbarkeit ausgewählter Stähle und Nickellegierungen. Dissertation, RWTH, Aachen
Mao K-W 1971 ECM study in a closed-cell system. *J. Electrochem. Soc.* **118(11)**, 1870–9
Mao K-W 1973 The anodic dissolution of mild steel in solutions containing both Cl and NO_3 ions. *Corrosion Sci.* **13**, 799–803
Masuzawa T 1990 Micro electro-discharge machining and its applications. *Proc. IEEE MEMS,* Napa Valley, CA, USA, pp. 21–6
Masuzawa T 1998 *Recent Trends in EDM/ECM Technologies in Japan.* VDI-Berichte, Düsseldorf VDI-Verlag, Vol. 1405, pp. 1–15
McGeough J A 1974 *Principles of ECM*. Chapman and Hall, London
Mirnoff N 1965 Die Elektroerosion – Ihre physikalischen Grundlagen und industriellen Anwendungen. *Microtechnic* **19**
Mirnoff N 1968 *Einführung in das Studium der Elektroerosion (Physikalische Grundlagen und praktische Anwendungen)*. Microtec (Hrsg.), Lausanne
Neubauer J 1984 Untersuchung deckschichtbestimmender Reaktionsmechanismen und ihrer Auswirkungen auf die ECM. Dissertation, RTWH, Aachen
N.N 1974 DIN 8580; DIN 8590, Fertigungsverfahren Einteilung. *Forschungsbericht*, Beuth Verlag, Berlin
N.N 1975 *VDI-Richtlinie Elektroerosive Bearbeitung. Begrie, Verfahren, Anwendung*. VDI-3400, VDI-Verlag, Düsseldorf
N.N 1976 *VDI-Richtlinie Elektroerosive Bearbeitung. Denitionen und Terminologie*. VDI-3402 Blatt 1, VDI-Verlag, Düsseldorf
N.N 1985 CNC-Ausbildung für die industrielle Praxis, Teil 6: Erodieren/AGIE SA, Losone, Schweiz. Forschungsbericht. .Carl Hanser Verlag, München
Nöthe T 2001 Funkenerosive Mikrobearbeitung von Stahl und Hartmetall durch Schneiden mit dünnen Drähten. Dissertation, RWTH, Aachen
Pahl D 1969 Über die Abbildungsgenauigkeit beim elektrochemischen Senken. Dissertation, RWTH, Aachen
Powers R, Wilfore J 1970 *Some observations on the anodic dissolution of titanium at high current. Fundamentals of Electrochemical Machining*. The Electrochemical Society, Pennington, NJ, pp. 135–52
Pramanik D K 1975 Effect of heat treatment on ECM productivity. *Mach. Prod. Eng.* **126**, 224–6
Ritzhaupt-Kleissl E, Müller C, Sossalla A, Ertl S, Gluche P 2005 Fabrication of diamond micro tools for ultra precision machining. *Microsystemtechnologies.* Springer, Berlin, Vol. 11(4–5), pp. 278
Sato T, Mizutani T, Kawata K 1985 Electro-discharge machine for micro hole drilling. National Technical Report 31 Jg., pp. 725–33
Sauschkin B, Shiriyev V, Plarksin V, Atanasyants A 1995 Finish electrochemical machining of the large-scale punches for hot die forging. *Proc. ISEM*, Lausanne, Switzerland, Vol. 11, pp. 603–10
Schoepf M 2001 ECDM Abrichten metallgebundener Diamantschleifscheiben. Dissertation, ETH, Zürich
Schönbeck J 1992 Analyse des Drahterosionsprozesses. Dissertation, TU, Berlin
Schumacher B 1975 Bahngesteuertes, funkenerosives Schneiden mit Drahtelektroden
Schwabe K 1966 Über die Passivität der Metalle. *Angew. Chem.* **78(4)**, 253–66
Schwabe K 1974 *Physikalische Chemie*. Akademie Verlag, Berlin, Vols. 1 and 2
Schwabe K, Dietz G 1958 Zur Passivität des Nickels. *Z. Elektrochem.* **62(6/7)**, 751–9
Siebers F-J 1993 Funkenerosives Senken mit wässrigen Arbeitsmedien Grundlagen, Technologie und Wirtschaftlichkeit. Dissertation, RWTH, Aachen
Spur G, Stöferle T 1984 *Handbuch der Fertigungstechnik*. Carl-Hanser-Verlag, München, Vol. 4/1
Staelens F 1990 Overall on-line optimization of planetary electro discharge machining. Dissertation, KU, Leuven
Stanek J 1970 Perspektiven der Entwicklung der EDM- und ECM-Technik. *Proc. Int. Symp. Electrical Machining (ISEM)*. Wien, Austria
Stempel G 1977 Einsatz und Anforderungen an Elektroden für die verschiedenen Funkenerosionsverfahren. *Werkstatt Betrieb* **110**
Thornton R F 1992 Electrochemical machining of a titanium article. *Pat. No.* 5, 171–408
Tsuchiya H, Inoue T 1982 Generation and propagation of pressure wave by spark discharge in liquid. *Ann. CIRP* **31/1**, 107–10
Vetter K J 1961 *Elektrochemische Kinetik*. Springer Verlag, Berlin, Göttingen, Heidelberg
Walsch G 1979 Elektrochemische Metallbearbeitung – Die Spalt- und Oberflaechenausbildung beim elektrochemischen Senken von Staehlen mit Natriumnitratloesung. Dissertation, Universität Stuttgart
Weck M, Hennig J, Walter O 2003 Long stroke fast tool servo with dynamic mass compensation and hydrostatic bearings. *Proc. Euspen International Topical Conference,* Aachen, Germany, May, pp. 141–4
Weitzel W, Pompe R 1949 *Theorie elektrischer Lichtbögen und Funken*. Johann Ambrosius Barth Verlag, Leipzig
Wertheim R 1975 Untersuchungen der energetischen Vorgänge bei der funkenerosiven Bearbeitung als Grundlage für eine Verbesserung des Prozessablaufs. Dissertation, TH, Aachen
Wilson J F 1971 *Practice and Theory of Electrochemical Machining*. Wiley Interscience, New York
Wolf A 1997 Feine Sache. Die Mikrofunkenerosion ermöglicht beim Präzisionsspritzguss
Wolf A, Ehrfeld W, Gruber H P 1999 Mikrofunkenerosion für den Präzisionsformenbau. *wt Werkstatttechnik* **89(11/12)**, 499–502
Yu C K 1981 The relation between copying accuracy and electrolytes of electrochemical machining for titanium alloys. *Ann. CIRP* **30(1)**, 123–7
Zolotych B N 1955 *Physikalische Grundlagen der Elektrofunkenbearbeitung von Metallen*. SVT 175 VEB-Verlag Technik, Berlin
Zolotych B N 1957 *über die physikalischen Grundlagen der elektroerosiven Metallbearbeitung, Vol. 1 Elektroerosive Bearbeitung von Metallen*. Akademie der Wissenschaften der UdSSR, Moskau
Zolotych B N 1971 Theorie zum Phänomen der funkenerosiven Bearbeitung. *Fertigung* **2**, 185–91

Biographies

Holger Reinecke was born in 1964 in Bad Harzburg. From 1983 to 1988, he studied chemistry at the Technical University at Clausthal-Zellerfeld. From 1988 to 1990, he was scientific assistant at the Institute for Inorganic and Analytical Chemistry, and graduated in the field of electrochemical analytics in 1990.

In August 1990, Mr. Holger Reinecke started as a scientific assistant at the company microParts GmbH in the electroplating group, which he became head of in 1991. From 1993, he took over the department of chemical process technology, and the complete department of process technology in 1995. During this time, he developed, qualified, and established processes for the fabrication of microstructured components and tools. Among other things, lithographical, electrochemical, vacuum, or laser technical methods were implemented. Furthermore, he has established complete process chains for mass production of silicon-based medical products as well as for cleaning and surface coating of polymeric components. These components were used in medical devices. These processes were designed, installed, validated, and operated according to medico-technical and pharmaceutical requirements of European and US-American approval authorities. In 1999, as an area manager Mr. Reinecke additionally became head of the product branches microfluidics and microoptics. Since November 2004, Holger Reinecke is Head of the Chair of Process Technology at the Institute for Microsystem Technology (IMTEK) at the Freiburg Albert Ludwigs University. Since May 2005, Holger Reinecke is also Speaker of the Executive Board of the HSG-IMIT in Villingen-Schwenningen.

Claas Müller studied physics from 1986 to 1991 at the University of Karlsruhe. Following the physics diploma, he earned his doctorate in 1994 at the Forschungszentrum Karlsruhe, Institute for Microstructure Technology, for his work on a miniaturized spectrometer system, fabricated by LIGA technology. Meanwhile, the micro spectrometer is introduced to a broad range of industrial applications by the company microParts. At the Forschungszentrum, the prerequisites for a small-scale production were achieved. As a responsible project manager, Claas Müller was considerably involved in these activities. Since 1996, he is a graduate council at the Chair of Process Technology of the IMTEK. In 1999, he was appointed substitutional manager, and in 2004, the managing director of the Chair of Process Technology.

1.14 Self-Assembly

Jiandong Fang and Karl F Böhringer, University of Washington, Seattle, WA, USA

1.14.1 Introduction and Motivation

1.14.1.1 The Growing Need for Assembly

Microassembly is a technique for producing complex microstructures and systems by integrating microcomponents made in separate manufacturing processes. Two-dimensional microassembly mounts components on a flat substrate, and 3D microassembly produces structures with a more intricate spatial geometry.

Both 2D and 3D microassemblies are needed for mass production of microdevices. For example, a radio frequency identification (RFID) tag consists of a silicon microchip and an antenna. The antenna powers the tag and exchanges data between the chip and a neighboring RFID reader. The key to efficient

fabrication of RFID tags lies in the rapid replacement of the small, more expensive silicon microchips with the larger, cheaper antenna substrate (Silicon Chip Online 2006). Another example for a microassembly is smart dust, i.e., collections of tiny distributed nodes that integrate sensing capabilities with signal processing and can organize into a communications network (Warneke *et al.* 2001). These systems are emerging as commercial products, and microassembly techniques play a very important role in mass production of such microscale devices.

1.14.1.2 Monolithic Fabrication vs. Hybrid Integration

Monolithic fabrication constructs a whole device step by step from a single substrate. Hybrid integration constructs a device by combining various components from different techniques or processes. For devices with simple structures, monolithic fabrication has advantages such as compact layout and easy encapsulation. But many fabrication processes and materials are not compatible with one another, and thus microdevices having complex functionalities or structures must be constructed by hybrid integration of multiple units from several different fabrication processes.

Additionally, hybrid integration has another significant advantage for product developers. A complex microsystem can be broken down into multiple distinct units. Each unit is then assigned to a separate process engineer or group. Each unit can be separately optimized in design and fully characterized before integration. Units of the same type can be fabricated at the highest spatial density on a single substrate to save fabrication costs.

1.14.1.3 Packaging: Die Level vs. Wafer Level

Typically, packaging of microdevices involves three process steps:

(1) Placement of microdevices with correct face and in-plane orientations on receptor sites
(2) Permanent bonding of microdevices with electrical interconnections to the substrate
(3) Encapsulation of the bonded device components for protection from the environment

The choice of bonding method depends on the placement of microdevices: flip chip bonding is used for microdevices with electrical interconnects facing the substrate and wire bonding is used for microdevices with electrical interconnects facing the same direction as their counterparts on the substrate. A flip chip bonding process achieves electrical and mechanical connections simultaneously, and the electrical connections are established in parallel. A wire bonding process attaches a microdevice to the substrate to form mechanical connections before serial establishment of electrical connections. Die-level packaging assembles individual microcomponents after they are diced from a wafer. Wafer-level packaging simultaneously assembles all the devices on the same substrate before they are diced into individual devices, i.e., wafer-level packaging process is parallel assembly whereas die-level packaging is serial assembly.

1.14.2 State of the Art in Microassembly

1.14.2.1 Deterministic vs. Stochastic

Microassembly approaches, aimed at achieving highly efficient assembly of a very large number of microcomponents, can be classified into two major categories: deterministic and stochastic (**Figure 1**). In a deterministic assembly process, each part is assigned to a specific receptor site, e.g., as in wafer-to-wafer transfer of microstructures. In a stochastic assembly process, each part can be attached to any of the specifically designed identical receptor sites. The term self-assembly usually describes such stochastic assembly approaches, owing to the property that the assembly takes place in a spontaneous manner with components of a specific design. Self-assembly processes are discussed in detail in the remaining sections.

1.14.2.2 Serial vs. Parallel

The number of parts assembled at a time on a single platform defines the property of an assembly process: serial or parallel. In a serial assembly process, only one part is assembled at each time, for example, as in robotic pick-and-place assembly methods. In a parallel assembly process, multiple parts are assembled simultaneously, which can achieve a higher throughput. Both wafer-to-wafer transfer of microcomponents and self-assembly are parallel processes.

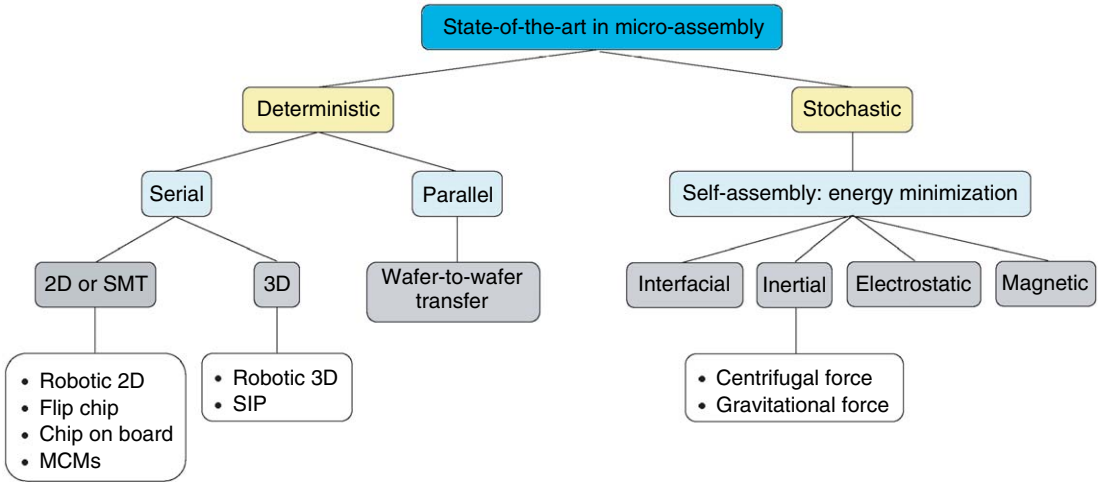


Figure 1 Categories of state-of-the-art microassembly techniques. MCM, multichip module; SIP, system in package.

1.14.2.3 Surface Mount vs. 3D Assembly

Surface mount technology (SMT) was fully established in the 1980s, when circuits became more and more complex so that through-hole component-mounting techniques were no longer economically or technologically feasible. The limitations in conventional printed circuit (PC) board technology motivated circuit design engineers to develop SMT. SMT keeps components and their interconnecting leads on one PC board surface, rather than feeding the component leads through the circuit board. SMT employs a solder to provide electrical and mechanical connections between components and PC boards. Compared with conventional through-hole mounting techniques, SMT possesses many prominent benefits such as reduced component size, increased circuit density, reduced PC board size, reduced weight, increased interconnecting leads density, and improved high-frequency performance. No single electronic assembly technology is perfect for satisfying all circuit design constraints. Current SMT also has some significant limitations such as poor heat dissipation, thermal mismatch, and decreased mechanical bonding strength of solder. SMT has a wide range of applications in packaging IC or microelectromechanical systems (MEMS) components, e.g., chip-on-board (COB) and multichip modules (MCMs).

Robotic assembly is a serial deterministic method widely used in industry to assemble a variety of components. This assembly approach has three major steps:

(1) Feed parts
(2) Pick and place parts
(3) Affix parts in specified position and orientation

During a typical part feeding step, randomly agitated components move through a series of mechanical filters so that finally they show appropriate face and in-plane orientations. Then robotic grippers pick up these components and transfer them to targeted sites to complete the assembly. This is a serial assembly process. Different part feeding mechanisms are required by various types (geometry or material fragility) of components. Zyvex Corporation has constructed small semiautomated robotic systems for microscopic and nanoscopic assemblies (Skidmore *et al.* 2000). Flat silicon microcomponents are a big challenge for part feeding since they are symmetrical except for negligible differences in geometric features on some surfaces, such as interconnecting pads. Depending on the degrees of freedom of the gripper, robotic assemblies can achieve 2D surface mount or 3D assembly.

A system in package (SIP) comprises multiple components integrated into a single package. In a typical SIP, components can be stacked vertically with 3D electrical interconnections; thus SIPs tend to be compact with much less parasitic capacitance and inductance.

1.14.3 Self-Assembly

Self-assembly techniques for component-to-substrate assembly are mainly based on energy minimization (**Figure 2**). A substrate is patterned with an array of energy traps. The types of energy traps include inertial, interfacial, electrostatic, and magnetic energies. On an assembly substrate, agitated

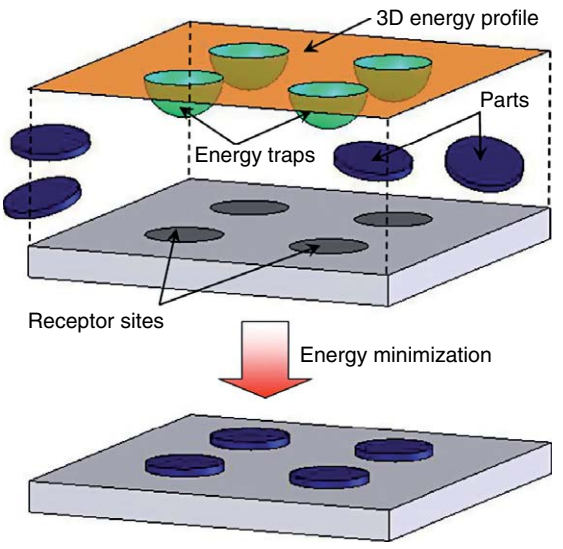


Figure 2 Schematic diagrams showing a typical self-assembly process based on energy minimization: on a substrate, patterned receptor sites are energy traps to attract the surrounding agitated microcomponents.

components are attracted to these energy traps and then permanently attached. Thus, components self-assemble to receptor trap sites on the substrate.

Compared with SMT and robotic pick-and-place assembly methods, self-assembly techniques have several major advantages such as easy handling of components with size in microdomains, fast parallel assembly, and self-alignment with high accuracy.

1.14.3.1 2D-Conformal-3D

A 2D conformal self-assembly process positions microcomponents on a rigid or flexible substrate, and the flat components attach directly to the substrate, i.e., assembled components are conformal with the substrate surface. This is the most common type of self-assembly, since patterning of energy traps on a substrate is straightforward with current microfabrication and surface modification techniques.

Three-dimensional self-assembly of microcomponents poses more challenges to currently available microfabrication techniques than does 2D self-assembly. Some simple 3D microstructures have been constructed by self-assembly processes based on shape matching and interfacial energy minimization (Gracias *et al.* 2000, Zheng *et al.* 2004), but this bottom-up fabrication method for complicated 3D microstructures still requires more breakthrough innovations in microfabrication techniques and assembly mechanisms.

1.14.3.2 Hard-Coded vs. Programmed

Depending on whether the probability for parts to attach to receptor sites can be adjusted by varying conditions, self-assembly processes can be categorized into hard-coded or programmed types. In a hard-coded assembly process, each receptor site has a constant probability to attach a part at any time. In a programmable assembly process, a receptor site has different affinities for a part under different conditions such as assembly environment and surface hydrophobicity, i.e., a receptor site can be turned on (an energy trap) or off (at the same energy level as its background or even higher). A typical usage of programmed self-assembly is to integrate different types of microcomponents on an array of receptor sites patterned by a single fabrication process.

1.14.3.3 Electrostatic and Electromagnetic Self-Assembly

1.14.3.3.1 Electrostatic 2D assembly

Electrostatic 2D assembly is based on the electrostatic attraction of microcomponents, and the electrostatic attraction is due to the polarization of the microcomponents in electrical fields. According to Coulomb's law, electrostatic force is inversely proportional to the square of the distance between charges, and thus it is a short-range interaction. Patterned surface areas with charges or localized electrical fields can be binding sites for microcomponents, and the binding strength increases with surface charge density or electrical field intensity.

Tien *et al.* (1997) exploited electrostatic interactions for direct patterning of 10-μm diameter gold disks on functionalized substrate surfaces. They fabricated the small charged gold disks by electroplating gold into photoresist molds and derivatizing these disks with charged self-assembled monolayers (SAMs), and patterned the planar or curved substrate with surface charges by microcontact printing or photolithography. By agitation, the charged gold disks aggregated on the regions with the opposite charges. Finally, they obtained selective and dense assembly in solutions such as methanol, ethanol, IPA, and dioxane (**Figure 3**).

Cohn *et al.* (1995) demonstrated a self-assembly process with alignment capabilities over electrostatic traps. The experiment was performed in an aqueous environment. The electrostatic traps were fabricated on a silicon substrate: a layer of 2-μm low-temperature oxide (LTO) was first deposited

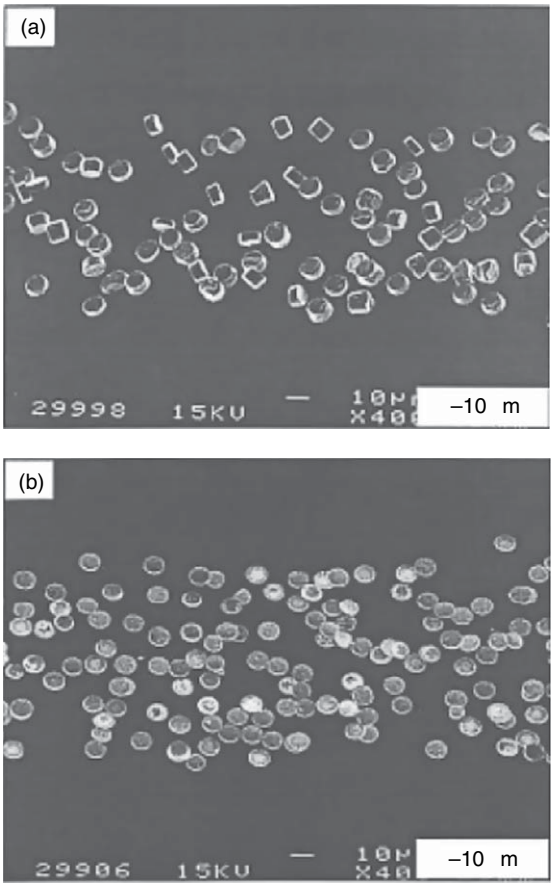


Figure 3 Scanning electron micrographs (SEMs) of PO_3H-terminated gold disks assembled on patterned gold. Gold wafers were patterned with HS-$(CH_2)_{15}$COOH and washed with 1 mM $HS(CH_2)_{11}NMe^{3+}$ and 1% HCl/EtOH. (a) Disks with an aspect ratio near 1:1 did not overlap each other. (b) Disks with an aspect ratio of 1:15 overlapped occasionally. (Photo courtesy: Tien J, Terfort A, Whitesides G M 1997 Microfabrication through electrostatic self-assembly. *Langmuir* **13**, 5349–55.)

as a dielectric layer by using low-pressure chemical vapor deposition (LPCVD), and then a layer of Au was deposited and patterned with an array of apertures (in the size ranges from 2 to 100 μm). The silicon substrate and the Au layer acted as two driving electrodes, which resulted in fringing electric fields from the patterned apertures in the Au layer. If the microcomponents stick to the substrate or to each other, there is no way to distribute or to assemble them to the specified binding sites or to the electrostatic traps on the substrate. They demonstrated a method to avoid such sticking phenomena by levitating the microcomponents a short distance (0–100 μm) above the target electrostatic traps before letting them settle, and thus accurate placement and orientation of the microcomponents with respect to the binding traps became feasible with a relatively low applied field of $\sim 10\,V\,\mu m^{-1}$. Each microcomponent (a SiO_2–*p*-Si–SiO_2 sandwich) had an average dielectric constant of 10, and the assembly liquid environment was hexane with a relatively low permittivity; therefore, microcomponents were easily attracted to the charged electrodes. The levitation of microcomponents was achieved by adding a small amount of a more polar solvent (acetone, with a relatively high dielectric constant of ~20), i.e., acetone shielded the electric field from the microcomponents and prevented contact of microcomponents with the binding sites. When all trap sites were occupied with floating microcomponents, the polar solvent was titrated out to lower the microcomponents to contact the trap sites. Finally, permanent bonding could be accomplished with sintering metals.

Fringing electric fields out of patterned apertures on a substrate can be exploited for assembling microparts in an air environment (Böhringer *et al.* 1998). The experimental setup is shown schematically in **Figure 4**. A glass substrate was coated with a layer of Cr/Au and the Cr/Au was patterned with an array of square holes to expose the glass underneath. The glass substrate was mounted on an aluminum vibratory platform driven by a piezoelectric actuator. Bulk parts were placed on the glass substrate. For effective agitation of these parts (initially parts tended to stick to the substrate and to each other because of the combination of electrostatic interaction and capillary and van der Waals forces), the vibrating frequency was chosen in the 20-kHz range. A DC voltage was applied on the aluminum platform and the Au electrode, and the resulting fringing electric field from the apertures in the upper electrode induced polarization in neutral parts and caused them to be attracted to the apertures. When a part covered an upper electrode aperture, the fringing electric field was significantly reduced, which prevented attraction of more parts to this occupied site. Assembly experiments were performed in both air and low vacuum, and the results indicated that the vibrating energy required to overcome adhesive forces decreased with air pressure, probably due to squeeze film effects (Fearing 1995) and due to the vacuum created between the flat-part bottom surface and the substrate when operated at ultrasonic frequencies.

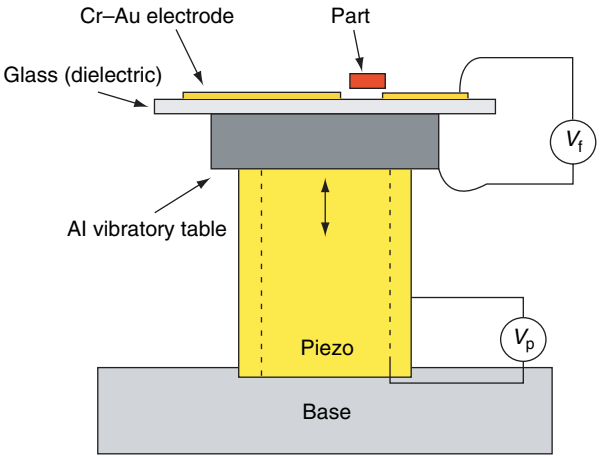


Figure 4 Schematic experimental setup for self-assembly with electrostatic traps: a vibratory table with a gold-covered dielectric is attached to a piezoelectric actuator. The aperture in the upper electrode creates a fringing field that causes polarization in the part. The part is attracted to the aperture. (Source: Böhringer, K F, Goldberg K, Cohn M, Howe R T, Pisano A 1998 Parallel microassembly with electrostatic force fields. *Proc. Int. Conf. Robotics and Automation (ICRA)*, Leuven, Belgium, pp. 1204–11.)

1.14.3.3.2 Toward 3D by triboelectricity

Kaajakari and Lal (2001) demonstrated a method based on triboelectricity for batch assembly of polysilicon hinged structures (**Figure 5**). Triboelectricity on hinged plates and the substrate was achieved by use of ultrasonic vibrations generated with an attached piezoelectric actuator to vibrate polysilicon plates on silicon nitride or polysilicon substrate surfaces. Such contact electrification charges resulted in hinged flaps being stabilized vertically on the substrate. Furthermore, they also observed a memory effect in which assembly occurred even without further ultrasonic vibration after the initial ultrasonic assembly.

1.14.3.3.3 Magnetic assembly

Magnetic force is a type of short-range interaction, and it can be utilized to attach microcomponents to magnetized sites on a substrate. Perkins *et al.* (2002) developed a new assembly approach called magnetically assisted statistical assembly (MASA): the compound semiconductor device heterostructures, 'nanopills' covered with a soft magnetic material, were trapped in the shallow recesses patterned into the surface of an integrated circuit wafer, and the short-range magnetic attractive forces kept the parts in the recesses (**Figure 6**). The 'nanopills' were agitated by liquid flow. **Figure 7** shows the

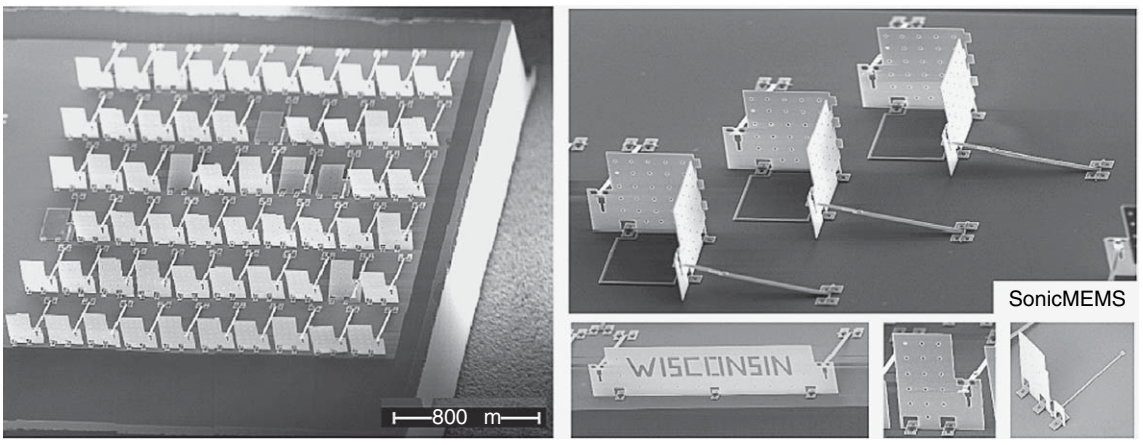


Figure 5 Examples of devices assembled with triboelectricity: array of assembled micromachined flaps, polysilicon corner cube reflectors (200 μm × 200 μm) with lock-in structure, and micro-art Wisconsin banner. Lower right shows a flap with a retaining spring latch anchored to the substrate. (Photo courtesy: Kaajakari V, Lal A 2001 Electrostatic batch assembly of surface MEMS using ultrasonic triboelectricity. *14th IEEE Int. Conf. Micro Electro Mechanical Systems*, pp. 10–13.)

relationship between the magnetic force per unit area and the separation between a part's permalloy surface and the magnetized Co–Pt stripes at the bottom of a recess.

Magnetic fields can also be exploited to construct 3D microstructures. Iwase *et al.* demonstrated a sequential batch assembly method (Iwase and Shimoyama 2005) based on the magnetic field-assisted self-assembly (**Figure 8**): an external magnetic field perpendicular to the substrate was used to lift the hinged ferromagnetic microstructures, and the magnetic field was increased gradually so that the plate with softer hinges was lifted first. The softness of a plate's hinge can be controlled by adjusting the length and the number of hinge beams.

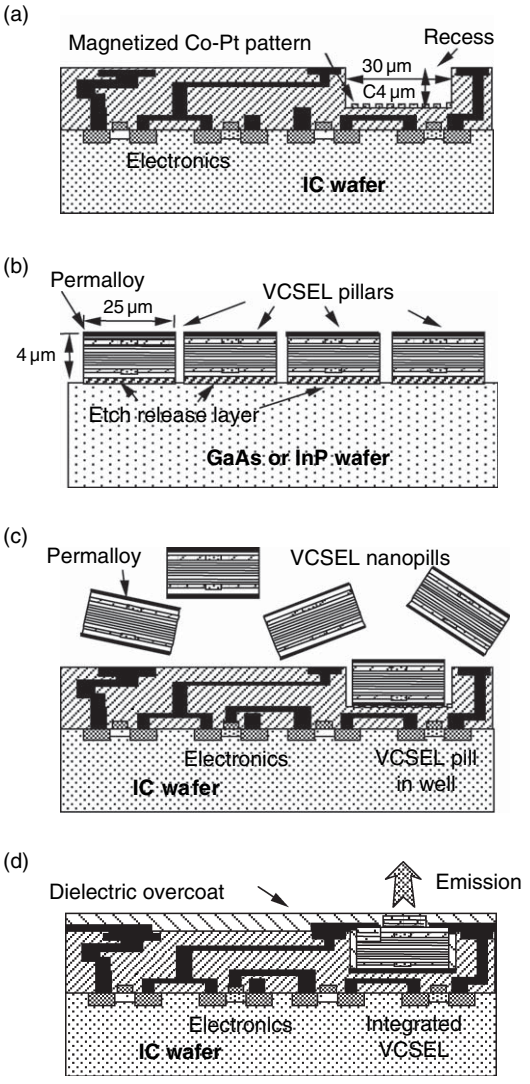


Figure 6 A schematic overview of the magnetically assisted statistical assembly (MASA) process: (a) the processed IC wafer with the recesses prepared; (b) the p-side down device wafer (in this case vertical cavity surface-emitting laser (VCSELs)) with pillars etched in a close-packed array; (c) stochastic assembly of the freed nanopills in the recesses on the IC wafer; (d) after completing device processing and integration. (Photo courtesy: Perkins J, Rumpler J, Fonstad C G 2002 Magnetically assisted self-assembly – A new heterogeneous integration technique. *MIT Microsystems Technology Laboratories Annual Report*.)

1.14.3.4 Shape Matching

1.14.3.4.1 Fluidic self-assembly

Fluidic self-assembly relies on shape matching between microcomponents and recessed receptor sites (Yeh and Smith 1994). Bulk microcomponents are agitated by liquid flow until they fall into receptor holes. GaAs vertical cavity surface-emitting laser (VCSEL) components were demonstrated to be integrated on a silicon substrate. The silicon host substrate was patterned with etched holes of trapezoidal shape, and the GaAs light-emitting diode (LED) components were also of the same trapezoidal shape. Carrier fluid containing the GaAs microcomponents was dispensed over the host silicon substrate. Because of the trapezoidal shape design, the GaAs blocks fit preferentially into the holes in the desired face orientation. Random mechanical vibration was performed to enable the precise positioning of a large number of microparts in the recessed receptor sites. More than 90% of the etched holes on the substrate were correctly filled with the GaAs blocks before the carrier fluid evaporated.

Alien Technology Corporation has commercialized this fluidic self-assembly technique (**Figure 9**). Fluidic self-assembly can work on either rigid (glass or plastic) or flexible (polyester, polyimide, polycarbonate, etc.) substrates. They have been using fully automated in-line reel-to-reel (web) processing on a continuous flexible substrate. Fluidic self-assembly and web processing offer many advantages over conventional processing on rigid substrates, such as those given below:

(1) Reduced manufacturing facility investment
(2) Reduced material and manufacturing costs
(3) Increased manufacturing throughput

Stauth *et al.* also used a fluidic self-assembly technique to integrate silicon microcomponents on a plastic substrate (Stauth and Parviz 2005). Receptor

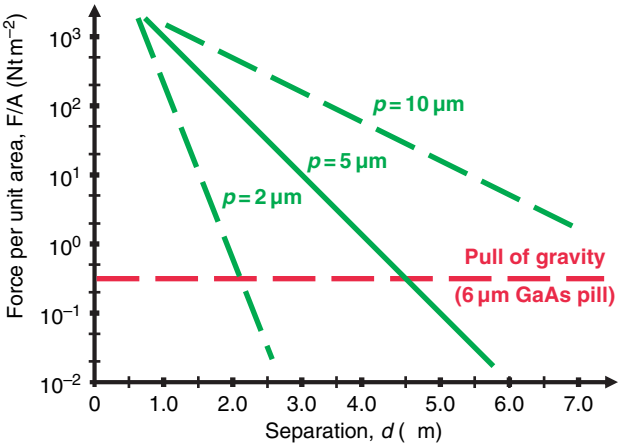


Figure 7 The attractive magnetic force per unit area as a function of separation for various spatial periods p of magnetic stripes at the bottom of the recesses. For comparison the pull of gravity on a 6-μm-thick GaAs nanopill is shown by the horizontal dashed line. (Photo courtesy: Perkins J, Rumpler J, Fonstad C G 2002 Magnetically assisted self-assembly – A new heterogeneous integration technique. *MIT Microsystems Technology Laboratories Annual Report*.)

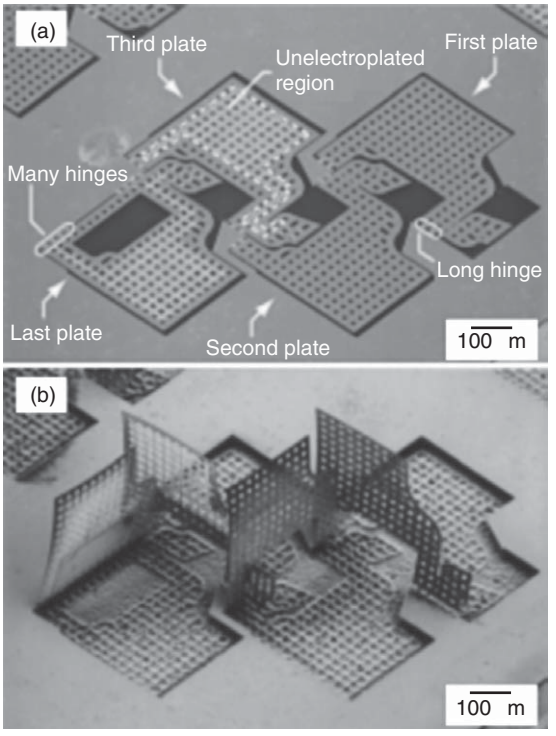


Figure 8 Scanning electron micrograph (SEM) of four-step sequential batch assembly driven by magnetic field: (a) before assembly; (b) after assembly. (Photo courtesy: Iwase E, Shimoyama I 2005 Multi-step sequential batch self-assembly of three-dimensional micro-structures using magnetic field. *Proc. IEEE Int. Conf. Micro Electromechanical Systems*, pp. 588–91.)

wells on the plastic substrate were formed with a lithographically patterned thick SU8. At the bottom of these receptor wells, electrical terminals were coated with low-temperature melting solder. Interconnect pads on silicon microcomponents are patterned Au areas with high solder-wetting capability. In a hot and acidic aqueous environment, microcomponents slid on the tilted substrate and fell into receptor wells, and then the microcomponents were anchored when their interconnect Au pads contacted melting solder droplets in the wells.

1.14.3.4.2 Dry self-assembly

Shape matching can also take place in a dry assembly environment such as air or vacuum. Dry assembly environments avoid potential physical damage to or chemical erosion of microcomponents. Fragile structures on a microcomponent, for example, cantilever beams, can be easily immobilized or even broken by surface tension forces of liquid residue from a wet assembly environment. Dry self-assembly techniques based on shape matching are discussed in Section 1.14.3.7.

1.14.3.5 Centrifugal Force-Driven Self-Assembly and toward 3D

For microscale components, inertial forces, for example, gravity, are usually neglected when compared with surface forces such as surface tension and electrostatic forces. As an inertial force, a centrifugal force can be increased to overcome surface forces because it is proportional to the square of the rotating speed.

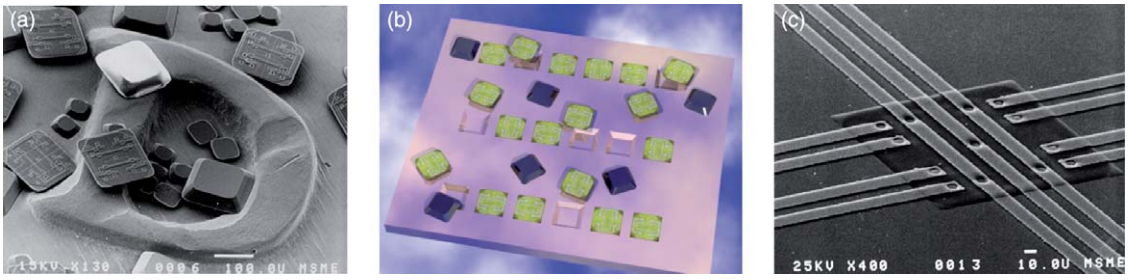


Figure 9 Fluidic self-assembly technique commercialized by Alien Technology: (a) microphotograph of individual nanoblocks on the surface of a dime; (b) artist's rendition of nanoblocks falling into substrate holes; (c) microphotograph of a passivated metallized nanoblock. (Photo courtesy: Alien Technology whitepaper published in 1999.)

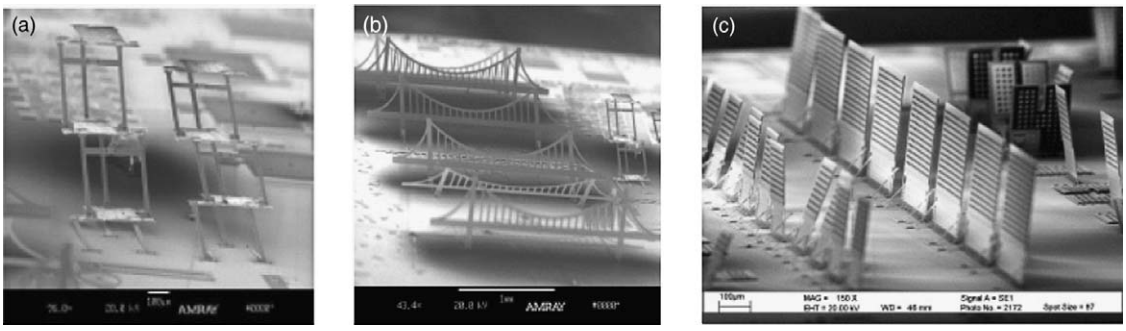


Figure 10 Microstructures assembled by centrifugal force: (a) tri-level poly-tower; (b) an array of Tsing-ma Bridge (a famous bridge in Hong Kong) structures; and (c) an array of microplates. (Photo courtesy: Lai K W C, Hui A P, Li W J 2002 Non-contact batch micro-assembly by centrifugal force. *Proc. IEEE Conf. Micro Electro Mechanical System*, pp. 184–7.)

Lai *et al.* (2002) demonstrated noncontact batch assembly of 3D microstructures by centrifugal force. Hinged microstructures fabricated in a multi-user MEMS process (MUMPs) were assembled with the help of latching features by centrifugal forces (**Figure 10**).

1.14.3.6 Capillary Self-Assembly

1.14.3.6.1 Toward 3D with solder/photoresist reflow

In microdomains, surface tension forces dominate over gravitational forces, and they have been explored and exploited by some researchers to lift hinged plates for constructing 3D structures (Syms *et al.* 2003). To achieve this type of 3D assembly, a thick layer of solder or photoresist is usually deposited and patterned at the joint of hinged structures. When the solder or photoresist is reflown, it changes into a spherical shape to minimize interfacial energies (a spherical shape has the minimum surface-to-volume ratio), and thus the hinged plate attached to the reflowed droplet rotates. With appropriate latching features, some complex 3D microstructures can be constructed (**Figure 11**).

Yang *et al.* (2006) developed a method to achieve local assembly by reflowing solder with magnetic induction welding. The heating temperature depends on the area of the magnetic film underneath the solder, and thus different heating temperature regions on a substrate can be photolithographically defined (the magnetic field is assumed to be uniform across the substrate). They demonstrated this technique by separating Ni plates from Ti plates (**Figure 12**).

Zheng *et al.* (2004) used a low-temperature melting solder to construct functional circuits that consisted of 3D structures including a LED component, a chip carrier with solder bumps, and an encapsulation body with solder bumps. The assembly sequence was as follows (**Figure 13**):

(1) LED components and chip carrier components were put in a hot and acidic liquid environment, where tumbling agitation caused each LED component to attach to a chip carrier component. Finally, the attached components with the solder bonding were sorted out.

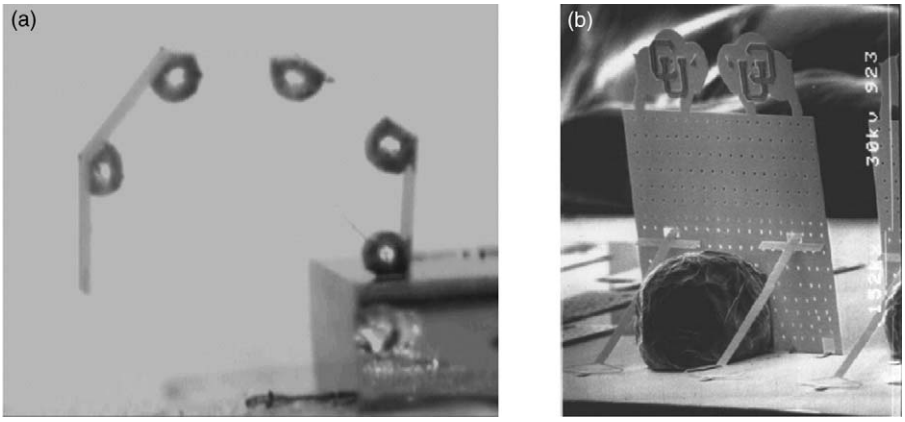


Figure 11 Optical graphs of some 3D self-assembled structures by reflowed solder: (a) 5-bar multiple link assembly. (Photo courtesy: Harsh K F, Kladitis P E, Zhang Y H, Dunn M L, Bright V M, Lee Y C 2000 Tolerance and precision study for solder self-assembled MEMS. *Proc. SPIE* **4075**, 173–84.) (b) A flap vertically assembled. (Photo courtesy: Kladitis P E, Bright V M 2000 Prototype microrobots for micro-positioning and micro-unmanned vehicles. *Sens. Actuators* **A80**, 132–7.)

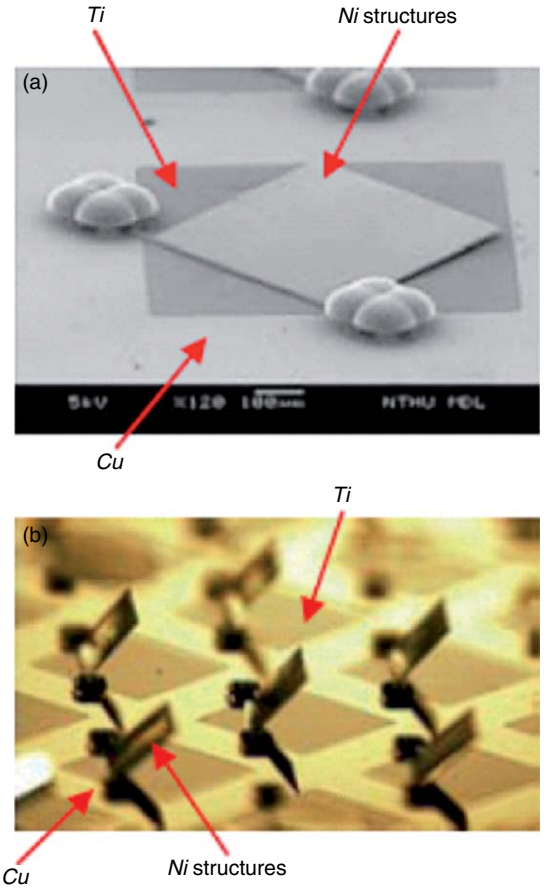


Figure 12 Optical micrographs of the tested structure to show separation of Ni and Ti films: (a) before and (b) after applying magnetic field. (Photo courtesy: Yang H, Lin C, Fang W 2006 Wafer level self-assembly of microstructures using global magnetic lifting and localized induction welding. *IOP J. Micromech. Microeng.* **16**, 27–32.)

(2) The attached components from the previous step were placed in a hot and acidic liquid environment together with encapsulation components for another batch assembly.

1.14.3.6.2 Parallel 2D assembly

In capillary-driven 2D self-assembly processes, agitated flat microcomponents contact and attach to adhesive liquid droplets on receptor sites, and then self-align to receptor sites to minimize interfacial energies. Depending on the assembly environment, capillary-driven self-assembly processes can be classified into two types: in an aqueous environment and in an air environment. Different adhesive liquids require different hydrophilicity or surface coatings for the receptor sites. For an acrylate-based adhesive liquid, receptor sites should be hydrophobic for aqueous assembly environments and hydrophilic for air assembly environments (**Figure 14**). For low-temperature melting solder, receptor sites should be thin film metals with good solder-wetting capability, e.g., copper and gold.

Low-temperature melting solder can be used to assemble flat microcomponents with both electrical and mechanical connections on a substrate. Jacobs *et al.* (2002) developed a self-assembly process to mount LED arrays on flexible cylindrical templates by using a low-temperature melting solder. The assembly template was patterned with copper squares, and a simple dip coating process left melting solder on these copper squares since copper has very good wetting capability for solder. The dip coating process was performed in an acidic aqueous

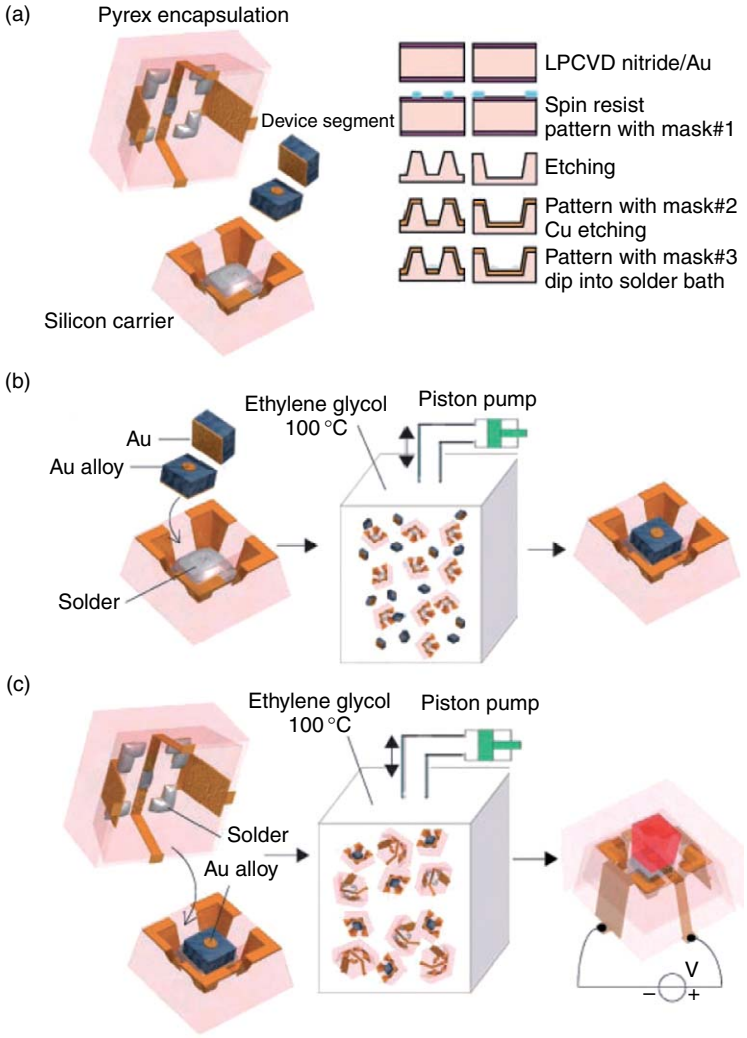


Figure 13 Schematic overview of the solder-directed 3D assembly of LED components. (a) Layouts of carrier, device, and encapsulation components fabricated by surface micromachining and etching. The illustrated device segment is a LED that has two contacts: a small circular anode on the front, and a large square cathode covering the back. The silicon carrier has a solder-coated area in a tapered opening to host a single semiconductor device segment. The encapsulation unit has five solder-coated copper areas inside a tapered opening to connect to corresponding contact pads on the device and the carrier. (b) Chip-on-carrier assembly and (c) chip encapsulation in an ethylene glycol solution at 100 °C, where the solder is melted. The components, agitated by using a piston pump, self-assemble and form a 3D circuit path between device layers that allows testing in a surface mount device configuration. (Photo courtesy: Zheng W, Buhlmann P, Jacobs H O 2004 Sequential shape-and-solder-directed self-assembly of functional microsystems. *Proc. Natl. Acad. Sci. USA* **101**, 12814–17.)

environment to avoid oxidation of the melting solder. Since a LED segment has electrical polarity, i.e., it can be lit only with an appropriate positive DC bias, two electrodes on either side of each LED segment should be distinguished with some pattern: the gold covered the whole bottom surface and 1/9th of the top surface. Hundreds of LED components and a flexible assembly template were placed inside a vial filled with water at a temperature above the melting point of the solder. The LED components were tumbled inside the vial, and then they were attracted, aligned, and bonded to the receptor sites on the template by the surface tension forces of the melting solder. By controlling the agitation intensity, only the bottom surfaces 100% covered gold instead of the top surface with 1/9 covered gold were bonded to solder droplets because the adhesion is roughly proportional to the gold-covering area on the LED segments. Therefore, the LED components self-assembled with unique face orientations. Gold pads

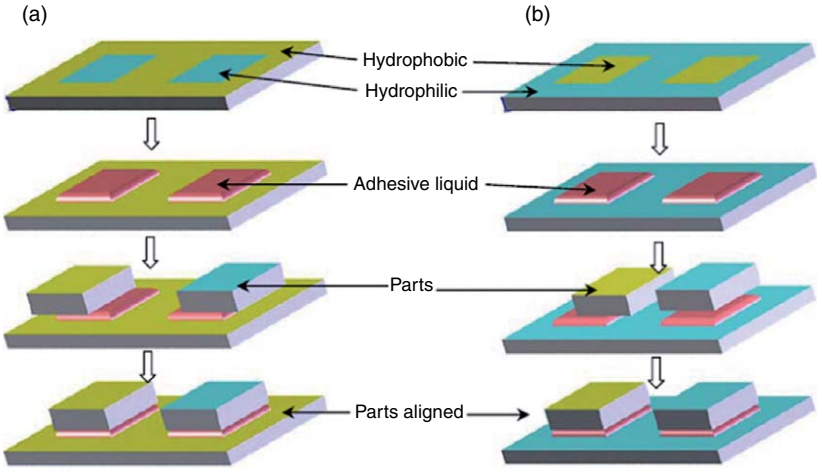


Figure 14 Schematic process overviews of capillary-driven self-assembly of flat microcomponents: (a) in an aqueous environment and (b) in an air environment.

on the top surfaces of the bonded LED components were coated with melting solder by another dip coating. As a final assembly step, a flexible transparent film patterned with electrical circuits and solder bumps was roughly and manually aligned to the assembly template, i.e., a wafer-level flip chip bonding process, and the film was aligned with high accuracy to the assembly template by surface tension of the melting solder. Thus, the electrical connections to the top electrodes of the LED segments were established. Assembly was successful by testing with a low defect rate of ~2% (**Figure 15**).

Srinivasan *et al.* (2001) developed and demonstrated another fluidic self-assembly technique. They patterned a substrate with an array of hydrophobic thiolated Au binding sites. When the substrate was inserted into water through a film of hydrophobic adhesive floating on the water surface, the adhesive covered only the hydrophobic binding sites. Then they introduced microparts fabricated from silicon-on-insulator wafers through a pipette to the substrate in the water. When the hydrophobic pattern on the microparts came in contact with the adhesive, shape matching occurred spontaneously due to interfacial energy minimization. Finally, the adhesive was polymerized by heat or UV light depending on the type of the adhesive, so that the bonding became permanent. Binding sites of shapes with in-plane rotational symmetries such as squares gave alignment yields up to 100% (**Figure 16(a)**). The translational and rotational misalignments were <0.2 μm and within ~0.3°, respectively. Binding sites without in-plane rotational symmetries (aimed at a

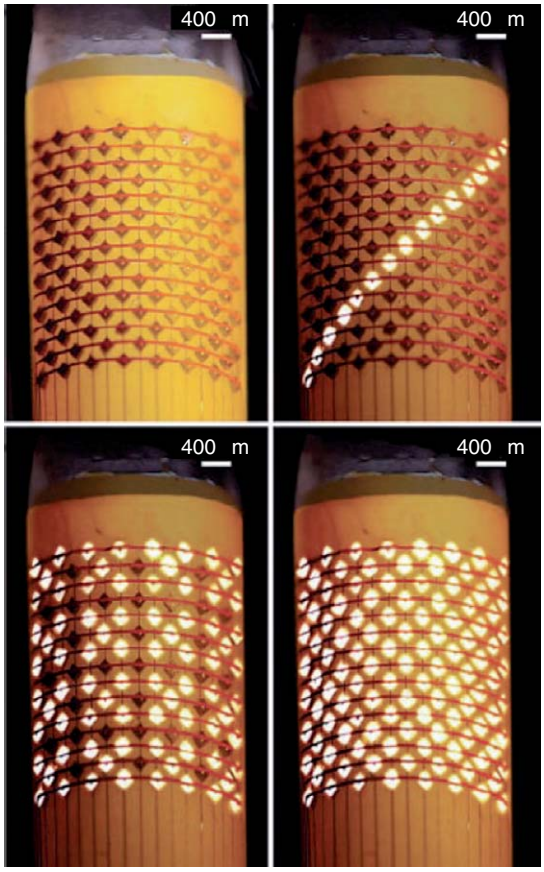


Figure 15 Optical photographs of tested light-emitting diode (LED) arrays self-assembled on a flexible substrate by capillary forces from low-temperature melting solder. (Photo courtesy: Jacobs H O, Tao A R, Schwartz A, Gracias D H, Whitesides G M 2002 Fabrication of cylindrical display by patterned assembly. *Science* **296**, 323–5.)

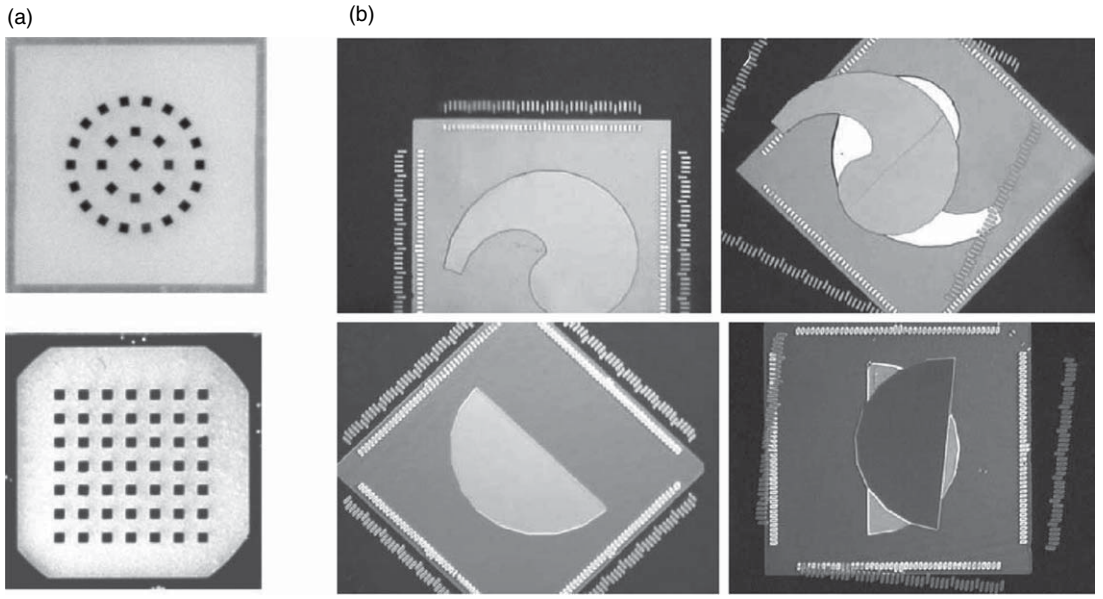


Figure 16 Optical micrographs of capillary-driven self-assembly of flat microparts: (a) square parts on quartz substrates in ring and grid configurations and (b) correct and wrong alignments for semicircle- and comma-shaped binding sites. (Source: Srinivasan U, Liepmann D, Howe R T 2001 Microstructure to substrate self-assembly using capillary forces. *J. Microelectromech. Syst.* **10**, 17–24.)

uniquely oriented alignment) such as semicircles and commas gave alignment yields of approximately 30–40% (**Figure 16(b)**).

To achieve self-assembly with unique in-plane orientations, Liang *et al.* (2004) designed and demonstrated a capillary-driven self-assembly process with special geometrical designs for binding sites. Both the parts and the binding sites are of an offset ring shape, which provides a unique global interfacial energy minimum (**Figure 17**): the energy calculation is based on the assumption that the interfacial energy has an approximately linear relationship with the overlap area between a part and the binding site (Böhringer *et al.* 2001).

Capillary-driven self-assembly can also proceed in an air environment as long as the receptor sites are coated with liquid droplets that act as a medium to attach flat microcomponents. Fang *et al.* (2006) have demonstrated a capillary-driven self-assembly method to surface mount PZT (lead zirconate titanate) actuators for microfluidic devices. The hydrophilic receptor sites are precisely recessed wells with an opening smaller than the PZT actuators, and the bonding face of each PZT actuator has a hydrophilic area of identical size at the center as the recess's opening. All the surfaces except the receptor wells and the bonding areas on the PZT actuators are hydrophobic. A simple dip coating process left adhesive droplets only in receptor wells. When PZT actuators were attached to these adhesive droplets, they self-aligned to minimize interfacial energies. Then, the aligned PZT actuators were pressed down such that their rims touched the substrate with excess adhesive squeezed out, and the electrically insulated adhesive was polymerized by heating to form permanent bonds. Electrical and mechanical connections were established at the rim and at the center, respectively, of the PZT actuators.

To introduce bulk microcomponents to receptor sites during dry assembly, Liang *et al.* (2005) developed a method using an agitated diaphragm to bounce microcomponents until some were caught by water droplets on the receptor sites on the downward-facing assembly substrate (**Figure 18**). This process demonstrated fast assembly (<30 s) and high yields (>90%), and fast recycling of redundant components. The glass substrate was coated with hydrophobic thiolated Au, which was patterned with an array of openings (exposure of the glass substrate). The exposed glass receptor sites were hydrophilic and covered with water droplets after a dip coating process. The capillary-driven alignment was a subsecond process (**Figure 19**).

1.14.3.6.3 Multibatch assembly

Xiong *et al.* demonstrated a method for controlled multibatch self-assembly, based on the fluidic self-assembly technique developed by Srinivasan *et al.*, of microparts by capillary force. They patterned an

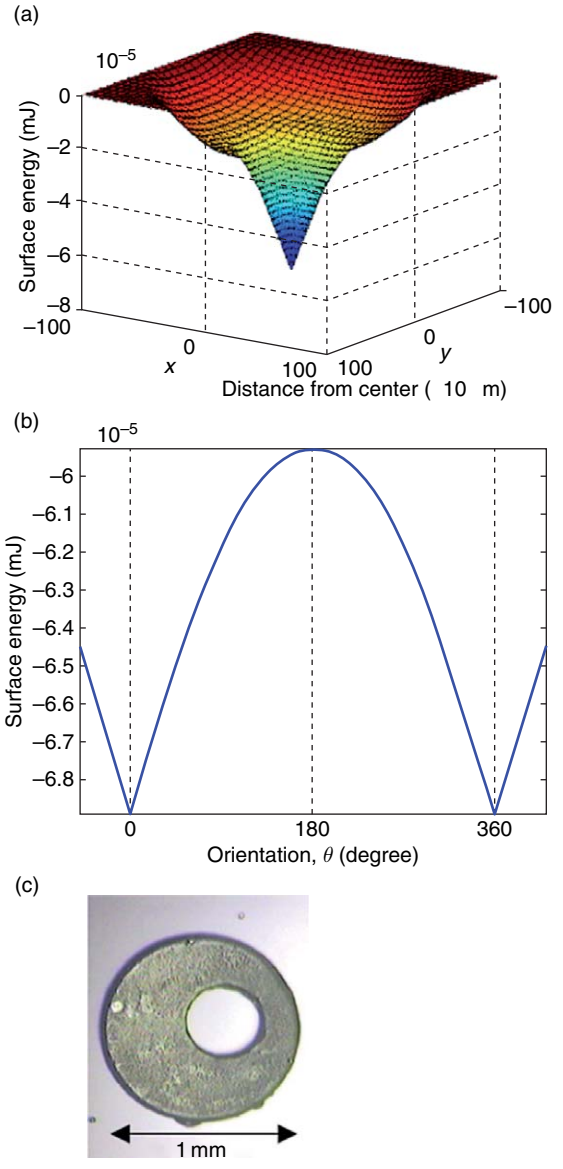


Figure 17 Energy profiles for an offset ring design: (a) the part with the correct in-plane orientation moves in a translational way across the binding site; (b) the part rotates about its center, which coincides with the binding site's center; and (c) an experimental result shows exact alignment of the silicon part over the thiolated gold binding site in water. (Photo courtesy: Liang S, Xiong X, Böhringer K F 2004 Towards optimal designs for self-alignment in surface-tension driven micro-assembly. *Proc. IEEE Conf. Micro Electro Mechanical Systems (MEMS)*, Maastricht, The Netherlands, pp. 9–12.)

oxidized silicon substrate with an array of gold binding sites, and these hydrophilic gold binding sites became hydrophobic after adsorbing a SAM. To be active, i.e., to adsorb an adhesive lubricant droplet for later

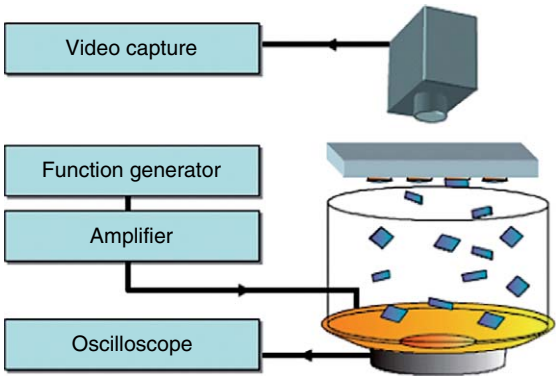


Figure 18 Schematic experimental setup: a sinusoidal wave signal from a function generator drives the speaker diaphragm to bounce microcomponents up until they make contact with the binding sites of the downward-facing assembly substrate. (Photo courtesy: Liang S, Wang K, Böhringer K F 2005 Self-assembly of MEMS components in air assisted by diaphragm agitation. *Proc. IEEE Conf. Micro Electro Mechanical Systems*, pp. 592–5.)

anchoring of parts, a binding site should be hydrophobic. They demonstrated an electrochemical method to desorb the SAM from the gold binding sites: an appropriate electrochemical potential was applied between the gold binding sites and an aqueous environment. Before the first batch assembly, some binding sites were selectively deactivated by SAM desorption, to be reactivated for the next batch assembly. An array of commercial LED components was assembled on the substrate with the heat polymerizable adhesive, and the electrical connections between the substrate and the LED components were established by solder electroplating (**Figure 20**).

1.14.3.7 Multistage Self-Assembly

1.14.3.7.1 Semidry, uniquely orienting, self-organizing parallel assembly

Fang and Böhringer have developed a semi-DUO-SPASS (dry, uniquely orienting, self-organizing parallel assembly) process for batch assembly of flat microcomponents. This semidry assembly process, based on the assembly strategy shown in **Figure 21**, includes two wet process steps followed by two dry process steps.

The parts used in demonstration were dummy flat silicon components with dimensions of 2 mm × 2 mm × 0.5 mm. Each part had a circular peg (ϕ 0.3 mm) located offset from the center of mass. The part surface opposite to the peg was the only hydrophobic face coated with thiolated gold (**Figure 22(a)**). Correspondingly, each

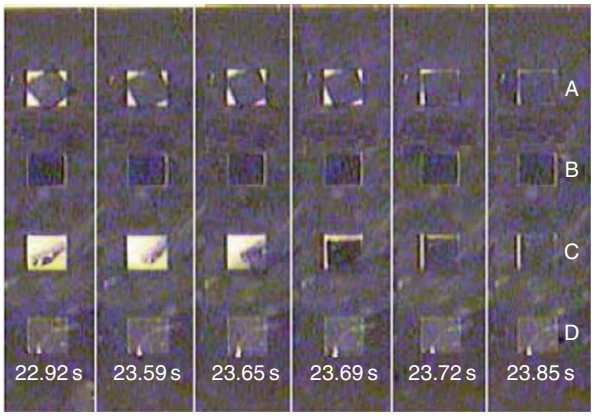


Figure 19 Video frames of one column of a receptor site array from an assembly experiment: A, alignment driven by surface tension; B and D, correctly aligned components; C, vertical attachment corrected by impacts. (Photo courtesy: Liang S, Wang K, Böhringer K F 2005 Self-assembly of MEMS components in air assisted by diaphragm agitation. *Proc. IEEE Conf. Micro Electro Mechanical Systems*, pp. 592–5.)

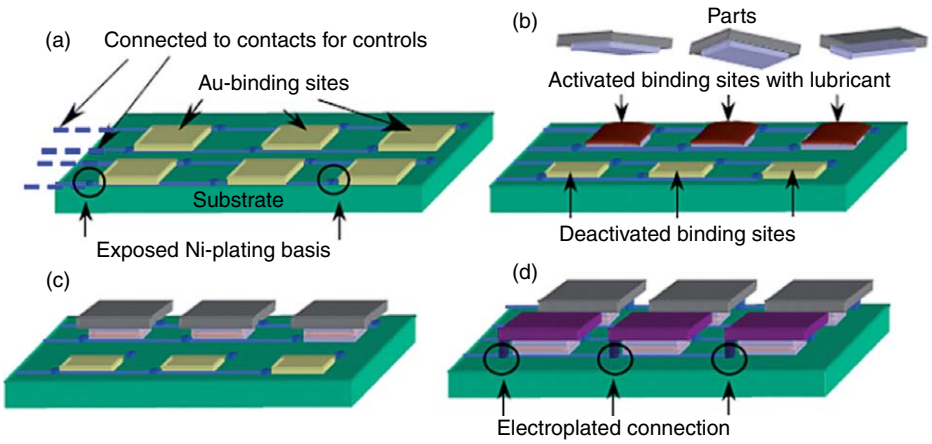


Figure 20 Multibatch assembly schematic flow: (a) A fabricated substrate for assembly with Au binding sites and Ni-plating basis. (b) A substrate prepared for first batch assembly. The substrate is immersed in water with lubricant wetting exclusively the activated binding sites. (c) First batch assembly. (d) By repeating the assembly process, the second batch of components is assembled. Electroplating is performed afterward to establish electrical interconnections. (Photo courtesy: Xiong X, Hanein Y, Fang J, Wang Y, Wang W, Schwartz D T, Böhringer K F 2003 Controlled multi-batch self-assembly of micro devices. *IEEE J. Microelectromech. Syst.* **12**, 117–27.)

receptor site on an assembly substrate had a circular well with a large diameter for easy adoption of a peg, but the well's diameter should have an upper limit to ensure that one well adopts exactly one peg (**Figure 22(b)**).

The first wet process step is to uniquely face-orient bulk silicon parts. An agitated water surface was exploited to selectively float silicon components with only their hydrophobic faces upward. For floating parts, the other two poses, horizontal with the hydrophilic face upward and vertical, are not stable due to surface tension forces. It was observed that the uniquely face-oriented floating parts stayed together in a single layer after agitation was turned off for several minutes, and this monolayer assembly phenomenon is driven by potential energy minimization.

A hydrophobic carrier wafer was employed to transfer the uniquely face-oriented floating parts. When the hydrophobic carrier wafer entered water vertically, the water surface curved downward near the hydrophobic surface to approach a contact angle $>90^\circ$. Such downward-curving surface resulted in a water valley near the carrier wafer surface, and this valley attracted the nearby floating

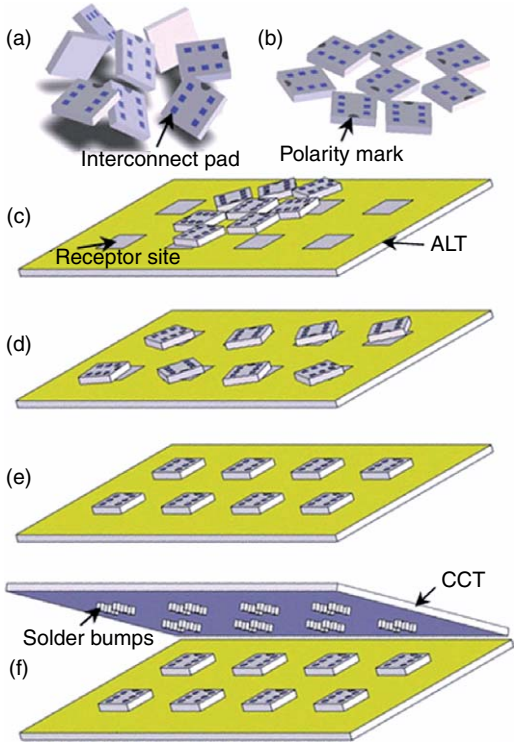


Figure 21 Batch assembly strategy: (a) randomly oriented bulk parts; (b) bulk parts are uniquely face-oriented and spread in a single layer; (c) parts are palletized on an alignment template; (d) parts are distributed one-to-one to receptor sites; (e) parts are aligned to the receptor sites with unique in-plane orientations; and (f) parts are bonded to a chip carrier template (CCT) via wafer-level flip chip bonding. (Source: Fang J, Böhringer K F 2006 Wafer-level packaging based on uniquely orienting self-assembly (the DUO-SPASS processes). *ASME/IEEE J. Microelectromech. Syst.* **15**, 531–40.)

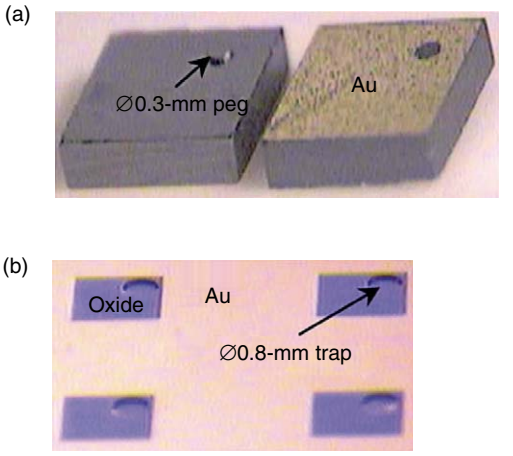


Figure 22 Microscope images of parts and receptor sites: (a) top and bottom views of 2-mm square silicon parts; (b) a 2×2 section of an array of 2-mm square receptor sites on a 4-in. alignment template. (Source: Fang J, Böhringer K F 2005 High yield batch packaging of micro devices with uniquely orienting self-assembly. *Proc. IEEE Conf. Micro Electro Mechanical Systems*, Miami, FL, USA, pp. 12–15.)

parts to minimize their potential energies (**Figure 23(a)**). As the carrier wafer continued to enter water, floating parts started to adhere to the carrier wafer and be submerged in water. The adhering of parts is due to significant capillary forces introduced by trapped air bubbles between the parts and the carrier wafer (**Figure 24**). After all the floating parts adhered to the carrier wafer, the carrier wafer was withdrawn from water, and more than ~98% parts stayed attached. The final step to transfer these parts was to bring the carrier wafer approximately parallel to an alignment template on a 120 °C hotplate (**Figure 23(b)**), and these parts were then released to touch the alignment template with their pegs when water evaporated.

The next process step for the assembly sequence is to distribute the palletized silicon parts to an array of receptor sites on the alignment template. Orbital shaking was utilized to drive the parts to move until their pegs were trapped to receptor sites. Since orbital shaking provides a centrifugal force field, which is uniform in both magnitude and direction, and since the force direction is rotating in phase with the orbital shaker's rotating arms, parts are randomly moving locally instead of unidirectionally. Therefore, each unoccupied trap has an equal chance to have some parts to pass by, i.e., have a chance to trap the moving parts. Our trapping experiments with varying part redundancies gave yield–time curves shown in **Figure 25**.

The last process step was achieved by gravitationally driven self-alignment of anchored parts. When the alignment template was tilted in an appropriate orientation, gravitational forces drove the parts to rotate about their anchoring pegs to approach the equilibrium position with the lowest potential energies as a simple pendulum does (**Figure 26(a)** and **26(b)**). Kinetic frictional forces became smaller when the slope of the alignment template increased. Some parts can get stuck due to dust contamination or adsorbed moisture from the environment. A slight agitation at the edge of the alignment template (the impacting force must be in the plane of the alignment template) can break the sticking and free the parts to rotate. If the tilting orientation of the alignment template is precisely controlled, high accuracy of alignment can be achieved. Without good control of

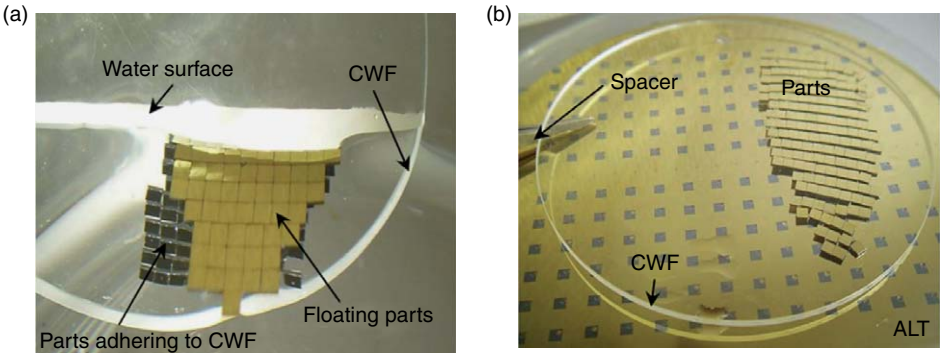


Figure 23 The palletizing process: (a) a carrier wafer inserted into water, and floating parts adhere to the hydrophobic PFC802 surface firmly; (b) the carrier wafer brought approximately parallel to an alignment template to release the adhering parts when water evaporates. (Source: Fang J, Böhringer K F 2005 High yield batch packaging of micro devices with uniquely orienting self-assembly. *Proc. IEEE Conf. Micro Electro Mechanical Systems*, Miami, FL, USA, pp. 12–15.)

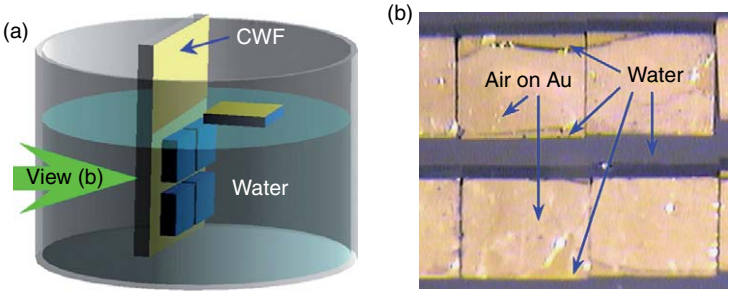


Figure 24 Mechanism of parts adhering to the hydrophobic surface: (a) a schematic view of a carrier wafer inserted vertically into water to pick up floating parts; (b) a microscopic image of trapped air bubbles underneath the parts adhering to the transparent carrier wafer in water. (Source: Fang J, Böhringer K F 2005 High yield batch packaging of micro devices with uniquely orienting self-assembly. *Proc. IEEE Conf. Micro Electro Mechanical Systems*, Miami, FL, USA, pp. 12–15.)

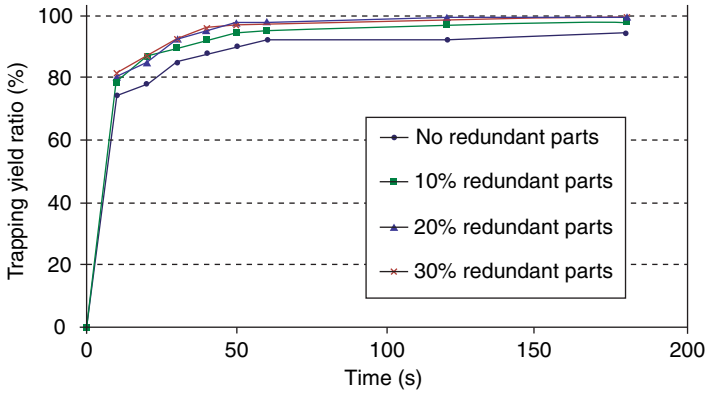


Figure 25 Trapping results of 2-mm square silicon parts on a 164 receptor site alignment template without and with 10–30% redundancy of parts. (Source: Fang J, Böhringer K F 2005 High yield batch packaging of micro devices with uniquely orienting self-assembly. *Proc. IEEE Conf. Micro Electro Mechanical Systems*, Miami, FL, USA, pp. 12–15.)

the template tilting, an additional wet fine alignment step was used: steam condensate was introduced into the parts and the template where steam formed film-wise and droplet-wise condensation on hydrophilic and hydrophobic surface areas, respectively. The receptor sites were hydrophilic with a hydrophobic background. The capillary gaps underneath the parts attracted steam condensation, so that parts self-aligned

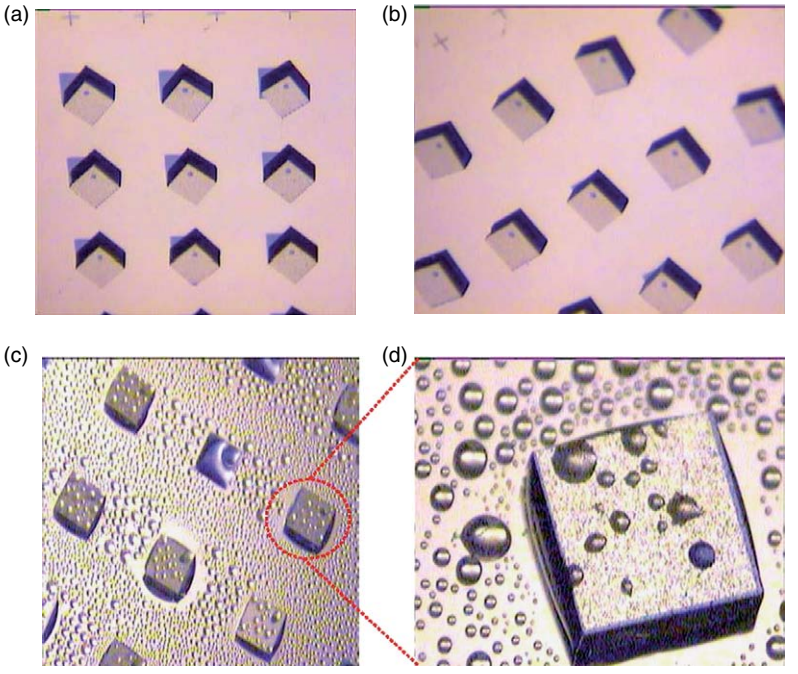


Figure 26 Uniquely aligning of 2-mm square silicon parts: (a) parts roughly aligned by gravity on a tilted alignment template; (b) parts roughly aligned to the receptor sites on an appropriately tilted alignment template; (c) misalignment corrected by capillary-driven self-alignment when water steam is introduced to condense on the alignment template and parts surface; and (d) a zoom view of a part with water condensate. (Source: Fang J, Böhringer K F 2005 High yield batch packaging of micro devices with uniquely orienting self-assembly. *Proc. IEEE Conf. Micro Electro Mechanical Systems*, Miami, FL, USA, pp. 12–15.)

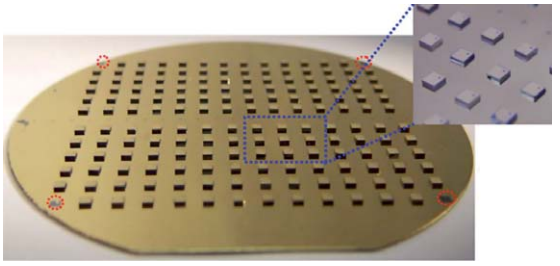


Figure 27 The assembly result: one hundred and sixty-four 2-mm square silicon parts correctly assembled on a 4″ alignment template (the four corner receptor sites were blocked by a covering petri dish to define moving range of the parts during the trapping experiment). (Source: Fang J, Böhringer K F 2005 High yield batch packaging of micro devices with uniquely orienting self-assembly. *Proc. IEEE Conf. Micro Electro Mechanical Systems*, Miami, FL, USA, pp. 12–15.)

to receptor sites by minimizing interfacial energies (**Figure 26(c)** and **26(d)**). After the parts were self-aligned, the water was evaporated by slow heating (rapid heating can cause nonuniform capillary stress on parts and can result in misalignment) to leave a dry aligned array of parts on the alignment template (**Figure 27**). At this stage, the aligned parts are ready to be transferred to a chip carrier template (CCT) via wafer-level flip chip bonding, and then dicing separates these simultaneously assembled or packaged devices.

1.14.3.7.2 Dry, uniquely orienting, self-organizing parallel assembly

To avoid aqueous assembly environments for microcomponents with released microstructures that can be easily immobilized or even broken by surface tension forces, Fang and Böhringer (2006b) developed a completely dry assembly process called DUO-SPASS. A schematic overview of this assembly process is shown in **Figure 28**. The main mechanism for this dry assembly is two-stage shape recognition: each part has one tall circular peg (CP) and one low cross peg (XP); each receptor site has correspondingly a circular trap (CT) and a cross trap (XT); the first stage assembly is achieved by anchoring a CP into a CT; the second stage assembly fixes the part in a specific in-plane orientation when its XP falls into an XT.

The pegs can be fabricated with either an additive or a subtractive method. A possible additive pattern

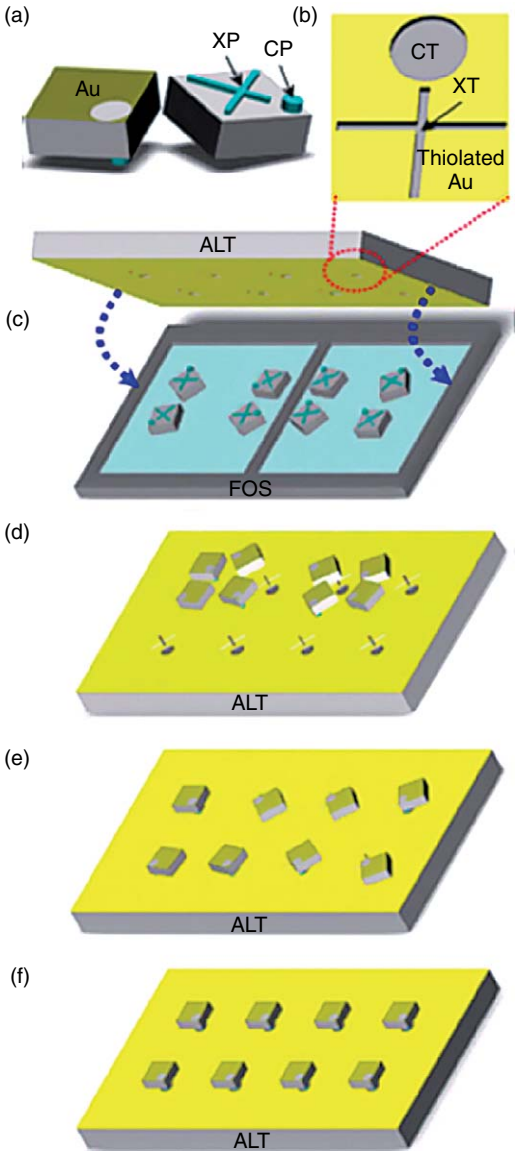


Figure 28 Schematic overview of the dry process to parallel assemble square microparts: (a) top and bottom views of a silicon part (the bottom face has a CP and an XP, and the CP has twice the height of the XP; the top face is coated with gold, and the opening in the gold layer marks the position of the CP); (b) trenches of a receptor site on the alignment template; (c) bulk parts face-oriented on an orbital shaken face-orienting substrate and then sandwiched by adding an alignment template; (d) parts palletized to the alignment template with their peg sides facing downward; (e) parts one-to-one anchored to receptor sites by orbital shaking: the CPs fall into the CTs (the first shape recognition); (f) parts rotated by orbital shaking introduce torques until their XPs fall into the XTs (the second shape recognition). (Source: Fang J, Böhringer K F 2006 Wafer-level packaging based on uniquely orienting self-assembly (the DUO-SPASS processes). *ASME/IEEE J. Microelectromech. Syst.* **15**, 531–40.)

method consists of double-layer SU8 casting (**Figure 29**), and a subtractive method is deep reactive ion etching (DRIE). The traps are fabricated with DRIE, with the depth of the traps being greater than that of the circular peg for stable anchoring.

A unique face orientation of bulk parts based on an asymmetry in dynamic stability was achieved. On an orbitally shaken plate, it is much more stable to rest bulk parts on their flat surfaces than on the pegs, and thus bulk parts are uniquely face-oriented to point upward with their pegs. Uniquely face-oriented parts are then transferred to an alignment template by the method shown in **Figure 28(c)** and **28(d)**.

Centrifugal forces generated by orbital shaking drive the parts to move until they are anchored to the circular traps (the first shape recognition). Since the anchored circular peg is located offset from the center of mass, a torque is exerted on each anchored part. The torque rotates each part until the second shape recognition is achieved: the cross peg falls into a cross trap. During this two-stage shape recognition, parts should move without flipping over, which can be achieved by the following:

(1) Lower orbital shaking speeds (~110 rpm for 1-mm square silicon parts)
(2) Thiolated gold coating on the alignment template for lower sliding friction
(3) An ultrasonic vibrating stage attached to the alignment template to reduce sticking effects and sliding friction.

Assembly results for 1-mm square silicon dummy parts are shown in **Figure 30**.

1.14.3.7.3 Parallel component-to-substrate assembly with controlled poses

Most of the reported parallel microassembly techniques had flat microcomponents assembled in only one pose: horizontal mounting on a substrate with the largest possible footprint. In some special cases, such as radio frequency MEMS (with MEMS = microelectromechanical system) and optical MEMS components, vertical mounting can benefit in the following ways:

(1) Less footprints and higher spatial density
(2) Functional surfaces (assuming to be the large flat surfaces) perpendicular to the assembly substrate.

Especially for VCSEL components, vertical mounting can direct a laser beam parallel to the substrate,

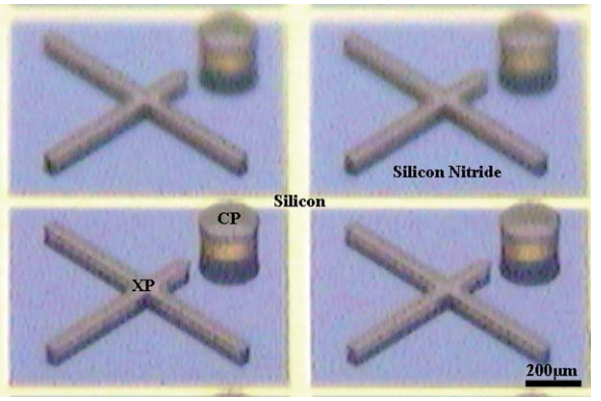


Figure 29 An optical microscope image of a 2 × 2 section of an array with SU8-2025 pegs on a 3″ silicon substrate, before dicing into individual parts. (Source: Fang J, Böhringer K F 2006 Wafer-level packaging based on uniquely orienting self-assembly (the DUO-SPASS processes). *ASME/IEEE J. Microelectromech. Syst.* **15**, 531–40.)

which enables the construction of free-space micro-optical systems and optical detectors for lab-on-chip microsystems.

A strategy to mount flat microcomponents parallelly on a substrate with either vertical or horizontal poses is shown schematically in **Figure 31**. Only a vertical pose is allowed for trapped components in the apertures due to geometrical constraints (**Figure 31(a)**). The surface tension of water is exploited to transfer the trapped vertically standing components onto a palletizing plate having an array of water-covered receptor sites (**Figure 31(b)** and **31(d)**). The transfer process leaves components vertically attached to the palletizing plate, i.e., components are assembled with vertical poses.

Assembly with a horizontal pose for a flat-edge component can be achieved by surface tension torques. The component has one hydrophobic surface, and the opposite surface is hydrophilic. After attaching to a water droplet, the component has more water on its hydrophilic face than does the other faces. The water droplets on the palletizing plate are formed by dip coating, and thus their volumes are limited by the dynamic contact angle of water and the receptor site area. For a 790 μm × 790 μm × 330 μm silicon component, the volume of a dip-coated water droplet cannot introduce enough surface tension torque on the hydrophilic surface to lay down the component. Steam condensation can introduce more water to increase the surface tension torque on the hydrophilic surface, and then the component will lie down on its hydrophilic surface as shown in **Figure 31(d)**. Optical images of an assembly process are shown in **Figure 32**.

1.14.4 Theoretical Aspects of Self-Assembly

The purpose of modeling, as discussed in this section, is twofold:

(1) To develop a theoretical understanding of the physical processes that are essential to self-assembly.
(2) To develop design tools that can help us to optimize engineered self-assembly processes and to establish the theoretical and practical bounds on their performance. The model presented here is very general. It simply assumes the existence of reaction rate constants that specify the speed at which components organize into aggregate structures, and assumes that this speed is proportional to the number of available parts and/or binding sites. Thus, it is applicable to all self-assembly approaches described in this review. However, the rate constants are highly dependent on the specific technique chosen, and they vary significantly with particular system parameters such as, for example, materials properties and scale of the components. Determining these rate constants can be achieved empirically by performing systematic sets of experiments. Alternatively, they can be derived from first principles; because of the complex interplay between the numerous effects that may influence a typical self-assembly process, such an *ab initio* approach remains a challenging task.

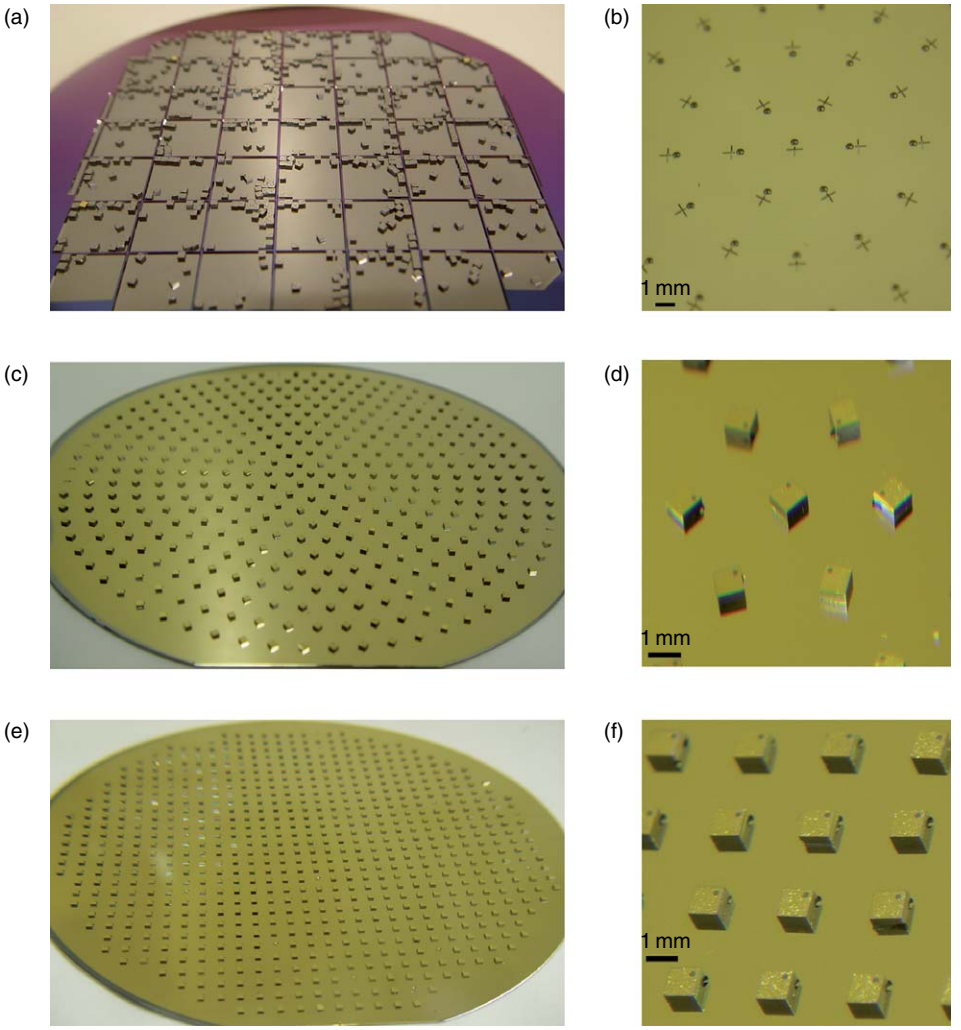


Figure 30 Optical photographs of templates and assembly results: (a) after 1 min of orbital shaking, almost all parts come to rest on their flat faces on the ϕ100-mm face-orienting substrate; (b) a partial view of an alignment template with a polar array of receptor sites; (c) 388 silicon parts assembled on the ϕ100-mm alignment template with a polar array of 397 receptor sites; (d) zoom-in view of the center of the alignment template; (e) 710 silicon parts assembled on the ϕ100-mm alignment template with an orthogonal array of 720 receptor sites; (f) zoom-in view of a 3×3 section of the array of receptor sites. (Source: Fang J, Böhringer K F 2006 Wafer-level packaging based on uniquely orienting self-assembly (the DUO-SPASS processes). *ASME/IEEE J. Microelectromech. Syst.* **15**, 531–40.)

1.14.4.1 Modeling

1.14.4.1.1 Parallels to chemical kinetics

A self-assembly process is, in principle, very similar to a chemical reaction. Typically, a large but finite number of reactants are available to interact and form bonds with each other. In chemistry, these reactants are atoms or molecules; in microscale self-assembly, they are, for example, silicon chips or hydrophobic binding sites. In either case, the reaction is of a stochastic nature and driven by energy minimization; however, the energy minimization for a microscale self-assembly relies on a wide range of possible sources, as described in Section 1.14.3. The equivalent of temperature is usually provided by some kind of mechanical agitation, such as the orbital shaker in the DUO-SPASS process described in Sections 1.14.3.7.1 and 1.14.3.7.2. Hosokawa *et al.* (1994) illustrated this connection to chemical kinetics with assemblies of tiles with moving permanent magnets that provided the binding force. A more detailed investigation followed in Cohn's (1997) Ph.D. thesis.

Whether or not a reaction takes place depends not only on the energy difference between the

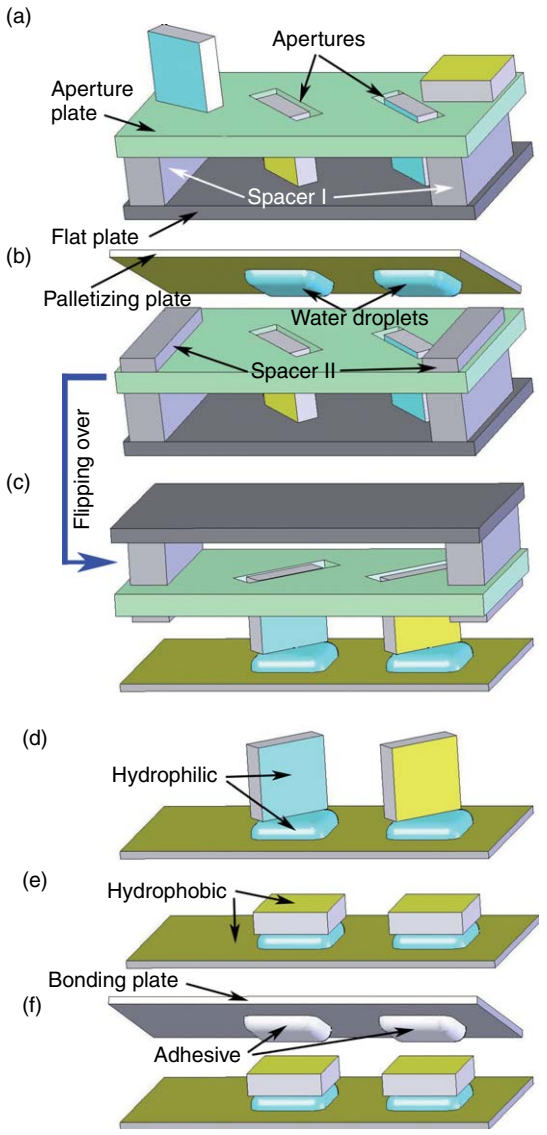


Figure 31 Schematic overview of the assembly process: (a) bulk parts fall into apertures vertically; (b) a palletizing plate carrying water droplets is aligned with the aperture plate; (c) the plates are flipped over to transfer parts on water droplets; (d) parts stand on receptor sites; (e) parts rotate to adhere to the palletizing plate with their hydrophilic oxide faces; (f) parts are permanently bonded to a bonding plate via wafer-level flip chip bonding. (Source: Fang J, Böhringer K F 2006 Parallel micro component-to-substrate assembly with controlled poses and high surface coverage. *IOP J. Micromech. Microeng.* **16**, 721–30.)

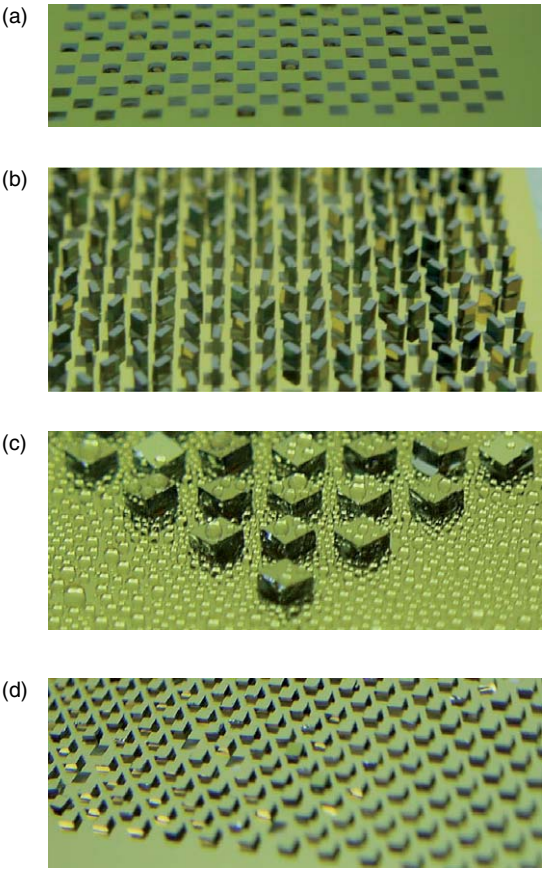


Figure 32 Optical images of 790-μm square parts transferred to a palletizing Pyrex plate via water droplets: (a) a partial view of an array of 790-μm square hydrophilic receptor sites covered with water droplets; (b) parts were transferred and are standing vertically; (c) water steam condensation was introduced on the palletizing plate, where steam formed filmwise and dropletwise condensation on hydrophilic and hydrophobic areas, respectively; (d) parts attached to receptor sites horizontally after steam condensation, with their only hydrophobic thiolated Au surface facing outward. (Source: Fang J, Böhringer K F 2006 Parallel micro component-to-substrate assembly with controlled poses and high surface coverage. *IOP J. Micromech. Microeng.* **16**, 721–30.)

unassembled and the assembled states, but also on the activation energy and the overall temperature of the system. Stiction between parts and substrates is a typical example of an energy barrier that has to be overcome either with careful engineering of the self-assembly system (e.g., nonstick hydrophobic surfaces and low humidity conditions) or with activation energy (e.g., vibration). This activation energy can be understood as the temperature of the system, and agitation with a piezoactuator or an orbital shaker is the analog of Brownian motion.

Consider the semi-DUO-SPASS self-assembly system described in Section 1.14.3.7.1 as an example. Here, gravity provides the potential energy difference that causes the parts to assemble in the cavities. If we have n binding sites and rn parts ($r \geq 1$ is the

part redundancy) we can initially expect a forward reaction rate of $K_f rn^2$, where K_f is the rate constant of the forward reaction. As parts are consumed and binding sites fill up, this rate changes to $K_f(rn-x)(n-x)$ with x being the number of already occupied sites. We can thus express x through a differential equation: $dx/dt = K_f(rn-x)(n-x)$. This equation has a closed-form solution

$$x(t) = \frac{rn(1-\exp(-K_f(r-1)nt))}{r-\exp(-K_f(r-1)nt)} \quad [1]$$

and x approaches 100% as t goes toward infinity. In **Figure 33**, the rate constant is set to $K_f = 1$ part per time unit and $r = 1.1$, 1.2, and 1.3.

In practice, 100% yield is difficult to achieve because (a) assembly has to be terminated in a finite – and usually short – amount of time and (b) there may also exist a nonzero reaction in the reverse direction that causes occasional breakup of already assembled parts. Let K_r be the rate constant for the reverse reaction, and we expect that usually $K_r < K_f$. Then eqn [1] needs to be updated to include the reverse reaction whose rate is $K_r x$. Note that here x is a linear term and not a quadratic term as in the forward reaction. This is due to the combinatorics of this particular self-assembly system, where the number of free part–free binding site combinations is quadratic in x but the number of occupied sites is linear in x. Thus we obtain a new differential equation $dx/dt = K_f(rn-x)(n-x) - K_r x$. This equation, again, can be solved exactly, and we can derive the equilibrium equation

$$x = \frac{K_{rf} + (r+1)n}{2} - \sqrt{\left(\frac{K_{rf} + (r+1)n}{2}\right)^2 - rn^2} \quad [2]$$

where K_{rf} stands for K_r/K_f. For $K_{rf} = 0.1$ and $r = 1.1$, 1.2, and 1.3, we now obtain equilibria at 97.0%, 98.2%, and 98.7%, respectively (**Figure 34**).

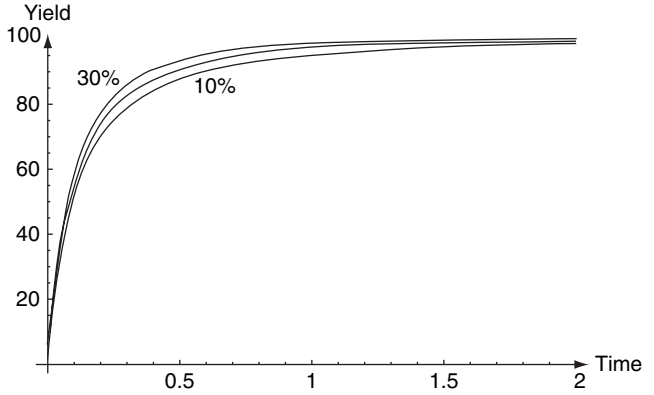


Figure 33 Theoretically determined yield of a self-assembly process described by eqn [1] as a function of time, with part redundancy at 10%, 20%, and 30%.

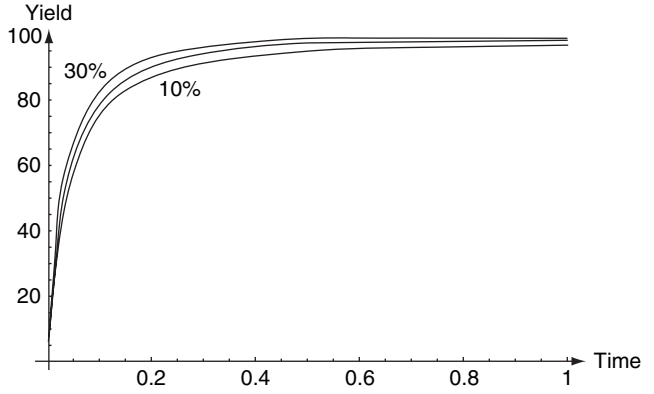


Figure 34 Theoretically determined yield of a self-assembly process that takes into account forward reaction rate $K_f = 1$ and reverse reaction rate $K_r = 0.1$. Again, part redundancy is 10%, 20%, and 30%. The self-assembly processes reach a yield of 97.0%, 98.2%, and 98.7%, respectively.

This analysis could be further refined in many ways. For example, note that eqns [1] and [2] assume a reaction-limited process. In practice, this means that parts are sufficiently mobile to reach all binding sites. If instead a diffusion-limited process is chosen (i.e., the parts' mobility is more constrained such that only parts in the immediate vicinity of a binding site can react) then the equation must be adjusted again. Another limitation of eqns [1] and [2] is the application of a continuous equation to a discrete system. Alternatively, one can express the assembly in terms of discrete states, and each reaction corresponds to a state transition. Such a system can be represented by a directed graph where each node represents a state and each edge a (probabilistic) transition. The theory of Markov processes provides a powerful and well-developed foundation for this approach.

1.14.4.1.2 Predicting and optimizing performance

When describing a self-assembly system, we can distinguish between process parameters and performance parameters. Process parameters describe the physical conditions under which the system operates: geometry, material properties, environmental conditions, part counts, etc. Performance parameters provide a metric to evaluate the self-assembly process: yield, assembly time, accuracy, wear on parts, etc.

The model equations in Section 1.14.4.1.1 indicate process parameters that influence specific performance parameters. If our objective is to maximize the number of assemblies x in a given time t, then intuitively we would choose a large part redundancy r and a large ratio between the rate constants K_f and K_r. The equations confirm this strategy but also give bounds on the yield, which we can expect from a practical experimental setup.

Calculating exact values for the rate constants K_f and K_r for self-assembly processes from first principles is difficult in general. The challenge lies in the usually complex interplay of different phenomena (gravity, capillarity, van der Waals forces, etc.) and in the difficulty to find the equivalent of temperature (i.e., kT) in reaction kinetics. Further research is needed in this area to establish a generally applicable theory of microscale self-assembly.

1.14.4.2 Parallels between Assembly and Computation

Another fascinating aspect of self-assembly is its close relationship to computation. Adleman's seminal paper showed that nondeterministic polynomial time (NP)-hard problems can be encoded into DNA molecules (Adleman 1994). Moreover, it is well-known that self-assembly systems consisting of simple 2D tiles can implement a Turing machine, and thus can perform (in principle) any computation (Grünbaum and Shephard 1986, Wang 1963). This result may not lead to practical computers based on self-assembly, but it indicates that virtually arbitrarily complex structures could be created via self-assembly, given a sufficiently sophisticated program that guides the assembly toward the desired product. Real-time error correction is essential because, as we have seen throughout Section 1.14.3, stochastic processes cannot be expected to produce perfect yield. To date, biological growth in nature remains by far the best proof of principle for successful self-assembly.

1.14.5 Discussion and Outlook

Self-assembly techniques enable rapid and parallel construction processes for MEMS devices. With appropriate driving forces, self-assembly techniques can assemble components ranging from submicrons to millimeters or larger scales, while conventional robotic pick-and-place assembly methods face mounting difficulties below 500 μm due to sticking effects between grippers and components. The following research directions are worthy of efforts to further extend the capabilities of self-assembly techniques.

In the previous sections, several major self-assembly techniques driven by inertial, electrical, magnetic, and interfacial energy minimization have been discussed. Each of these self-assembly techniques can achieve both 2D and 3D assemblies with appropriate structure designs. Current and potential applications of these self-assembly techniques are given in **Table 1**.

Three-dimensional self-assembly methods will most likely play a key role in building MEMS devices with more complicated architectures. Most of the currently available self-assembly techniques mount microcomponents on a substrate surface, i.e., a 2D assembly process. To achieve multiple functions or stand-alone applications, a MEMS device can be integrated with various components such as MEMS sensing units, MEMS actuating units, or complementary metal oxide semiconductor (CMOS) signal processing units. Three-dimensional assembly

Table 1 Current and potential applications of each self-assembly technique

Inertial	2D	• Mass fabrication of RFID tags by fluidic self-assembly • Self-assembly of microcomponents by single- or double-stage shape recognition
	3D	• Construction of 3D structures by centrifugal forces driving hinged structures • Potential construction of 3D structures by feature mating
Interfacial	2D	• Patterning and aligning of microcomponents on a substrate by capillary forces of melting solder or other liquids
	3D	• Integration of different types of microcomponents to form a hybrid composite component • Construction of 3D structures by reflowing solder or photoresist deposited at hinges
Electrical	2D	• Patterning of micro- and nanoscale components on a substrate • Programmable assembly of hybrid components
	3D	• Construction of 3D structures by triboelectricity • Potential aligning of nanowires in desired orientations
Magnetic	2D	• Integration of semiconductor components to recessed sites with magnetized strips on a substrate • Potential programmable assembly of heterogeneous components
	3D	• Construction of 3D frames by sequential pricking up of hinged structures • Potential building of 3D structures by attaching components of specific geometric design

enables dense packaging of these components, which can greatly reduce environmental disturbances and signal transmission losses among different functional units.

Assembly programmability will greatly widen the application range of self-assembly techniques and enable construction of more versatile MEMS devices. This programmability can indicate several situations such as the following:

(1) Integration of different types of components by multiple batches of self-assembly
(2) Self-assembly of same types of components into various patterns in different batches
(3) Selective disassembly of some components by breaking up their bonds to receptor sites
(4) Assembly of multiple components in appropriate sequences to form 3D structures

A consequence of the stochastic nature of self-assembly is that in most cases, perfect yields of 100% cannot be expected. As a remedy, redundancy could be incorporated into the assembly, for example, by including more than the minimum required amount of components. Intentional disassembly of incorrectly placed components is also a possibility, but it would require binding sites that can be deactivated even after assembly has already taken place. Finally, a precision robotic assembly step could also be employed to place any remaining defects.

Assembly process automation enables running the self-assembly process continuously for maximum throughput. To this end, one needs to consider how the process can be integrated into the overall production flow, including parts feeding, redundant parts recycling, assembly defect recognition, and recovery. The high capital investment and specialized equipment used in prior and subsequent processing steps should not be affected by the introduction of a new process step. Thus, ideally, the inputs and outputs of the self-assembly process module should be identical with the conventional assembly module that it replaces. This may imply additional constraints on how the substrate and parts are delivered to and from the self-assembly module.

Acknowledgments

The authors would like to thank Kerwin Wang, Xiaorong Xiong, Yanbing Wang and Sheng-Hsiung Liang in the MEMS research group of the University of Washington, Rajashree Baskaran at Intel Corporation, Vaidyanathan Kripesh and his team at A*STAR IME for helpful comments and discussions. This work was supported in part by NSF award ECS-05-1628, NIH Center of Excellence in Genomic Science and Technology grant 1-P50-HG002360-01, DARPA DSO award FAA9550-04-0257, and US Department of Justice award 2000-DT-CX-K001.

References

Adleman L 1994 Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021–4

Böhringer K F, Goldberg K, Cohn M, Howe R T, Pisano A 1998 Parallel microassembly with electrostatic force fields. *Proc. Int. Conf. Robotics and Automation (ICRA)*, Leuven, Belgium, pp. 1204–11

Böhringer K F, Srinivasan U, Howe R T 2001 Modeling of fluidic forces and binding sites for fluidic self-assembly. *Proc. IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS)*, Interlaken, Switzerland, pp. 369–74

Cohn M 1997 Assembly techniques for microelectromechanical systems. Electrical Engineering and Computer Sciences. Ph.D. thesis, University of California

Cohn M B, Howe R T, Pisano A P 1995 Self-assembly of microsystems using non-contact electrostatic traps. *Proc. ASME Int. Cong. Exposition, Symp. Micromechanical Systems, (IC '95)*, San Francisco, CA, USA, pp. 893–900

Fang J, Böhringer K F 2005 High yield batch packaging of micro devices with uniquely orienting self-assembly. *Proc. IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS)*, Miami, FL, USA, pp. 12–15

Fang J, Böhringer K F 2006a Parallel micro component-to-substrate assembly with controlled poses and high surface coverage. *IOP J. Micromech. Microeng.* **16**, 721–30

Fang J, Böhringer K F 2006b Wafer-level packaging based on uniquely orienting self-assembly (the DUO-SPASS processes). *ASME/IEEE J. Microelectromech. Syst.* **15**, 531–40

Fang J, Wang K, Böhringer K F 2006 Self-assembly of PZT actuators for micro pumps with high process repeatability. *ASME/IEEE J. Microelectromech. Syst.* **15**, 871–8

Fearing R S 1995 A planar milli-robot on an air bearing. *Proc. Int. Symp. Robotics Research*, Herrsching, Germany, pp. 570–81

Gracias D H, Tien J, Breen T L, Hsu C, Whitesides G M 2000 Forming electrical networks in three dimensions by self-assembly. *Science* **289**, 1170–2

Grünbaum B, Shephard G C 1986 *Tilings and Patterns*. Freeman, New York

Harsh K F, Kladitis P E, Zhang Y H, Dunn M L, Bright V M, Lee Y C 2000 Tolerance and precision study for solder self-assembled MEMS. *Proc. SPIE* **4075**, 173–84

Hosokawa K, Shimoyama I, Miura H 1994 Dynamics of self-assembling systems – Analogy with chemical kinetics. *Artif. Life* **1**, 413–27

Iwase E, Shimoyama I 2005 Multi-step sequential batch self-assembly of three-dimensional micro-structures using magnetic field. *Proc. IEEE Int. Conf. Micro Electromechanical Systems (MEMS)*, Miami Beach, FL, USA, pp. 588–91

Jacobs H O, Tao A R, Schwartz A, Gracias D H, Whitesides G M 2002 Fabrication of cylindrical display by patterned assembly. *Science* **296**, 323–5

Kaajakari V, Lal A 2001 Electrostatic batch assembly of surface MEMS using ultrasonic triboelectricity. *14th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS)*, Interlaken, Switzerland, pp. 10–13

Kladitis P E, Bright V M 2000 Prototype microrobots for micro-positioning and micro-unmanned vehicles. *Sens. Actuators* **A80**, 132–7

Lai K W C, Hui A P, Li W J 2002 Non-contact batch micro-assembly by centrifugal force. *Proc. IEEE Int. Conf. Micro Electro Mechanical System (MEMS)*, Las Vegas, NV, pp. 184–7

Liang S, Xiong X, Böhringer K F 2004 Towards optimal designs for self-alignment in surface-tension driven micro-assembly. *Proc. IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS)*, Maastricht, The Netherlands, pp. 9–12

Liang S, Wang K, Böhringer K F 2005 Self-assembly of MEMS components in air assisted by diaphragm agitation. *Proc. IEEE Conf. Int. Micro Electro Mechanical Systems (MEMS)*, Miami Beach, FL, USA, pp. 592–5

Perkins J, Rumpler J, Fonstad C G 2002 Magnetically assisted self-assembly – A new heterogeneous integration technique. *MIT Microsystems Technology Laboratories Annual Report*

Silicon Chip Online 2006 *RFID tags – How they work*. http://www.siliconchip.com.au/cms/A_30750/article.html

Skidmore G D, Ellis M, Parker E, Sarkar N, Merkle R 2000 Micro-assembly for top-down nanotechnology. *Proc. Int. Symp. Micromechatronics and Human Science*, Nagoya, Japan, pp. 3–9

Srinivasan U, Liepmann D, Howe R T 2001 Microstructure to substrate self-assembly using capillary forces. *J. Microelectromech. Syst.* **10**, 17–24

Stauth S A, Parviz B A 2005 Self-assembled silicon networks on plastic. *Proc. 13th Int. Conf. Solid-State Sensors and Actuators,* Vol. 1, pp. 964–7

Syms R R A, Yeatman E M, Bright V M, Whitesides G M 2003 Surface tension-powered self-assembly of microstructures – The state-of-the-art. *IEEE J. Microelectromech. Syst.* **12**, 387–417

Tien J, Terfort A, Whitesides G M 1997 Microfabrication through electrostatic self-assembly. *Langmiur* **13**, 5349–55

Wang H 1963 *Proc. Symp. Math. Theory of Automata,* New York, NY, USA

Warneke B, Last M, Liebowitz B, Pister K S J 2001 Smart Dust: Communicating with a cubic-millimeter computer. *Computer* **34**, 44–51

Xiong X, Hanein Y, Fang J, Wang Y, Wang W, Schwartz D T, Böhringer K F 2003 Controlled multi-batch self-assembly of micro devices. *IEEE J. Microelectromech. Syst.* **12**, 117–27

Yang H, Lin C, Fang W 2006 Wafer level self-assembly of microstructures using global magnetic lifting and localized induction welding. *IOP J. Micromech. Microeng.* **16**, 27–32

Yeh H J, Smith J S 1994 Fluidic self-assembly for the integration of GaAs light-emitting diodes on Si substrates. *IEEE Photonics Technol. Lett.* **6**, 706–8

Zheng W, Buhlmann P, Jacobs H O 2004 Sequential shape-and-solder-directed self-assembly of functional microsystems. *Proc. Natl. Acad. Sci. USA* **101**, 12814–17

Biographies

Jiandong Fang received his Bachelor's degree in Physics from the University of Science and Technology of China in 1999 and his Master's degree in Physics and the Ph.D. degree in Electrical Engineering from the University of Washington (Seattle, WA) in 2001 and 2006, respectively. From September 2006, he worked as a postdoc research fellow at the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA.

His current research is mainly focused on self-assembly of microcomponents, design, fabrication, and packaging of microelectromechanical systems (MEMS) such as fluidic pumps, scanning mirrors, biochemical sensors, and piezoelectrically driven MEMS resonators.

Karl F. Böhringer received his Dipl.-Inform. degree from the University of Karlsruhe, Germany, in 1990 and his MS/Ph.D. degrees in Computer Science from Cornell University, Ithaca, NY, in 1993/1997.

He was a Visiting Scholar at Stanford University in 1994–1995 and a Postdoctoral Researcher at the University of California, Berkeley from 1996 to 1998. He joined the Electrical Engineering department at the University of Washington in Seattle, WA, in 1998, where he is currently Associate Professor. He also held visiting faculty positions at the Universities of Tohoku, Tokyo, Kyoto (Japan), and São Paulo (Brazil). His research interests include microelectromechanical systems (MEMS), manipulation and assembly from macro- to nanoscales, microfluidic systems for the life sciences, and microrobotics. He has created, among others, multibatch self-assembling systems, massively parallel microactuator arrays, and a walking microrobot.

Karl F. Böhringer is a member of the Society for Nanoscale Science, Computing and Engineering (ISNSCE), the American Society for Engineering Education (ASEE), and the German Society for Information Sciences (GI). He was awarded a Long-Term Invitational Fellowship for Research in Japan by the Japan Society for the Promotion of Science (JSPS) in 2004, an IEEE Robotics & Automation Society Academic Early Career Award in 2004, an NSF CAREER Award in 1999, and an NSF Postdoctoral Associateship in 1997. His work was listed among the Top 100 Science Stories of 2002 in *Discover* magazine. He is an associate editor of ASME/IEEE Journal of Microelectromechanical Systems and IEEE Transactions on Automation Science and Engineering and has served, among others, on technical program committees for the IEEE MEMS and Transducers conferences.

1.15 Packaging

O. Brand, Georgia Institute of Technology, Atlanta, GA, USA

Glossary

ADC Analog to digital converter
ASIC Application specific integrated circuit
BCB Benzocyclobutene
BGA Ball grid array
C4 Controlled collapse chip connection
CBGA Ceramic ball grid array
CC Chip carrier
CMOS Complementary metal oxide semiconductor
CMP Chemical mechanical polishing
COB Chip on board
CPGA Ceramic pin grid array
CQFP Ceramic quad flat package
CSP Chip scale package
CTE Coefficient of thermal expansion
CVD Chemical vapor deposition
DIP Dual in line package
DMD Digital micromirror device
DRIE Deep reactive ion etching
EEPROM Electrically erasable programmable read-only memory
HDTV High definition television
I/O Input/output
IC Integrated circuit
LCC Leadless chip carrier
LCP Liquid crystal polymer
LDCC Leaded chip carrier
LGA Land grid array
LLCC Leadless chip carrier
LTCC Low temperature cofired ceramic
MCM Multi-chip module
MEMS Microelectromechanical system
MLF MicroLeadFrame
MOEMS Microoptoelectromechanical system

OTP One time programmable
PBGA Plastic ball grid array
PCB Printed circuit board
PGA Pin grid array
PLCC Plastic leaded chip carriers
PMMA Polymethylmethacrylate
PPGA Plastic pin grid array
PQFP Plastic quad flat package
QFN Quad flat no leads
QFP Quad flat package
RAM Random access memory
RF Radio frequency
SIP System in package
SIP Single in line package
SMD Surface mount device
SMT Surface mount technology
SOC System on chip
SOI Silicon on insulator
SOIC Small outline integrated circuit
SOP Small outline package
SRAM Static random access memory
SSOP Shrink small outline package
SVGA Super video graphics array
SXGA Super extended graphics array
TAB Tape automated bonding
TC Thermocompression
TO Transistor outline
TQFP Thin quad flat package
TS Thermosonic
TSOP Thin small outline package
UBM Underbump metallization
US Ultrasound
UV Ultraviolet
VOC Volatile organic compound
WLP Wafer level packaging
XGA Extended graphics array

1.15.1 Introduction

Packaging is one of the most important aspects in microsystem development. It is typically a major cost component, regularly exceeding the common 30% for integrated circuit (IC) packages, and ultimately determines the reliability and long-term stability of the microsystem. Therefore, an excellent packaging concept is crucial for a successful microsystem product. Nevertheless, packaging is often considered far too late in the development cycle of a microsystem. A simple search on the engineering and sciences publication database INSPEC reveals roughly 20 000 publications for the keyword search *microsystem* or *MEMS* (microelectromechanical systems) (as of November 2006), but only about 2600 publications or less than 15% if the keywords *package* or *packaging* are added to the above search string. The reader might argue that packaging is the job of the companies, not of academia, and companies generally do not publish their results. This is certainly true, but there is more to it. Packaging research is typically time-consuming, requiring long-term testing, and generally does not yield a publication as easily as a new device or a new fabrication process. Nevertheless, the importance of packaging has been widely acknowledged in recent years and several books (Gilleo 2005, Hsu 2003, Spangler 2007) and book chapters (e.g., Bright *et al.* 2004, Monajemi *et al.* 2007) have been dedicated exclusively to this topic.

The goal of the present chapter is not a comprehensive review of all packaging work done in the microsystems area (this would be a very challenging task, looking at roughly 2500 publications published over the past 25 years), but rather to provide an overview on common packaging concepts. The chapter starts with a summary of microelectronics packaging, highlighting the concept of different packaging levels, chip interconnection techniques, and standard IC packages. Section 1.15.2 concludes with a comparison of microelectronics and microsystems packaging requirements. Section 1.15.3 covers zero-level packaging concepts, i.e., wafer-level packaging (WLP) processes, which encapsulate the fragile MEMS/microsystem components before the dicing step. Section 1.15.4 is dedicated to first-level packages, covering dicing, die attach and interconnect, and various concepts for nonelectrical feedthroughs. This section is enriched with a number of microsystem examples, which utilize ceramic packages, metal packages, plastic packages, or chip-on-board (COB) technologies. Based on the background of the author, the focus is on the packaging of microelectromechanical components and topics like the packaging of optoelectronic components are not covered.

1.15.2 Microelectronics Packaging

Microsystems packaging strongly borrows from designs and processes originally developed for the microelectronics industry. The purpose of this section is to provide a brief overview on chip interconnection technologies and packages used to encapsulate ICs. For more detailed treatments of microelectronics packaging, the reader is referred to one of the many handbooks in this area (e.g., Harper 2004, Lau *et al.* 1995, 1998, 2000, Tummala 2001, Tummala and Rymaszewski 1989, 1997, Ulrich and Brown 2006). It should be noted that the microelectronics packaging industry, just as the microelectronics field in general, is developing fast, with a constant flow of new package types and packaging processes. Therefore, most up-to-date information on the latest developments is likely to be found on the web pages of major microelectronic packaging companies, such as Amkor Technologies (http://www.amkor.com).

1.15.2.1 Packaging Hierarchy

The package for an electronic system has four major functions, namely power distribution, signal distribution, heat dissipation, and circuit protection (Tummala and Rymaszewski 1989). As electronic systems can be rather complex, they typically involve several levels of packaging, often referred to as a packaging hierarchy. Figure 1 illustrates such a packaging hierarchy: the encapsulation of single IC chips is widely considered the first level of packaging. It should be noted that modern first-level packages might not only contain one but several IC chips and are then referred to as multichip modules (MCM). Single-chip modules and MCM are assembled together with other components, such as resistors and capacitors, on multilayer cards. Typically, the components are soldered onto a printed circuit board (PCB), but the cards may also have sockets to plug-in additional components. These cards make up the second level of packaging. In case of a personal computer as an example of an electronic system, second-level packages include the processor card (if not directly mounted onto the motherboard), random access memory (RAM) modules, or a graphics card. Most simple electronic systems typically only comprise first- and second-level packages, i.e., consist of a PCB with attached, encapsulated ICs and passive components. In the example of the personal computer, the motherboard, again a multilayer circuit board, constitutes the third level of packaging, providing sockets for various cards to be plugged in, but also incorporating single-chip packages and other components.

Not surprisingly, the traditional packaging hierarchy is not hammered into stone and many variations are found in today's microelectronic

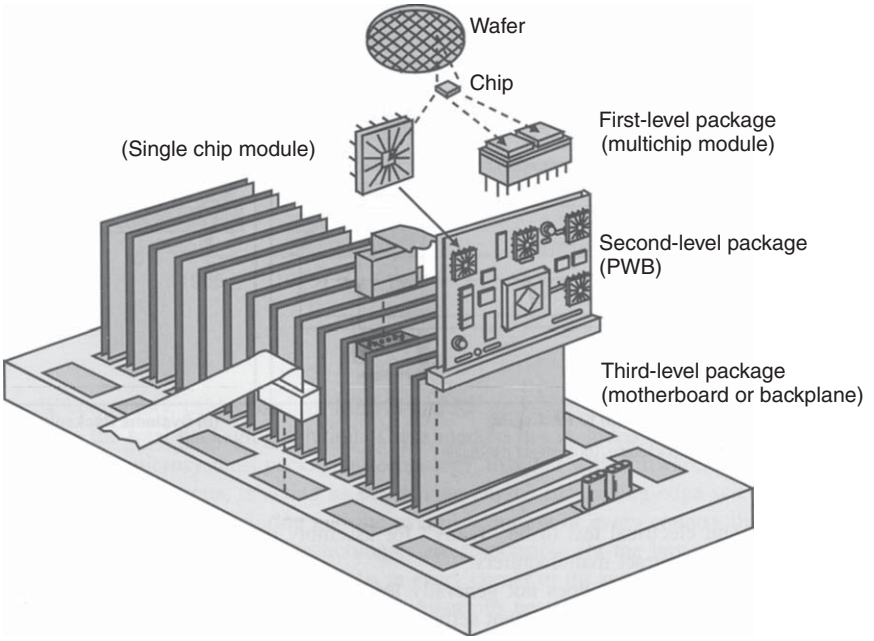


Figure 1 Schematic of microelectronic packaging hierarchy with first-, second-, and third-level packages. (Source: Tummala R R 2001 *Fundamentals of Microsystems Packaging*. McGraw-Hill, NewYork. Copyright: The McGraw-Hill Companies, Inc.)

systems to accommodate the customer's needs. For low-cost consumer applications, bare silicon chips are directly mounted onto PCBs, effectively skipping the first-level package. On the other hand, systems on a chip (SOC) or systems in a package (SIP) refer to systems where all necessary components for their operation are incorporated onto a single chip and into a single package, respectively (Brand 2006), thus combining first- and second-level packaging functions into a first-level package.

To protect the fragile mechanical structures of today's MEMS, special fabrication processes have been developed to cap these parts on a wafer level. This is often referred to as zero-level packaging and will be discussed in detail in Section 1.15.3. For the remainder of this chapter, we are now focusing on zero- and first-level packages.

1.15.2.2 Chip Interconnection Techniques

Electrical connections between chip and package are commonly performed by wire bonding, flip chip bonding, or tape-automated bonding (TAB) (Baldwin 2001). Even though wire bonding is a serial process, i.e., one interconnect is formed after the other, it is still today the most common chip interconnect technology. In case of flip chip bonding and TAB, all interconnects are formed at once (referred to as gang bonding), which constitutes a tremendous advantage for microsystems with very large input/output (I/O) count. To understand the advantages and disadvantages of these three chip interconnection technologies for the purpose of microsystem packaging, they are briefly reviewed in the following subsections.

Comparing the three chip interconnection technologies (**Figure 2**), the main advantages of wire bonding are its wide installation base and its process flexibility. With current bond speeds (>15 bonds s^{-1}), the wire bonding process is cost competitive even for chips with several hundred I/Os (Baldwin 2001). One of the key advantages of the flip chip technology are the better electrical characteristics of the interconnects with lower interconnect resistances, capacitances, and inductances. This can be decisive for high-frequency applications including radiofrequency (RF)-MEMS. Flip chip packages can be made smaller than wire-bonded packages, because no lateral space is needed for the fanout of the wires and no extra vertical space is needed to accommodate the wire loops.

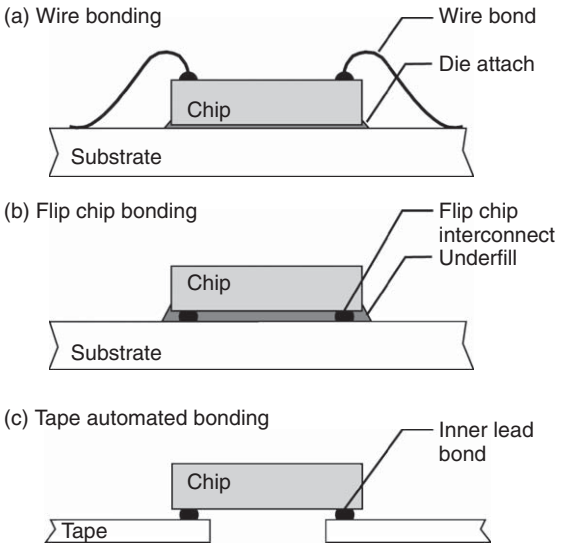


Figure 2 Comparison of chip interconnection technologies: (a) wire bonding, (b) flip chip bonding and (c) tape-automated bonding.

1.15.2.2.1 Wire bonding

Wire bonding always requires a preceding die attachment process to mount the die onto the substrate, such as a leadframe or multilayer substrate. One distinguishes polymer die bonding, solder die bonding, and glass die bonding (Tummala and Rymaszewski 1989, 1997), with polymer die bonding being the most common method. In case of polymer die bonding, an adhesive, commonly an epoxy, a polyimide or a silicone, is used to attach the chip to the substrate. The characteristics of the adhesive, namely its rheological, mechanical, electrical, and thermal properties, can be controlled by the polymer matrix itself, but also by adding filler materials, such as gold, silver, alumina (Al_2O_3), or boron nitride particles. The die attach process is typically highly automated: a die bonder feeds the substrates, dispenses the die attach adhesive, picks the chip from a carrier, and attaches it to the substrate. Subsequently, the polymer is cured and the assembly is ready for wire bonding. The advantages of polymer die bonding are its flexibility with respect to material properties, the low process temperature, and the relatively low cost. Many MEMS are sensitive to packaging-induced stresses and soft die attach adhesives are widely used to minimize the stress acting on the sensor chip (see Section 1.15.4). Solder and glass die bonding generally form a more rigid interface between chip and substrate and typically require

higher process temperatures. In solder die bonding, different solder systems, such as AuSi (eutectic), AuGe, PbSn, AuSn, and PbAgIn, are used to attach the chip to the substrate. The solder can be applied by dispensing from a hot nozzle, by dispensing and printing solder pastes or by applying solder performs. Typical process temperatures range from 200°C to 400°C. Solder die attach provides an excellent thermal conductivity, high-temperature stability in case of hard solders, and is typically used in conjunction with metal and ceramic packages. Glass die bonding uses a glass frit material, which is either screen/stencil printed or dispensed onto the substrate (see also glass frit bonding in Section 1.15.3.1). Typical process temperatures are around 400°C, which limits its application mainly to ceramic packages.

After the die attach, thin metal wires are bonded, one after the other, from the chip's bond pads to the substrate pads. One distinguishes two wire bonding sequences, ball–wedge bonding and wedge–wedge bonding, and three basic bonding processes, ultrasonic (US), thermosonic (TS), and thermocompression (TC) bonding (Baldwin 2001). In case of ball–wedge bonding, the bonding wire is fed through a capillary; in case of wedge–wedge bonding it is fed through a bonding wedge. The ball–wedge bonding sequence (**Figure 3(a)**) starts with the formation of a ball at the end of the bonding wire, e.g., with a high-voltage electric discharge (electronic flame off). The first bond of a bonding cycle is achieved by pressing the ball onto the bonding pad on the chip surface, while applying either heat only (TC bonding) or a combination of heat and US energy (TS bonding). This initial bond is referred to as the ball bond. Along a programmed trajectory, the wire is now looped to the location of the second bond on the substrate (loop formation). The second bond, a wedge bond, is performed by pressing the capillary onto the package pad, applying either thermal energy only or thermal energy and US excitation to assist bond formation. While raising the bonding tool after the second bond, the wire tail is

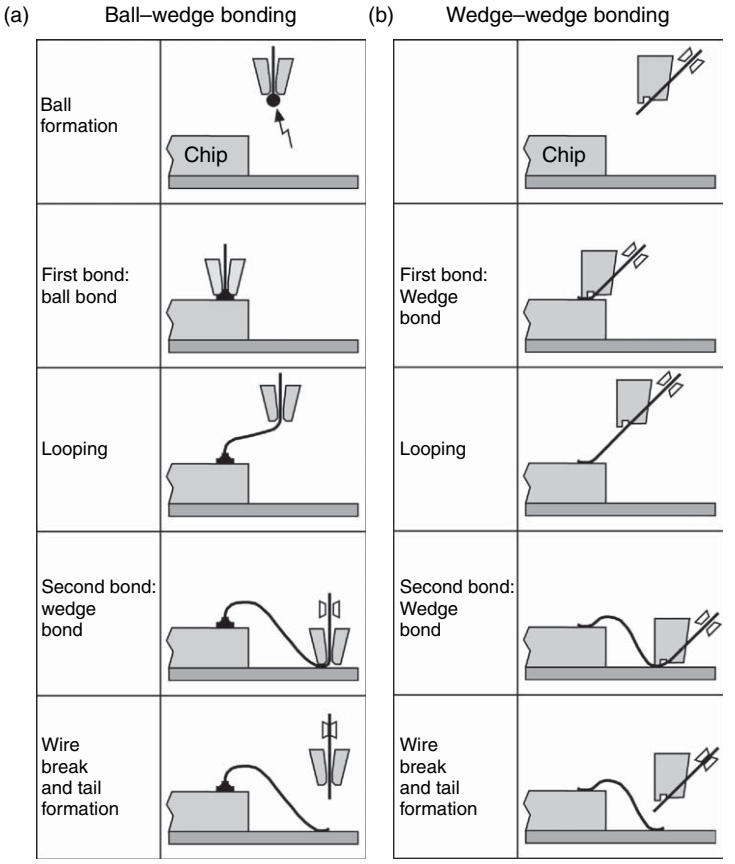


Figure 3 Schematics of (a) ball–wedge and (b) wedge–wedge bonding cycles; for clarity, only the tips of the bonding tools and a schematic of the wire clamp are shown.

broken off and the next bonding cycle can start with the formation of a new ball. Because of the circular symmetry of the first ball bond and the bonding capillary, the bond head can move in any direction during loop formation, i.e., no rotation of the substrate holder or bond head is required to bond all the wires along the perimeter of a chip. This allows for very high bond speeds and automatic ball–wedge bonders can execute between 10 and 20 bond cycles per second (see, e.g., Oerlikon ESEC Semiconductor, Cham, Switzerland, http://www.esec.com). Gold wires with a typical diameter of 25 µm are widely used for ball–wedge bonding. In fine-pitch applications, 15-µm gold wires enable high-speed wire bonding with a pad pitch of only 35 µm. To be able to bond ICs with advanced copper metallization, ball–wedge bonding processes based on copper wires have been developed in recent years.

In case of wedge–wedge bonding, both bonds of a bonding cycle are wedge bonds (**Figure 3(b)**). In contrast to the ball–wedge bonding, the wedge–wedge process is unidirectional, i.e., the direction of the looping must be along the centerline of symmetry of the first bond. This generally requires the rotation of the substrate holder for each wire bond. Using aluminum wires, wedge–wedge bonding can be done at room temperature by applying a vertical bond force and US energy only (US bonding). However, TS and TC bonding are possible as well. Because of the dimensions of the first wedge bond, wedge–wedge bonding generally enables smaller bonding pitches. Wedge–wedge bonding is especially popular in laboratory and prototyping environments, where manual bonders enable unsurpassed process flexibility in terms of substrate choice, pad layout, and wire material. In microsystem applications, gold wires are often preferred over aluminum wires because of their superior corrosion resistance.

1.15.2.2.2 Flip chip bonding

In flip chip bonding, the chip is flipped before being attached, and solder or conductive polymer bumps between chip and substrate serve as both an electrical and a mechanical interconnection. The original technology, known as Controlled Collapse Chip Connection or C4, has been demonstrated in the late 1960s by IBM. Main drivers for this process are ultrahigh I/O counts and the ever-increasing electrical performance requirements for high-frequency applications. Today, customers can choose between a growing number of different flip chip interconnection systems depending on their particular application (Lau 1995, 2000, Oppermann *et al.* 2000). Flip chip services are offered by a number of companies, including FlipChip International (http://www.flipchip.com) and PacTech (http://www.pactech.de). Flip chip connections can be made to ceramic and polymer substrates using either solder or adhesive interconnection systems. The basic flip chip interconnection can be subdivided into four functional areas (Baldwin 2001):

(1) the underbump metallization (UBM),
(2) the IC bump and bond material,
(3) the substrate metallization, and
(4) the encapsulant.

The UBM is a compatible metal sandwich connecting the chip metallization (usually aluminum or copper) to the bump metallization, and generally consists of an adhesion layer, a barrier layer, a wetting layer, and an antioxidation barrier. Depending on the application and the actual interconnection system, various UBMs are available. For solder systems, an electroless nickel–gold UBM is a frequently used low-cost option: the aluminum metallization of the chip is activated with a zincate layer (adhesion layer); subsequently, a nickel bump (typically 5 µm) is formed by electroless plating (barrier and wetting layer), and finally, a thin gold film (typically 50 nm) is deposited as an antioxidation barrier. In case of a solder interconnect system, the IC bump and bond material consist of a controlled amount of solder per interconnect, deposited by evaporation, electroplating, stencil printing, or solder jetting. Alternatively, the solder balls can be placed and reflown with a Laser Solder Ball Jet System (see PacTech, Nauen, Germany; http://www.pactech.com). The chosen solder system again depends on the application: if the flip chip packaged die requires withstanding a subsequent soldering step (e.g., in case of a surface-mountable package with a flip chip die), a higher melting temperature solder is needed for the flip chip attachment. If, on the other hand, the flip chip die is mounted onto a laminate substrate together with other surface-mount device (SMD) components, a low melting temperature solder might be preferable to form all interconnects on the board simultaneously.

For adhesive flip chip interconnection systems, nonoxidizing gold is widely used as bump material and combined with, e.g., a titanium–tungsten UBM. Thereby, the gold bump is either electroplated or provided by a ball wire bonder in the form of a so-called stud bump (i.e., a ball bond with short wire tail). The adhesives are either isotropically or anisotropically conductive. Isotropic conductive adhesives, such

as epoxies filled with silver flake particles, are printed or dip-transferred to either the IC bumps or the substrate bond pads. After assembly and curing, the isotropically conducting adhesive forms a mechanical bond between the substrate and the IC bump. This is in contrast to the generally more reliable solder system, where a metallurgical bond is formed. Anisotropic conducting adhesives consist of a polymer matrix with embedded conducting spheres and are applied not only to the contacts but also to the complete chip area as a paste or a film. While pressing the chip onto the substrate, the conducting spheres are compressed and captured between the IC bump and the substrate pad, making a reliable interconnect upon curing the polymer. The advantage of anisotropic conductive adhesives is that they serve as an underfill at the same time. Most of the other flip chip interconnection systems require the application of a polymer underfill material between chip and substrate to reduce the thermomechanical stress acting on the interconnects, caused by differences in thermal expansion coefficient between chip and substrate. The underfill is dispensed along one or several edges of the chip and drawn into the volume between chip and substrate by capillary forces (Baldwin 2001).

1.15.2.2.3 Tape-automated bonding

TAB is an interconnection process based on mounting and bonding an IC to a flexible conductor tape, either a single-layer metallic tape or a multilayer polymer/metal sandwich. Like in flip chip bonding, all interconnects are formed at the same time (gang bonding). To prepare for the bonding process, the bond pads on the chip are coated with a UBM and a metal bump, e.g., an electroplated gold bump. Alternatively, the tape instead of the chip can be bumped. The so-called inner lead bonding forms the actual interconnects between inner metal leads on the tape and the chip, using either TC (e.g., Au–Au) or solder bonding (e.g., eutectic Au–Sn) as the bonding process. After the inner lead bonding, the chip and interconnects are encapsulated by dispensing and curing a polymer, followed by an optional burn-in and test. Before assembly, the TAB packages are singulated and the outer leads are formed. Finally, the outer leads are boded to the substrate (outer lead bonding).

1.15.2.3 Standard IC Packages

The goal of this section is to provide an overview of first-level packages for ICs. Keeping in mind that the overall chapter concentrates on microsystem packaging, the focus is more directed toward low pin count packages. Single-chip packages can be categorized by a number of criteria, such as package material (plastic, ceramic, or metal), interconnect arrangement (peripheral or area array interconnects), assembly technology (through-hole or surface-mount technology (SMT)), or package type. Thereby, an industry-wide body, the Joint Electronic Devices Engineering Council (JEDEC, http://www.jedec.org), establishes standard package designators and geometries. In the following, the basic single-chip package types (**Figure 4**) are briefly discussed (Tummala 2001, Tummala *et al.* 1997).

Single in line packages: Single in line packages (SIPs) are less common than the dual in line packages (DIPs) and have pins for through-hole assembly only

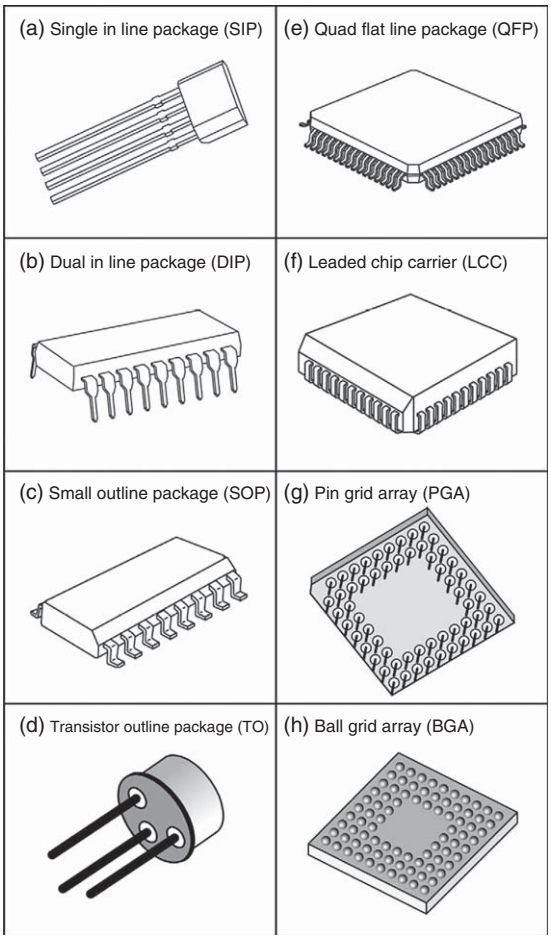


Figure 4 Overview of common single-chip package types used for integrated circuit packaging. (Images (a), (b), (c), (e), and (f) courtesy of: Allegro® MicroSystems Inc., Worcester, MA, USA.)

along one side of the package. SIPs are generally plastic packages with a pin count up to 48 and a pin pitch of 2.54 mm.

Dual in line packages: DIPs come in either plastic or ceramic versions and have two rows of interconnects along two opposite sides of the package. The DIP is a commodity through-hole package with up to 84 pins and a most common lead pitch of 2.54 mm (1.78 and 1.27 mm are less common). Ceramic DIPs are frequently used for testing MEMS/microsystem prototypes.

Small outline packages: The small outline package (SOP) is the most common package type for low pin count ICs. It is a leadframe-based plastic package with leads on two sides of the body and pin counts ranging from 8 to 84. The original lead pitch is 1.27 mm (SOP or SOIC (small outline integrated circuit)), but shrink small outline packages (SSOP) with smaller pitches of 0.8, 0.635, and 0.5 mm are available as well. Fully encapsulated SOP with a total package height of less than 1 mm are often termed thin small outline packages (TSOP).

Quad flat packages: Quad flat packages (QFPs) are available as ceramic (CQFP) or leadframe-based molded plastic (PQFP) packages with up to 356 peripheral leads. The surface-mountable QFP has leads on all four sides of the package and lead pitches ranging from 1 mm down to 0.3 mm. QFPs with overall thickness equal or below 1 mm are termed thin QFP or TQFP.

Chip carriers: Chip carriers (CCs) are, like QFPs, surface-mountable packages with peripheral leads on all four sides of the body. In contrast to the QFP with its gullwing leads, CCs feature either J-leads or no leads at all. Leadless chip carriers (LCCs or LLCCs) have a laminated ceramic substrate, leaded chip carriers (LDCCs) with J-leads are either leadframe-based molded plastic packages (PLCCs) or utilize ceramic substrates. CCs are often mounted in dedicated sockets and have I/O counts of less than 120 and typical interconnect pitches of 1.27 and 1.0 mm.

Pin grid arrays: Pin grid arrays (PGAs) are through-hole packages with an area array of interconnect pins using either laminated ceramic (CPGA) or laminated plastic (PPGA) substrates. Pin counts up to 750 are available with pin pitch of either 2.54 or 1.27 mm. CPGA packages are available in cavity-up or cavity-down configurations, i.e., the cavity for the IC is on the opposite or the same side, respectively, as the pin array. Several generations of Intel microprocessors have been packaged using PGAs. More recent versions use a land grid array (LGA), a package similar to the PGA (and BGA (ball grid array)), but having interconnect pads instead of pins.

Ball grid arrays: BGAs are the surface-mountable versions of the PGA with an area array of interconnects with attached solder balls. Plastic BGAs (PBGAs) based on laminated multilayer plastic substrates and ceramic BGAs (CBGAs) based on multilayer ceramic substrates are available. The chip interconnection is accomplished either by wire bonding or by flip chip bonding. Conventional BGAs have a solder ball pitch ranging from 1.0 to 1.5 mm and I/O counts up to 1100 (2600 for flip chip PBGA). Novel chip scale versions of BGA packages, such as Tessera's μBGA (http://www.tessera.com), feature ball pitches down to 0.5 mm, but typically have smaller I/O counts.

Single-chip package development in recent years is driven on one hand by higher and higher I/O counts, demonstrated by flip chip BGA packages with up to 2600 interconnects (see, e.g., Amkor Technologies, http://www.amkor.com), and on the other hand by smaller and smaller package form factors, witnessed by the growth in the area of chip scale packages (CSP). A characteristic of a CSP is a package footprint that is a maximum of 50% larger than the die area and a package perimeter no more than 20% larger than the chip perimeter. An example is the MicroLeadFrame (MLF) package developed by Amkor, a leadframe-based QFP with solderable lands instead of leads, thus also called quad flat no leads (QFNs). Single-chip packages are continuously developing and most up-to-date information on the newest CSP and other packages is found on the packaging manufacturer's and packaging service provider's web pages. The ultimate CSP will have the size of the actual die it protects. In this framework, considerable efforts in the area of WLP should be mentioned (see, e.g., Garrou and Tummala 2001). In WLP, some or all of the packaging steps, such as interconnect formation or encapsulation, are performed on a wafer level, before the singulation of the wafer into individual ICs. Section 1.15.3 introduces similar WLP concepts for microsystems and MEMS.

1.15.2.4 Microelectronics vs. Microsystem Packaging

The fabrication of microsystems and MEMS strongly relies on process steps borrowed from the IC industry. Can microsystem/MEMS packaging benefit in a similar way from developments in the microelectronics packaging area? The answer is Yes and No. Yes, because microsystem packaging indeed uses the same

chip interconnection techniques as the microelectronics industry to establish electrical interconnects and employs microelectronics packages wherever feasible. On the other hand, No, because a suitable microsystem/MEMS package usually requires more than protection, power distribution, signal distribution, and heat dissipation. In the following, the differences between microelectronic and microsystem packaging are briefly summarized.

One of the key features of microsystems is that they generally interact with their environment not only via the electrical energy domain. Microsensors need to sense the quantity of interest in the environment; microactuators generally need to transduce nonelectrical information through the package. As a result, a microsystem/MEMS package often requires not only electrical interconnects, but also feedthroughs for nonelectrical signals. At the same time, the package must still protect the microsystem from harmful environmental influences. The need for nonelectrical feedthroughs (or windows to the outside world) is fundamentally different from microelectronics packaging requirements and requires, depending on the applications, radically new packaging concepts. How can we sense a pressure inside a package? How can we transmit optical information in and out of the package? How can we detect chemical and/or biological analytes in air samples or, even more challenging, in a liquid sample and still protect our fragile MEMS structure? Packaging concepts developed in response to these and similar questions will be discussed in Section 1.15.4.

Microsystems/MEMS often comprise movable microstructures that cannot be encapsulated with a molding compound, the most common microelectronics packaging technique. In contrast, they need packages providing cavities while still protecting sensitive IC components and chip interconnect structures, such as wire bonds and bond pads. Frequently, MEMS structures also need extra protection during the wafer dicing process (see Section 1.15.4.1), prompting the development of wafer-level encapsulation techniques as discussed in Section 1.15.3. Moreover, microsystems/MEMS are usually more sensitive to mechanical stresses than are ICs and the influence of packaging-induced thermomechanical stresses must be carefully considered. Examples are piezoresistive pressure sensors, where packaging-induced stresses acting on the piezoresistive sensing elements can cause a sensor offset.

Yet another critical area influencing packaging cost are the manufacturing volumes of typical microsystem packages compared to microelectronics packages. Even though a number of very high-volume products exist (e.g., accelerometers, pressure sensors, inkjet print heads), most microsystem products have substantially smaller volumes or require packaging solutions that differ from customer to customer for a single product. An example is MEMS pressure sensors, where a given pressure sensor die might be packaged as an absolute or a differential pressure sensor, in a media-compatible package for operation in harsh environments or in a low-cost package with silicon gel for a consumer application. This calls for modular packaging approaches, which make use of basic packaging components for as many applications as possible.

Finally, the areas of assembly and testing/calibration cannot be emphasized enough. Single-chip packages are mounted onto PCBs in a highly automated fashion using pick-and-place systems. In order to benefit from these automated processes, microsystems packages must be handled by standard or customized pick-and-place tools and must be able to withstand standard assembly processes, such as a solder reflow. Otherwise, components must be assembled by hand, drastically increasing manufacturing cost. While the electrical testing of a microsystem might not be as involved as the electrical testing of an advanced IC, microsystem testing and calibration often involves other energy domains as well. Pressure sensors might be calibrated at different temperatures and pressures, requiring the development of dedicated testing equipment rather than the use of standardized off-the-shelf components.

1.15.3 Zero-Level Packaging for Microsystems

As mentioned earlier, various zero-level packaging technologies have been developed over the past decades to encapsulate microsystem/MEMS structures on a wafer level. Besides providing an obvious cost advantage by batch processing wafers rather than serial processing individual chips, the zero-level packaging steps offer two main advantages from a reliability point of view:

(1) The MEMS structures are (vacuum-) encapsulated in a clean room environment minimizing particle contamination that could lead to device failure.
(2) The fragile MEMS structures are protected by a capsule from water jets and sawing debris during the wafer dicing step.

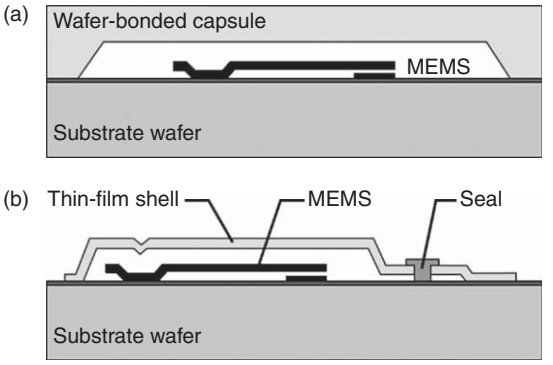


Figure 5 Zero-level packaging with (a) bonded capsule or (b) thin-film shell. (Source: Najafi K 2003 Micropackaging technologies for integrated microsystems: Applications to MEMS and MOEMS. *Proc. SPIE* **4979**, 1–19. Adapted with permission.)

In the following section, two basic zero-level packaging approaches (**Figure 5**), namely encapsulation by wafer bonding and encapsulation by thin-film shells, will be discussed in some detail. For additional information on zero-level packaging processes for MEMS/microsystems, the reader is referred to one of the review articles or books in this area (e.g., Hsu 2003, Iyer and Auberton-Herve 2002, Ko *et al.* 1985, Lin 2000, Najafi 2003, Schmidt 1998).

1.15.3.1 Encapsulation by Wafer Bonding

Wafer bonding refers to process technologies developed to bond two or more wafers to each other (Alex and Gösele 2004, Iyer and Auberton-Herve 2002, Schmidt 1998). Besides being used for zero-level packaging purposes, wafer bonding is used in the microsystems/MEMS field to build 3D microstructures, such as microturbine engines (Frechette *et al.* 2005), and to manufacture silicon-on-insulator (SOI) substrates. Wafer bonding techniques are not only used to bond two silicon wafers to each other, but can be employed for various substrates, including glass wafers or nonsilicon semiconductor (e.g., GaAs) wafers. The wafer bonding processes can be subdivided into direct bonding techniques, such as fusion and anodic bonding, and processes requiring an intermediate layer, a metal, dielectric, or adhesive film, for bond formation. Wafer bonding technologies based on intermediate metal films include metal TC bonding, solder bonding, and diffusion bonding; wafer bonding based on intermediate dielectric films include glass frit bonding. In the following, wafer bonding techniques pertinent to zero-level packaging are briefly introduced. For a more detailed treatment of bonding technologies, the reader is referred to Chapter 1.09. The characteristics of the most common wafer bonding technologies are summarized in **Table 5** of Chapter 1.09.

Direct bonding techniques typically require mirror-smooth wafer surfaces, while an intermediate bonding layer can level out surface topography resulting from, e.g., underlying electrical feedthroughs. The traditional silicon fusion bonding requires annealing temperatures around 1000°C, whereas most other bonding techniques have peak temperatures around or below 400°C and are, thus, compatible with standard complementary metal oxide semiconductor (CMOS) metallizations. Even silicon fusion bonding can be performed at CMOS-compatible temperatures by proper surface treatment before the bonding. The lowest bonding temperatures are typically achieved using adhesives as bonding film, but those bonds are not considered hermetic. State-of-the-art automated and manual wafer bonding tools are commercially available from Karl Suss (http://www.suss.com), EV Group (http://www.evgroup.com), and OAI (http://www.oainet.com). These bonding systems can handle different bonding processes and wafer diameters, and typically include a cleaning station, an alignments and a bonding tool.

Except adhesive bonding, all of the listed wafer bonding methods are considered hermetic and can potentially be used for vacuum packaging. A challenge for every vacuum package, however, is that gases are freed during the bonding process and desorption of gas molecules from the cavity walls takes place over time. Because of the relatively large surface-to-volume ratio of most microfabricated cavities, this outgassing can considerably increase the cavity pressure over time. This might not be tolerable for resonators requiring vacuum as low as 10^{-3} mbar for stable operation. A solution to this problem is getter films, which can be mounted or deposited inside the cavities and effectively getter residual gas molecules, thus, maintaining a constant cavity pressure over time (Moraja and Amiotti 2003, Sparks *et al.* 2003, 2004).

1.15.3.1.1 Fusion bonding

Fusion bonding or direct bonding (Alexe and Gösele 2004, Gösele and Tong 1998, Gösele *et al.* 1999, Schmidt 1998) is based on the fact that two mirror-flat, clean wafer surfaces will adhere to each other even at room temperature if brought into contact. While fusion bonding is mainly associated with the bonding of two silicon wafers, almost any material surfaces can be directly bonded if their surfaces are

sufficiently smooth. The room-temperature bond is, thereby, attributed to attractive van der Waals forces. Currently, the most important applications of silicon fusion bonding are in the manufacturing of SOI substrates and in the fabrication of silicon-based microsystems.

Silicon fusion bonding requires very smooth wafer surfaces with a typical surface roughness of better than 1 nm and a wafer bow of less than 5 μm on a 4″ wafer (Schmidt 1998). While prime-grade, polished wafer surfaces are generally sufficiently smooth for fusion-bonding applications, other surfaces can be prepared by chemical–mechanical polishing (CMP) steps. The fusion-bonding sequence comprises three process steps:

(1) wafer cleaning and surface conditioning,
(2) alignment and room-temperature bonding, and
(3) high-temperature anneal to improve bond strength.

Thorough wafer cleaning is essential to prevent nonbonded areas separated by small particles. The surface treatment generates either hydrophilic or hydrophobic wafer surfaces. Hydrophilic wafer surfaces are obtained either by growing a thermal oxide or by using the native oxide in combination with a proper surface treatment (e.g., a RCA cleaning procedure); the wafer surfaces are then terminated with –OH groups. If the surface oxide is removed by a dip in hydrofluoric acid, a hydrophobic surface with –H coverage is achieved (Gösele and Tong 1998). After surface preparation, the wafers are aligned in a clean room environment to reduce particle contamination and contacted by gently pressing them together at a central point. This way, the wafers will prebond without trapping air between the surfaces. Typical surface energies after the room-temperature prebond are 0.02 and 0.1 $J\,m^{-2}$ in the case of hydrophobic and hydrophilic silicon surfaces, respectively (Gösele and Tong 1998, Gösele *et al.* 1999). By annealing the wafer sandwich to temperatures between 800°C and 1200°C, the bond strength can be increased in both cases by one to two orders of magnitude with typical surface energies of 2–3 $J\,m^{-2}$. The surface chemistry involved in forming the bond at elevated temperature is described in detail in Gösele and Tong (1998). It is worth noting that hydrated wafer surfaces reach bonding energies >1 $J\,m^{-2}$ already with prolonged annealing steps at temperatures as low as 200°C (Gösele and Tong 1998, Gösele *et al.* 1999).

Considerable research activities are centered on reducing the required annealing temperature for silicon fusion bonding. Most approaches involve special pretreatment of the wafer surfaces, either using wet chemistries or by an appropriate surface plasma treatment (e.g., Berthold and Vellekoop 1997, Dragoi *et al.* 2005, Pelzer *et al.* 2004, Suni *et al.* 2006, Tong *et al.* 2003). While bonding temperatures around 200°C make silicon fusion bonding already compatible with IC processes and common surface metallizations, the ultimate goal is a reliable room-temperature bonding process. One of the main advantages of fusion bonding is its applicability not only to silicon surfaces, but to smooth material surfaces in general. This way, material combinations that cannot be epitaxially grown on top of each other (e.g., GaAs/Si or InP/Si) can be joined by wafer bonding. The main disadvantage of fusion bonding is the required surface smoothness and cleanliness.

One of the first applications of silicon fusion bonding in the microsystems area was the formation of reference pressure cavities for pressure sensors (Petersen *et al.* 1988). The proposed process involved etching of a reference cavity in one wafer and fusion bonding of a second wafer to seal the cavities. By thinning the second wafer, a membrane with desired thickness can be formed on top of the buried cavity. Piezoresistive sensing elements can then be implanted to make a pressure sensor. Compared to the conventional method of releasing the pressure-sensitive membrane with an anisotropic wet etching step from the back of the wafer, the fusion-bonding process can result in significantly smaller die sizes. Today, the SM5420 series pressure sensors of Silicon Microstructures (http://www.si-micro.com) utilize a similar technology to manufacture ultrasmall pressure sensor chips (Datasheet SM5420 Low-Cost, SO-8 Packaged Pressure Sensor, Silicon Microstructures Inc., Milpitas, CA, 2005). The underlying pressure sensor die SM5108 has a size of only 0.65 mm × 0.65 mm, allowing 24 000 die to fit on a 6″ wafer (Datasheet SM5108 Ultra-Small, Low-Cost, OEM Pressure Die, Silicon Microstructures Inc., Milpitas, CA, 2005). Other applications requiring sandwiches of several fusion-bonded silicon wafers include the above-mentioned microfabricated air turbine (Frechette *et al.* 2005) and microfabricated magnetic induction machines (Arnold *et al.* 2006).

1.15.3.1.2 Anodic bonding

Anodic bonding between a metal/semiconductor and a glass substrate was first demonstrated in 1969 (Wallis and Pomerantz 1969) and is today the most

widely used wafer bonding technique in the microsystems area. In anodic bonding, a silicon wafer is usually bonded to a sodium-rich glass wafer by applying voltages in the range 200–1500 V between the two wafers at temperatures in the range of 300–450°C (Najafi 2003, Obermeier 1995, Schmidt 1998). Thereby, the glass wafer is biased as the cathode, the silicon wafer as the anode. Under the applied electric field and elevated temperature, mobile sodium ions (Na^+) drift away from the bond interface toward the cathode, leaving behind fixed negative charges that form, together with mirror changes on the silicon side, a very strong electric field across the bond interface. The resulting electrostatic forces pull the wafers together and a strong chemical bond is formed at the elevated temperature (Schmidt 1998). A prerequisite for a good bond quality are clean and smooth wafer surfaces with a surface roughness of typically less than 1 µm. Because two different substrate materials are joined at an elevated temperature, the thermal expansion coefficient of the glass wafer must be carefully chosen in order to minimize mechanical stresses. For this reason, Pyrex 7740 of Corning Glass Works (http://www.slam.corning.com) and Borofloat 33 from Schott (http://www.schott.com) are widely used glass substrates with a room-temperature coefficient of thermal expansion of 3.3 ppm K^{-1} (compared to 2.6 ppm K^{-1} for silicon). Both substrates are available in different thicknesses and wafer sizes to match customer needs.

The most important commercial application of silicon-to-glass anodic wafer bonding is arguably in the area of piezoresistive pressure sensors. As an example, **Figure 6** shows a photograph of a Bosch pressure sensor, bonded to a glass constraint wafer and packaged in a transistor outline (TO) package (Marek 1999). The bonded glass wafer mainly serves two purposes:

(1) The formation of a vacuum reference cavity in case of absolute pressure sensors.
(2) The reduction of packaging-induced stresses acting on the piezoresistive sensing elements by increasing the physical distance from the sensor elements to the die bond area and package materials.

Triple-stack wafers with two Pyrex wafers bonded to a central silicon wafer (Grelland *et al.* 1999) and two silicon wafers bonded to a central Pyrex wafer (Despont *et al.* 1996) can be assembled as well.

Anodic bonding is not only possible between a silicon wafer and a glass wafer, but can be used to bond two silicon wafers with a sputtered or evaporated glass film in between (Berenschot *et al.* 1994, Obermeier 1995, Weichel *et al.* 1998). Moreover, Pyrex glass caps have been anodically bonded to passivated microsystem wafers with polysilicon bond rings (Ziaie *et al.* 1996). By reflowing the passivation layers underneath the bond ring, sufficiently smooth surfaces even over electric feedthroughs can be provided. To make the anodic bond less sensitive to the surface topography, a silicon wafer can be bonded to a glass wafer using a gold film in between (Harpster and Najafi 2003). At temperatures above the eutectic temperature of 363°C of the silicon–gold system, a liquid eutectic phase forms between the silicon and glass wafers. The electrostatic force pulls the wafers together and the liquid phase ensures that any surface topography is planarized. Because the Pyrex glass is transparent and biocompatible, the bonding technique is suitable for optical and biomedical applications. Finally, void-free glass-to-glass bonds have been achieved at temperatures as low as 375°C using anodic bonding (Wei *et al.* 2003).

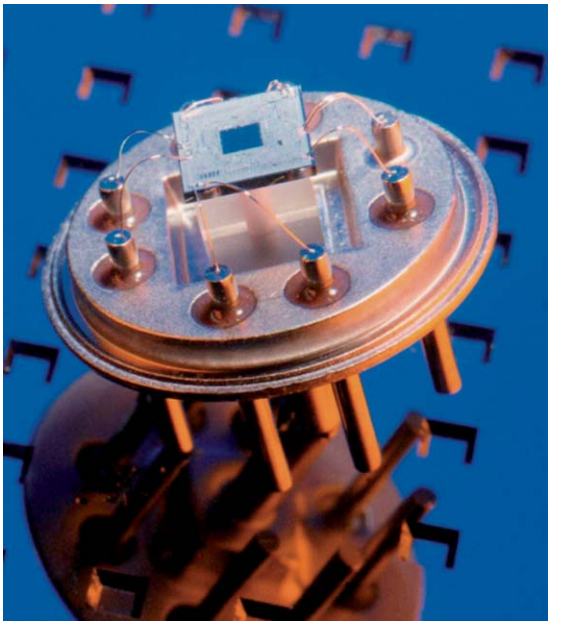


Figure 6 Piezoresistive pressure sensor by Robert Bosch GmbH; the sensor chip is anodically bonded to a Pyrex substrate and mounted in a transistor outline (TO) header. (Photo courtesy: Robert Bosch GmbH, Reutlingen, Germany.)

1.15.3.1.3 Glass frit bonding

Glass frit bonding (Knechtel 2005, Knechtel *et al.* 2006) is mainly used to join two silicon wafers, but can be applied to a wide range of surface materials

commonly used in microsystems technology, including silicon dioxide, silicon nitride, or aluminum. Glass frit bonding is based on glasses with low glass transition temperatures. Widely used are lead-containing glasses with process temperatures below 450°C. Typically, the bonding layers are applied by screen/stencil printing techniques to the capping wafers. Alternatively, thin films can be deposited by sputtering. The screen-printable pastes have the low melting glass powder embedded in an organic binder. Fillers can be incorporated into the paste to adjust the thermal expansion coefficient of the glass frit closer to the substrate materials. After applying the paste to one of the two substrates, typically the cap wafer, the glass layer is conditioned, i.e., the organic binder is burned out and the glass film is premolten to form a compact layer without gas inclusions. Subsequently, the two wafers are aligned to each other and the bond is performed in a bonding tool by applying pressure, heating the wafer pair above the glass transition temperature, and finally cooling it down under applied pressure. Because the glass frits are applied by thick-film technologies, minimal line width and spacing are of the order of 200 and 100 µm, respectively, with typical film thicknesses (after preconditioning) of micrometers (Knechtel 2005). The flow properties of the viscous glass films enable to cover steps from underlying electrical feedthroughs, while still providing a hermetic seal. A disadvantage caused by the screen printing of the glass frit is that film thickness and uniformity over the wafer are difficult to control.

Glass frit wafer bonding is commercially used, e.g., by Freescale Semiconductor (http://www.freescale.com), to attach a silicon constraint wafer to their piezoresistive pressure sensors (schematic in **Figure 7(a)**) and a protective cap wafer onto their surface-micromachined accelerometers (**Figure 7(b)**) (Frank *et al.* 1994). The advantage of using a silicon constraint instead of an anodically bonded Pyrex wafer is the better matching of the thermal expansion coefficient of the two substrates. Depending on the application in absolute or differential pressure sensors, the silicon constraint wafer seals a hermetic cavity or features an etched pressure port, respectively. In the case of accelerometers, the bonded silicon cap protects the fragile MEMS structures during wafer dicing and allows for complete encapsulation of the MEMS chip with a molding compound (see Section 1.15.4.4) More recently, Analog Devices has demonstrated a wafer-level encapsulation approach for their inertial sensors based on glass frit bonding (Felton *et al.* 2003, Yun *et al.* 2005). Their latest packaging technology provides a true wafer-level package with through-wafer interconnects and solder balls for direct flip chip attachment to a printed wiring board (Yun *et al.* 2005).

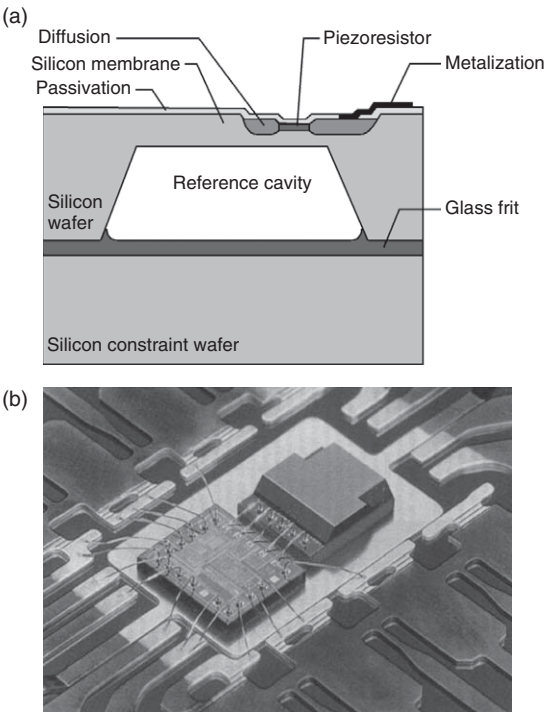


Figure 7 (a) Schematic of piezoresistive pressure sensor by Freescale Semiconductor (formerly Motorola) bonded to silicon constraint wafer using glass frit bonding. (Reproduced by permission from Ristic L (ed.) *Sensor Technology and Devices*, Norwood, MA: Artech House, Inc., 1994.); (b) Photograph of Freescale acceleration sensor and interface circuitry chip mounted on common leadframe; the surface-micromachined accelerometer is protected by glass-frit-bonded wafer cap. (Photo courtesy: Freescale Semiconductor, Austin, TX, USA.)

1.15.3.1.4 Metal bonding

Wafer bonding based on metallic bonding layers includes a wide range of technologies, such as TC bonding, solid-state diffusion bonding, solder bonding, eutectic bonding, and solid–liquid diffusion bonding. While it is beyond the scope of the present chapter to review all investigated metal-based bonding processes and material systems, the fundamental differences of different bonding schemes are discussed in the following based on characteristic examples. Generally, one can distinguish between metal bonding techniques that only involve solid-state material phases and those involving liquid phases. Bonding processes involving a liquid phase,

such as solder bonding, eutectic bonding, or solid–liquid diffusion bonding, more easily enable hermetic bonds even in the presence of surface topography because of the leveling behavior of the liquid phase. However, these processes are usually more prone to outgassing than processes involving only solid-state phases, such as TC bonding and solid-state diffusion bonding. The processes involving a liquid phase can be further subdivided into two categories:

(1) Technologies in which the liquid is solidified by cooling the wafer below its melting temperature (e.g., solder or eutectic bonding).
(2) Solid–liquid diffusion technologies in which the liquid solidifies at constant temperature because of formation of intermetallic compounds with higher melting temperature (e.g., transient liquid-phase bonding, also called isothermal solidification).

In general, all bonding processes involving different metals in the bonding process are experiencing some amount of interdiffusion of the materials even at room temperature. The formation of intermetallic phases must be studied carefully in order to avoid the formation of, e.g., brittle phases or voids over time that could reduce the lifetime of the interconnect. These challenges are well known in the microelectronics packaging area, from which many metal-based wafer bonding techniques have been derived.

1.15.3.1.4.(i) TC bonding In TC bonding, two materials are welded together by the applications of temperature and pressure. The process is used in the microelectronics packaging industry for wire bonding and TAB (see Section 1.15.2.2). In order to form a bond, the separation between the two surfaces must reach atomic dimensions. At an elevated temperature, the involved metals soften and, thus, the required pressure needed for a reliable bond is generally lowered. The most common metal used for TC bonds is gold, because it is relatively soft and does not oxidize. For other materials, any surface oxides on potential bonding partners would first have to be removed either mechanically or chemically.

Wafer-level TC bonds (Taklo *et al.* 2004, Tsau *et al.* 2002, 2004) are carried out at temperatures around 300°C and pressures ranging from about four to several tens of MPA (with respect to the actual bonding area, not the wafer area). By using a compliant polymer film underneath the gold layers, an Au–Au bond has been achieved even at room temperature (Zhang *et al.* 2005). The micrometer-thick gold films can be evaporated, sputtered, or electroplated, with higher surface roughness typically translating into higher required bonding pressures. If silicon wafers are joined by TC bonding, a good diffusion barrier, e.g., a silicon dioxide film, must be placed between silicon substrate and gold layers in order to avoid Au–Si interdiffusion (Tsau *et al.* 2002). Thereby, considerable interdiffusion already takes place at temperature well below the Au–Si eutectic temperature at 363°C.

Recently, a wafer-scale package for RF-MEMS (Min *et al.* 2004) in the DC-50 GHz regime has been demonstrated based on a gold–gold TC bond between a glass substrate and a low conductivity silicon cap. On the substrate side, a 2.5-μm-thick sputtered gold sealing ring is used; on the bonding cap side, a 3.5-μm-thick electroplated gold film is used, resulting in a cavity height of about 6 μm. Bonding was performed at a comparably high temperature of 360°C.

1.15.3.1.4.(ii) Solder and eutectic bonding Solder interconnects are widely used in the microelectronics packaging industry for electrical connections, die bonding, and cavity sealing and it is, thus, no surprise that they have also found their way into wafer bonding applications (see, e.g., Goyal *et al.* 2003, 2004, Harpster and Najafi 2003, Mei *et al.* 2002, Najafi 2003, Wolffenbuttel 1997). Basically, a suitable solder material is deposited in the bond area between a substrate wafer with microsystem components and the capping wafer. The two wafers are aligned to each other and brought into contact and the temperature is increased until the solder reflows and creates a bond between the wafers. The solders can be applied by screen or stencil printing, electroplating, or using solder preforms. Advantages of solder bonding are the relatively low process temperatures, the ability to level surface roughness or topography, and the self-alignment capability originating from the large surface tension of molten solder. A major shortcoming of many solder pastes is that they contain solder fluxes to chemically break surface oxides and ensure proper wetting of the bonding surfaces. The aggressive fluxes can be detrimental to the encapsulated MEMS structures and flux residues should be removed together with other impurities embedded in the solder after the reflow cycle. Because this is generally not possible inside sealed cavities, fluxes should be avoided in the first place, relying on fluxless solder systems and other ways of surface activation. Another disadvantage is that common solder systems only have a

temperature stability up to their melting point; therefore, if SMTs are used to solder the first-level package onto a circuit board, the zero-level packaging must use a solder system with higher melting point. In order to accommodate lead-free solder systems (e.g., the Sn/Ag/Cu system with typical melting temperatures around 220°C), board assembly reflow processes must now maintain peak temperatures of 240–260°C. This generally limits suitable solder systems for zero-level packaging to those having melting points around 300°C, such as 80Au–20Sn or 88Au–12Ge with melting points of 280°C and 356°C, respectively (Tummala and Rymaszewski 1989). An alternative and widely used bonding system is eutectic Au–Si with a eutectic temperature of 363°C (Mei *et al.* 2002, Najafi 2003, Wolffenbuttel 1997). To form the eutectic composition, gold and silicon have to be provided to the interface where the wafers are joined. Typically, gold is deposited on one wafer, whereas either bulk silicon or a polysilicon layer is used on the second wafer as the other bonding surface. For all mentioned systems, adhesion, barrier, and wetting layers have to be carefully chosen in order to achieve a suitable process yield. Wafer-level packaged microsystem examples include an accelerometer with a MEMS wafer and a CMOS circuitry wafer joined by gold–silicon eutectic bonding (Brandl and Kempe 2001), a capacitive pressure sensor with a MEMS wafer and a CMOS circuitry wafer joined by Au–Sn solder bonding (Rogge *et al.* 1998), and RF components sealed by Sn–Pb (Jourdain *et al.* 2005) or Au–Sn (Cai *et al.* 2005) solder bonding. A thin surface-mountable capping wafer with solder balls and vertical electrical feedthrough has been developed by Hymite (http://www.hymite.com, Shiv *et al.* 2006). The cap wafer is bonded to the microsystem wafer using a Au–Sn solder, providing simultaneously a hermetic seal and electrical connections.

1.15.3.1.4.(iii) Solid–liquid diffusion bonding

As mentioned above, a challenge of using conventional solder bonding for zero-level packaging is that the bond will only withstand temperatures up to the melting point of the solder, i.e., up to roughly the initial bonding temperature. If solder joints are also used for the first- and second-level packaging, the solder systems have to be carefully chosen in terms of their melting temperatures in order not to compromise prior packaging steps. If standard SMT is used for the final PCB assembly, the final process temperature for lead-free solder systems is about 240–260°C and all prior soldering steps must have higher melting temperatures. Wafer bonding based on solid–liquid diffusion bonding (also called isothermal solidification or transient liquid phase bonding) (Baltes *et al.* 2000, Du *et al.* 2005, Karpenkopf *et al.* 2005, Waelti *et al.* 2000, Welch *et al.* 2005a, b) provides a temperature stability of the bond well beyond the actual bonding temperature while still offering the advantages of a liquid phase during joint formation. Bonding by isothermal solidification is based on solid–liquid interdiffusion of a high- and a low-melting-point component. By applying a bonding temperature between the melting points of the two components, a liquid phase is formed. The fast solid–liquid interdiffusion leads to the formation of higher melting phases, while the liquid phase is consumed. After all of the liquid phase has reacted, the bond has isothermally solidified. If the bond has not yet reached equilibrium at this point, solid-state interdiffusion continues at a slower rate, leading to the formation of new phases and the dissolution of others. Possible material systems for isothermal solidification are Ag–In, Au–In, Cu–In, Ag–Sn, Au–Sn, Cu–Sn, and Ni–Sn (Welch *et al.* 2005a).

Figure 8 shows an example of an infrared sensor array with a silicon filter attached by Au–In isothermal solidification (Waelti *et al.* 2000, 2001). The CMOS-based sensor chip comprises an array of thermoelectric infrared pixels embedded in a single dielectric membrane, multiplexing circuitry to address individual pixels, and a chopper-stabilized low-noise amplifier (Schaufelbuehl *et al.* 2001). To thermally isolate the individual pixels and mechanically stabilize the millimeter-sized dielectric membrane, a grid of 25-μm-thick gold lines is electroplated on top of the membrane (**Figure 8(a)**). The sensor arrays have been fabricated with a commercial 1 μm CMOS technology of EM Microelectronic-Marin SA, Marin, Switzerland, which also provides the gold electroplating, normally used to prepare the chips for TAB. Au–In isothermal solidification has been chosen to bond a silicon infrared filter directly onto the gold grid of the sensor array (**Figure 8(b)**). To prepare for the bonding step, the silicon filter is coated with a stack of metal layers: a Cr adhesion layer and a Au plating seed layer, jointly patterned by a lift-off technology, an electroplated Ni wetting layer and an electroplated In bonding layer (**Figure 8(c)**). Both, the Ni and In layers are electroplated using photoresist as plating mold. Subsequently, chip and filter are aligned and the fluxless bonding is performed in forming gas atmosphere. The optimized process parameters for

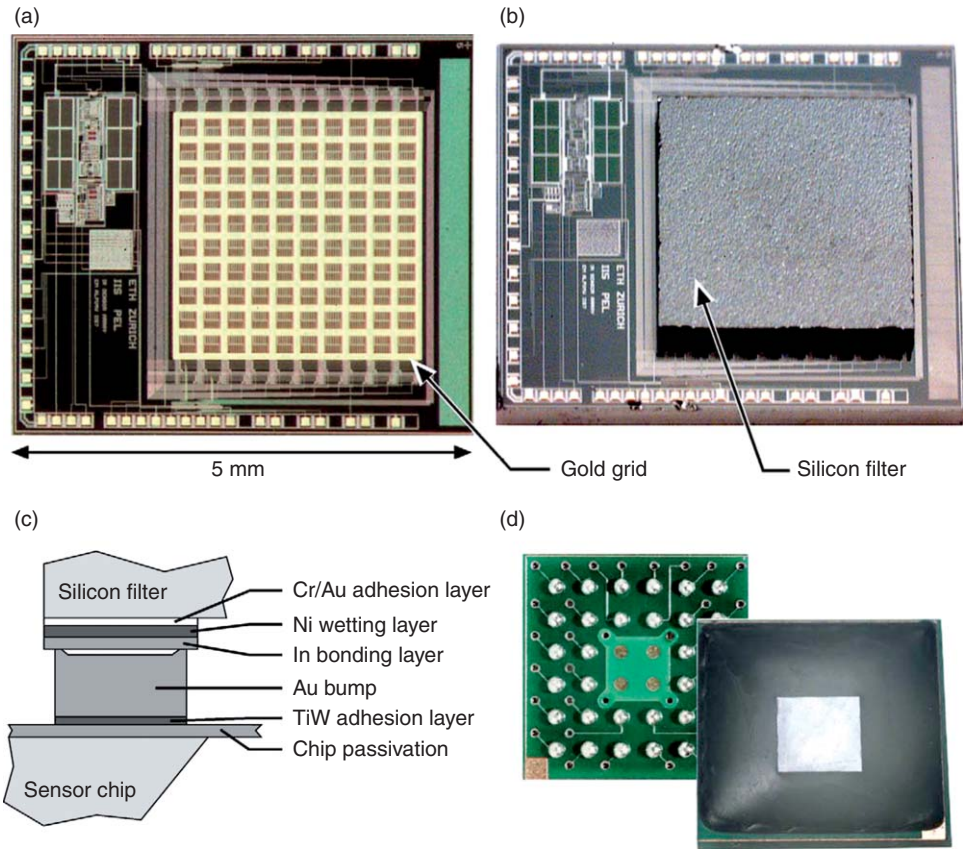


Figure 8 Infrared sensor array developed by ETH Zurich, Switzerland. (a) Chip photograph showing 100-pixel thermoelectric sensor, electroplated gold grid, and on-chip readout circuitry; (b) Sensor chip with silicon infrared filter bonded using Au–In isothermal solidification; (c) Schematic of bonding interface based on Au–In isothermal solidification; (d) Frontside and backside of fully packaged infrared sensor array in ball grid array format with wire bonds and on-chip circuitry protected by glob top. (Schaufelbuehl *et al.* 2001; Waelti *et al.* 2001.)

the isothermal solidification are 195°C for 10 min under a load of 8 MPa (Waelti *et al.* 2001). The completed bond exhibits a temperature stability of 473°C, measured by applying a constant shear stress of 0.5 MPa and raising the temperature until the bond fails. As first-level packaging for the infrared microsystem, a plastic BGA has been chosen (Waelti *et al.* 2001). After die attach and wire bonding, a glob top epoxy is dispensed using a dam and fill process to protect the wire bonds and the microsystem, while providing access to the infrared filter (**Figure 8(d)**).

1.15.3.1.5 Polymer bonding

If a hermetic seal is not required, low-temperature wafer bonds can be achieved using polymer adhesives as bonding layers. Possible adhesives are benzocyclobutene (BCB), polyimide, polymethyl methacrylate (PMMA), SU-8, fluorocarbon polymers, parylene, or Mylar. A summary of the bonding properties of the different films is given in Kim and Najafi (2005). Typical bonding temperatures are well below 200°C for most polymers. An exception is the high-temperature stable polyimide with a bonding temperature above 250°C. While most polymers are applied by spin coating, parylene is deposited from the vapor phase (Kim and Najafi 2005). Most polymers are available in photosensitive versions, facilitating the patterning of the bonding layer. A common material choice is BCB; it features a low loss tangent, high resistivity, and low dielectric constant, making it suitable for high-frequency applications, such as RF-MEMS (see, e.g., Jourdain *et al.* 2003, Kim *et al.* 2005). Because its curing process does not involve catalysts, no outgassing occurs, which generally improves the bond quality (Oberhammer and Stemme 2004).

1.15.3.1.6 Bonding by localized heating

So far, the described wafer bonding processes were performed by heating the overall wafer sandwich to the required process temperature. Therefore, the bonding temperature must be compatible with all on-chip components, such as microstructures, metallization lines, or integrated circuitry. The resulting maximum bonding temperature can severely limit the number of possible bonding processes. However, for reliable bonding, not the overall wafer sandwich but only the actual bonding interface would have to be heated to the desired process temperature. Based on this concept, a number of bonding processes utilizing localized instead of global heating have been developed (see, e.g., Lin 2000, Najafi 2003). A conceptional schematic of a localized heating and bonding process based on resistive heating is given in **Figure 9**. A resistive heating element is integrated in the bonding area either on the substrate or on the capping wafer. After bringing the wafers into contact, the bond is performed by locally heating the bonding area with the integrated heater. Simulations show that the temperature elevation is highly confined to the heater, if it is placed on top of a thermally insulating layer (e.g., an SiO_2 film), which has been deposited on a thermally conductive substrate (e.g., silicon) attached to a good heat sink (Lin 2000). This way, temperature-sensitive components on the microsystem wafer are not affected by the potentially high bonding temperatures in the bonding area. The concept of localized heating and bonding using resistive heating elements has been demonstrated for gold–silicon eutectic bonding, silicon–glass fusion bonding, indium solder bonding, and a localized chemical vapor deposition (CVD) sealing process (Lin 2000).

It is worth mentioning that the localized heating cannot only be achieved with resistive heaters, but also by using microwave heating, laser heating, magnetic induction heating, or RF heating (Najafi 2003). This way, the need to electrically contact the chips or wafers during bonding can be avoided. However, it must be verified that only the bonding areas are heated by these indirect methods and not other microsystem components, such as metallization lines. As an example, localized magnetic induction heating of a solder joint has been demonstrated using ferromagnetic Ni–Co spacers in the vicinity of the bond interface (Yang *et al.* 2005). An induction coil in proximity to the wafer sandwich is powered by a high-frequency power supply, heating the ferromagnetic materials due to eddy current losses.

1.15.3.2 Encapsulation by Thin-Film Shells

The second way to achieve a WLP for fragile MEMS/microsystems is the monolithic thin-film encapsulation (Najafi 2003). In this case, surface-micromachining and thin-film deposition steps are

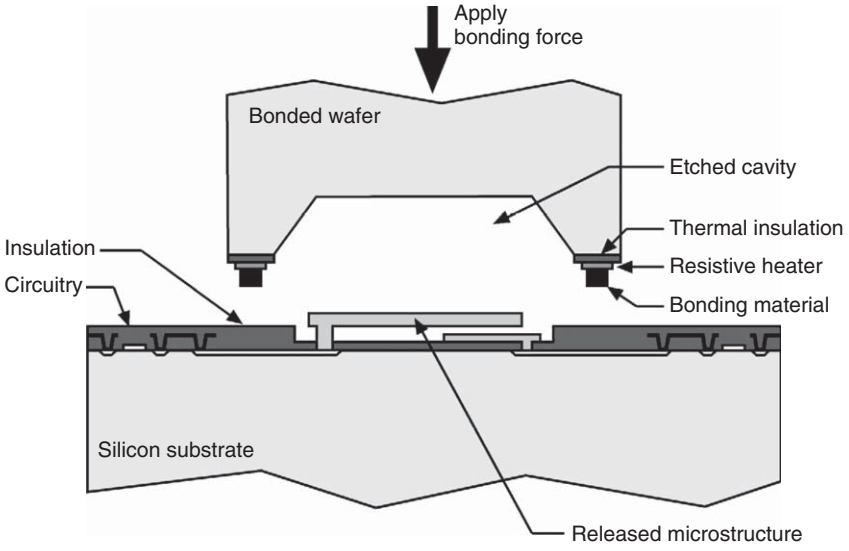


Figure 9 Conceptual schematic of wafer bonding by localized heating with a resistive heater; the heater is integrated in the bonding area of the capping wafer. (Source: Lin L 2000 MEMS post-packaging by localized heating and bonding. *IEEE Trans. Adv. Packag.* **23**, 608–16. Adapted with permission.)

combined to form a thin-film shell directly over the microstructures. These techniques avoid the processing and alignment of two wafers and eliminate the space consuming seal ring typical for wafer bonding. Early thin-film encapsulation processes were commonly developed to vacuum-encapsulate resonant microstructures (Brand and Baltes 1998) and employed polysilicon and silicon dioxide (Burns *et al.* 1995, Legtenberg and Tilmans 1994), highly and lightly p-doped silicon epitaxial layers (Ikeda *et al.* 1990), or silicon nitride and silicon dioxide (Lin *et al.* 1998) as shell and sacrificial layer material combinations. In all cases, the sacrificial layers are removed through etch channels in the shell structure, which must be sealed off in the last process step to provide a hermetic shell. Drawbacks of these processes are the involved high-temperature process steps, which limit the material choices for the microstructure itself. Moreover, the shells are typically rather thin and can be easily damaged themselves or can only be used to seal off rather small cavities.

More recent research is thus targeting fabrication processes, which use low-temperature steps only and can provide thicker shells. One possibility is the fabrication of electroplated metal shells, using photoresist as sacrificial material and electroplated nickel as structural material (Stark and Najafi 2004). With a metal shell thickness of 40 μm, 800 μm × 800 μm cavities have been hermetically sealed with a typical cavity pressure of 1.5 torr. The access channels for the sacrificial etching are sealed, e.g., with a solder plug, with the maximum process temperature being limited to about 250°C (Stark and Najafi 2004). Another recently developed process uses a porous Al_2O_3 thin film, obtained by postdeposition anodization of aluminum as shell material (He and Kim 2006). The sacrificial layer, a photoresist film, is removed though the nanopores of the shell, thus eliminating the need for larger access channels. Finally, the cavity is vacuum-sealed by depositing a low-temperature silicon nitride film over the Al_2O_3 shell. Thereby, the high aspect ratio of the nanopores prevents deposition of silicon nitride inside the cavity. The maximum process temperature is 300°C, but can be reduced by using a different sealing process. The pressure inside the cavity was found to be 8 torr (He and Kim 2006).

So far, all discussed on-chip shells were made from inorganic materials. Recently, an encapsulation method solely based on organic materials has been demonstrated for high-frequency filters and gyroscopes (Joseph *et al.* 2007, Monajemi *et al.* 2005, 2006). The process is based on the thermal decomposition of a sacrificial polymeric material, Unity 2000 (Promerus, LCC), through a solid, spin-coated polymer overcoat, Avatrel (Promerus, LCC). The sacrificial polymer is thermally decomposed at temperatures between 200°C and 300°C and the decomposition products diffuse through the solid overcoat film. Ultimately, an aluminum layer can be sputtered or electroplated over the overcoat to hermetically seal the encapsulated MEMS device (Monajemi *et al.* 2005).

The last process described in this section has been developed by the Bosch Group and Stanford University (Park *et al.* 2003) and was recently commercialized by SiTime (http://www.sitime.com) to encapsulate silicon resonators in epitaxially sealed epi-poly (epitaxially grown polysilicon) chambers buried under the wafer surface (Partridge *et al.* 2005). The MEMS-first process module completely defines, releases, and vacuum-encapsulates the resonant microstructures at the front end of the overall process sequence, providing a reliable zero-level packaging. The fabrication process is based on SOI wafers with a 10- to 20-μm-thick device layer (**Figure 10**). After defining the microstructure in the device layer by deep reactive ion etching (DRIE), an oxide layer is deposited and patterned to cover parts of the microstructure while providing openings for electrical contacts. A thin epitaxial silicon layer is grown and patterned to obtain openings to the oxide layer. After releasing the microstructure by sacrificial layer etching through the openings in the epi-silicon layer, the resonator chamber is sealed by growing a thick epitaxial silicon layer. Thereby, the epitaxial silicon growth yields polycrystalline silicon on top of the oxide layer covering the microstructures, but single-crystalline silicon where the oxide has been removed, i.e., in the potential circuit regions. After surface planarization and formation of contact isolation trenches, the wafers with encapsulated MEMS structures can be used as a starting material for a CMOS process (Partridge *et al.* 2005).

1.15.4 First-Level Packaging for Microsystems

Like conventional microelectronics packaging steps, the first-level packaging steps for microsystems include wafer dicing, die attach, and electrical interconnect formation. Depending on the previous fabrication steps, the wafers entering the first-level

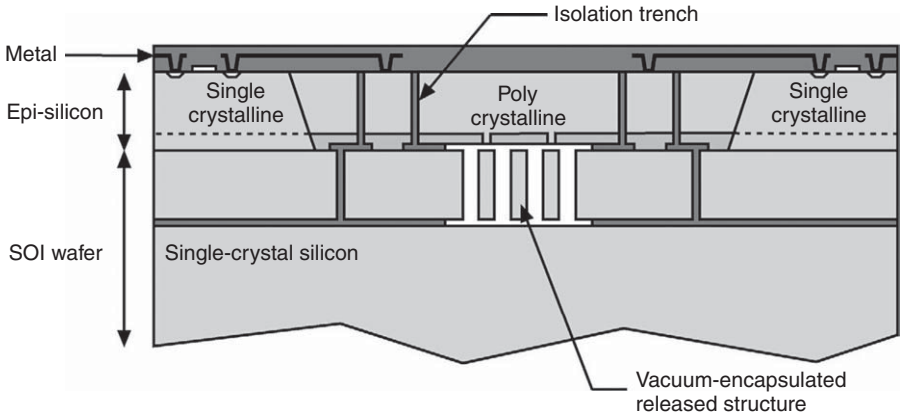


Figure 10 Cross section of EpiSeal technology by SiTime for fabrication of vacuum-encapsulated silicon resonators. (Source: Partridge A, Lutz M, Kim B, Hopcroft M, Candler R N, Kenny T W, Petersen K, Esashi M 2005 MEMS resonators: getting the packaging right. *Proc. 9th SEMI Microsystem/MEMS Seminar*, Semicon, Japan.)

packaging sequence might contain nonreleased microstructures, released and uncapped microstructures, released and capped microstructures, or no microstructures at all. While designing a particular process flow, one must consider that all process and testing steps after dicing are typically performed in a serial manner rather than the cost-effective wafer-scale batch processing. This way, if dicing is done before the final microstructure release in order to protect fragile mechanical structures, the final release step will have to be performed on a single-die level. To maintain the cost advantages of batch processing, the wafer dicing should be done as late as possible in the process flow; as an example, various manufacturers first perform device testing or even trimming steps on the wafer level before dicing.

The first-level package type is determined to a great extent by application and cost considerations: metal and ceramic packages provide a hermetic encapsulation, but are typically more expensive. Molded plastic packages and COB technologies, on the other hand, are more cost-effective packaging solutions if hermeticity is not required.

1.15.4.1 Dicing, Die Attach, and Interconnect

A dicing saw employs a high-speed spindle fitted with a thin diamond blade to dice the wafers. Water jets are used to cool the blade and remove dicing debris. The combination of these water jets and dicing debris can be a major challenge for wafers with released, fragile microstructures. Typically, thin, surface-micromachined microstructures are more prone to dicing damage than bulk-micromachined membrane structures. While the dicing yield can potentially be increased by reducing dicing speed and water pressure, the protection of fragile microstructures during dicing or a microstructures release after dicing might be required.

Several approaches to protect released microstructures during dicing have been discussed in the literature. Analog Devices has devised a backside sawing process for wafers containing released microstructures (Chau and Sulouff 1998). In this process, the wafer with released microsystems is mounted upside down onto a two-layer dicing tape. The top layer has punched openings at the location of the microstructures, thus providing protective air pockets during the dicing and cleaning process. A modified pick-and-place tool is then used to pick up the individual chips for die attach. Alternatively, fragile microstructures can be protected by coating the frontside and backside of the wafer with a protective film, e.g., a photoresist layer, before the sawing process (Agarwal *et al.* 2005). After dicing, the organic films are removed with a conventional resist stripper, washing away most of the sawing debris at the same time. Texas Instruments have implemented a different procedure to singulate their large Digital Micromirror Devices (DMD™) with hundreds of thousands of individual micromirrors (Mignardi 1998): before the mirror release, a protective layer is placed on top of the mirror level, the wafers are partially sawn and subsequently cleaned to remove sawing debris. Still on a wafer

level, the surface-micromachined mirror arrays are now released by removing the organic sacrificial layers in a dry etching step and subsequently tested in a wafer tester. Once the wafer-level tests are completed, the wafers are broken along the partially sawn scribe kerfs to separate the individual functional die.

Dicing of wafers with microstructures that have been capped in a zero-level packaging process is usually less critical. Here, the sawing process has to ensure access to the bonding pads, which might be covered by the capping wafer. In the case of wafer bonding with an intermediate layer, subsequent sawing steps with different sawing depths can individualize the wafer and provide access to the bond pads.

If wire bonding is used for the electrical interconnect formation, the individual dies are attached to the substrate/package after dicing. This can be done by a standard pick-and-place machine, equipped with a special pick-and-place tool to handle chips with released microstructures on their surface. In order to minimize packaging-induced stresses acting on the MEMS part, the soft die attach adhesives with a low modulus of elasticity are widely used. The die attach is followed by a conventional wire bonding process (Section 1.15.2.2). Alternatively, the microsystem can be attached and electrically connected to the substrate/package by flip chip bonding techniques. An example of this approach in the area of chemical sensors is presented in Section 1.15.4.5.

1.15.4.2 Windows to the Outside World

Unlike ICs, most MEMS interact with their environment. Therefore, not only electrical signals need to be routed through the package, but also information from other energy domains: a pressure sensor package must be able to transmit the pressure information to the actual sensor element and a chemical sensor package must provide access to the sensing elements for the analytes of interest. As discussed earlier, the need for windows to the outside world while still providing the typical packaging functions, in particular protection, is one of the major challenges for microsystem packages. Unfortunately, there is typically no *one size fits all* solution and the actual package for the same sensor might look completely different for two applications. A good example in this respect is the pressure sensor, where manufacturers typically offer a wide range of different packages for the same pressure range. Possible variations include through-hole and surface-mount packages, differential and absolute pressure sensors, shape and size of the pressure ports, and the way the actual silicon sensing element is protected. To illustrate this, **Figure 11** compares the schematic cross sections of two pressure sensors from opposite sides of the application spectrum: on the one hand, we have a low-cost pressure sensor in a premolded plastic package (**Figure 11(a)**), in which the sensor element and the wire bonds are protected by a silicone gel; on the other hand, we have a media-compatible pressure sensor (**Figure 11(b)**), in which a flexible steel membrane transmits the pressure into an oil-filled cavity containing the actual sensing element. In the following, a number of different windows or filters, transparent for the particular physical quantity of interest but opaque for others, are briefly discussed. The subsequent subsections then give examples of how such filters are integrated with various package types.

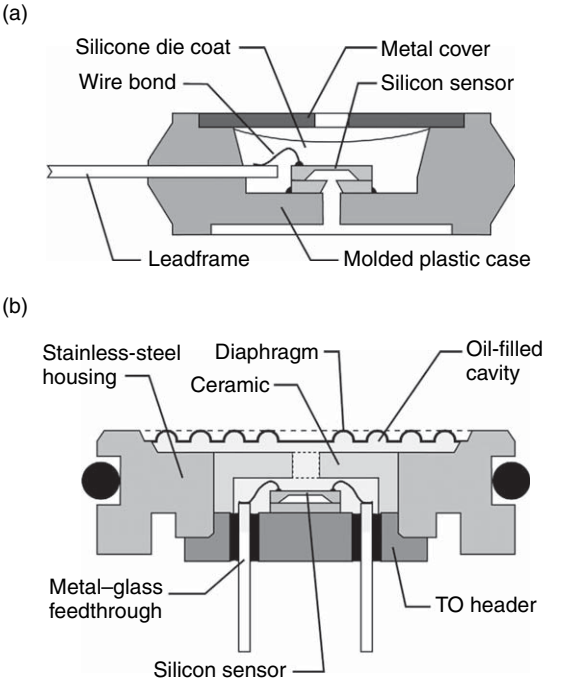


Figure 11 Schematic cross section of (a) low-cost pressure sensor in premolded plastic package. (Source: *Sensor Device Data*, Freescale Semiconductor, Rev. 6.1, 07/2005. Adapted with permission.) and (b) media-compatible pressure sensor embedded in oil-filled stainless-steel cavity bound by flexible membrane. (Source: Datasheet: Series 3L to 10L Pressure Transducers, Keller AG, Winterthur, Switzerland. Adapted with permission.)

Various optical windows are available for handling radiant signals. These windows/filters can be incorporated in metal or ceramic packages (see Section 1.15.4.3), thus providing hermetic microsensor packages. While standard optical glass is transparent in the visible wavelength regime, quartz glass is used to extend the transparent wavelength range into the UV range down to about 200 nm. On the other side of the visible spectrum, semiconductor filters, such as germanium and silicon, are transparent for wavelengths longer than about 1–2 μm, well suited for infrared sensing applications.

In the mechanical domain, the above-mentioned steel membrane or silicone gel are transparent for the pressure information, but protect the microsensor to a different extent from other influences. In particular, the media-compatible pressure sensors provide a hermetic package while still being able to transmit the pressure signal from the environment to the sensing element. Other common mechanical filters are meshes made from stainless steel or polymers, which can be used to protect the sensor element from direct physical contact, while providing access to the environment in, e.g., chemical sensing applications (see, e.g., Figaro gas sensors, http://www.figarosensor.com).

In the chemical domain, filters can reduce the influence of interference molecules on a chemical measurement. As an example, Figaro uses an active carbon filter in its TGS 2442 carbon monoxide sensor and a zeolite filter in its TGS 4161 carbon dioxide sensor to reduce the influence of interference gases (Product Information Sheets: TGS 2442 Carbon Monoxide Sensor and TGS 4161 Carbon Dioxide Sensor, Figaro USA Inc., 2000 and 2003). Zeolites are known to act as molecular sieves because of their microporous structure, while active carbon with its huge surface area efficiently adsorbs, e.g., volatile organic compounds, in the case of the Figaro sensor in particular ethanol.

It is worth noting that the package cannot only provide filtering functions, but also can concentrate the signal of interest. Examples are lenses for radiant signals, flow channels with varying cross-sectional areas to locally increase the flow velocity in mass flow sensors, or ferromagnetic concentrators to concentrate a magnetic field at the location of the sensor. In Popovic *et al.* (2006), such ferromagnetic elements change not only the magnitude but also the direction of the magnetic field, such that a magnetic induction in the plane of the chip can be sensed by conventional lateral Hall elements, sensitive only to field components perpendicular to the chip surface.

1.15.4.3 Metal and Ceramic Packages

Metal and ceramic packages are high-reliability packages, which are especially suited for harsh environments and are thus widely used for military applications. Both package types are considered hermetic and provide a cavity for the microsystem chip. As a result, they are mainly used for uncapped microsystems requiring a hermetic or even vacuum package. Metal and ceramic packages consist of a bottom part or header and a top part or lid. The cavity of a metal package is sealed by welding, soldering, or brazing the lid onto the header. Ceramic packages, on the other hand, frequently use glass frit seals, i.e., low melting temperature glasses with typical process temperatures around 350–400°C. While metal packages are generally through-hole components utilizing metal–glass seals to electrically insulate the individual feedthroughs, both through-hole and surface-mount ceramic packages are widely available.

TO packages are arguably the most common metal packages. Originally used to package single transistors or small circuits, they are widely used today for packaging small- and medium-volume microsystems. Companies such as Schott (http://www.schott.com) offer a wide range of TO headers with different sizes, pin layouts, and up to 16 I/O pins. Even more important for microsystem packaging, TO headers with additional nonelectrical ports are readily available; examples include headers with pressure ports or lids with embedded optical filters, lenses, or metal meshes. **Figure 12** shows a schematic of a TO package as well as images of a chemical sensor by Figaro (http://www.figarosensor.com), an accelerometer by Colibrys (http://www.colibrys.com), and an infrared sensor by Melexis (http://www.melexis.com). Figaro packages all of its chemical sensors in TO packages, generally using a lid with a metal mesh to protect the microsystem. As mentioned earlier, the carbon monoxide sensor TGS 2442 and the carbon dioxide sensor TGS 4161 feature chemical filters, which are integrated into the lid. In case of the MS7000 series accelerometers of Colibrys (Datasheet MS7000 series MEMS Capacitive Accelerometers, Colibrys, Switzerland, 2004), the TO package houses a small ceramic substrate carrying the sensor chip, an ASIC with the readout circuitry and an electrically erasable programmable read-only memory (EEPROM) chip.

Media-compatible pressure sensors, as discussed briefly in Section 1.15.4.2, are often based on TO

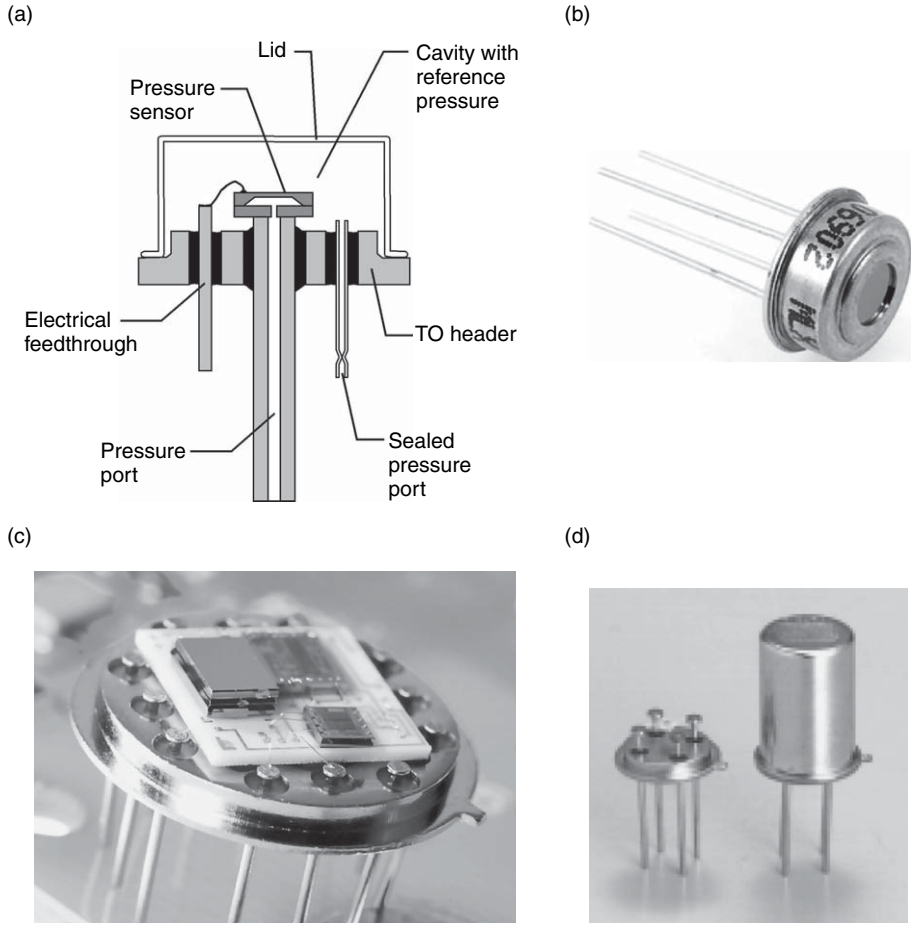


Figure 12 Examples of microsensors packaged in transistor outline (TO) packages. (a) Schematic cross section of TO package with pressure ports and mounted pressure sensor; (b) MLX90247 family of infrared thermopile sensors by Melexis. (Photo courtesy: Melexis, Ieper, Belgium.); (c) MS7000 series accelerometer by Colibrys SA. (Photo courtesy: Colibrys SA, Neuchatel, Switzerland); (d) TGS 4161 Carbon dioxide sensor by Figaro Engineering Inc. (Photo courtesy: Figaro USA, Glenview, IL, USA.)

packages (**Figure 11(b)**). Examples are the series 3L to 10L piezoresistive pressure transducers by Keller (http://www.keller-druck.com): each transducer features a piezoresistive pressure sensor, which is die and wire bonded to a TO header with glass–metal feedthroughs. The header is welded to a stainless-steel housing, which is filled with silicone oil and bound by a thin, laser-welded stainless-steel diaphragm. The diaphragm transmits the media pressure to the oil inside the cavity and, finally, to the silicon pressure sensor.

A large variety of metal and ceramic packages for special applications, in particular optoelectronic components such as fiber-optic transmitter/receiver modules, are available from, e.g., Kyocera (http://www.kyocera.com) and Schott (http://www.schott.com). These packages provide fiber-optic feedthroughs in addition to high-frequency electrical feedthroughs.

In contrast to metal packages, ceramic packages feature thermal expansion coefficients (CTE) that are better matched to those of silicon, helping to reduce packaging-induced stresses. Apart from this, they provide most of the advantages of metal packages, such as excellent mechanical strength, good thermal conductivity (especially in case of aluminum nitride (AlN)), heat durability, and hermeticity. The most common ceramics used in ceramic packaging technology are Al_2O_3, AlN, and low-temperature cofired ceramics (LTCC). LTCC are typically provided as sheets and consist of ceramic particles mixed with an organic binder. Multilayer ceramic circuit boards are manufactured using LTCC by laminating together individual sheets with filled through-holes and printed conductive lines

and cofiring the assembly at firing temperatures around 850°C (Tummala 2001). The cavity of a ceramic package is sealed by either a metal or a ceramic lid. Sealing with a solder system (typical process temperatures of 300–320°C) or a glass frit (typical process temperatures of 320–430°C) provides a hermetic package. If hermeticity is not required, the lid can be attached using an epoxy at a lower temperature between 120°C and 200°C. While through-hole ceramic packages, such as ceramic DIPs, are still widely used in research laboratories, leaded and leadless surface-mount packages dominate most commercial microsystems requiring ceramic packages, because of their compatibility with state-of-the-art circuit board assembly techniques.

Figure 13 shows a cross section of a typical leaded ceramic cavity package, as well as a photograph of a packaged gyroscope by Analog Devices (http://www.analog.com). Analog Devices packages its noncapped, high-volume micromachined accelerometers and gyroscopes in a variety of hermetic ceramic packages, including ceramic LCCs, ceramic LGAs, and ceramic BGAs (see http://www.analog.com/imems). The accelerometers and gyroscopes feature either polycrystalline or single-crystalline silicon sensing structures and are cointegrated with signal processing circuitry on a single chip (Judy 2004).

The DMD, developed and commercialized by Texas Instruments, is arguably one of the most complex MEMS devices. It features an array of 16 µm × 16 µm micromirrors fabricated on CMOS SRAM (static random access memory) cells (Mignardi 1998) and is used for image formation in a wide variety of projection displays. Depending on the image format, a DMD consists of up to 2 million individual mirrors, namely 1920 × 1080 pixels for high-definition television (HDTV) applications. The ceramic packages used to protect these delicate MEMS structures are described in Farris and Kocian (1998). The hermetic package for SVGA, XGA, and SXGA versions of the DMD chip is based on an Al_2O_3 header in a 114-pin LGA format (see schematic cross section in **Figure 14(a)**). A brazed Kovar (29%Ni/54%Fe/17%Co) seal ring forms the cavity for the DMD chip, which is die and wire bonded to the header with an adhesive and gold wire bonds, respectively. The lid consists of a stamped Kovar frame plated with gold/nickel and a projection-quality optical window with a hermetic glass-to-metal seal.

The hermetic window assembly is coated with antireflection coatings and finally seamwelded to the seal ring on the ceramic header, forming a hermetic package. Getter strips are mounted inside the hermetic cavity to getter residual gas molecules. Finally, a heat sink is attached with an adhesive to the ceramic header. All materials have been carefully chosen in order to minimize packaging-induced stresses acting on the mirror array. Besides describing the hermetic package, Farris and Kocian (1998) describe a nonhermetic package solution, in which the optical window is attached to the Kovar frame using an adhesive. A photograph of a packaged 1920 × 1080 pixel DMD chip is depicted in **Figure 14(b)**.

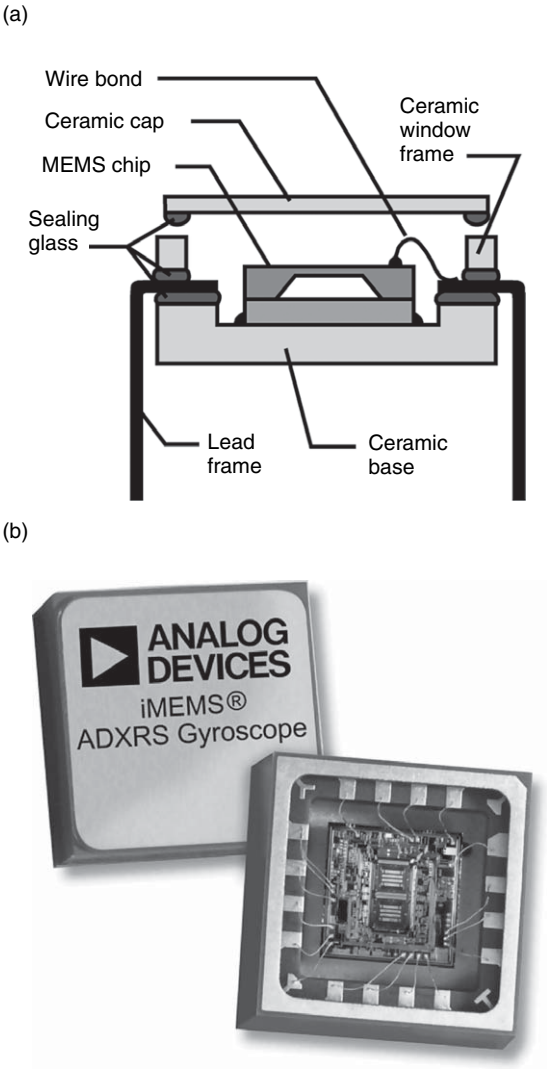


Figure 13 (a) Schematic cross section of leaded ceramic package and (b) inertial sensor ADXRS300 by Analog Devices in ceramic leadless chip carrier. (Photo courtesy: Analog Devices, Norwood, MA, USA.)

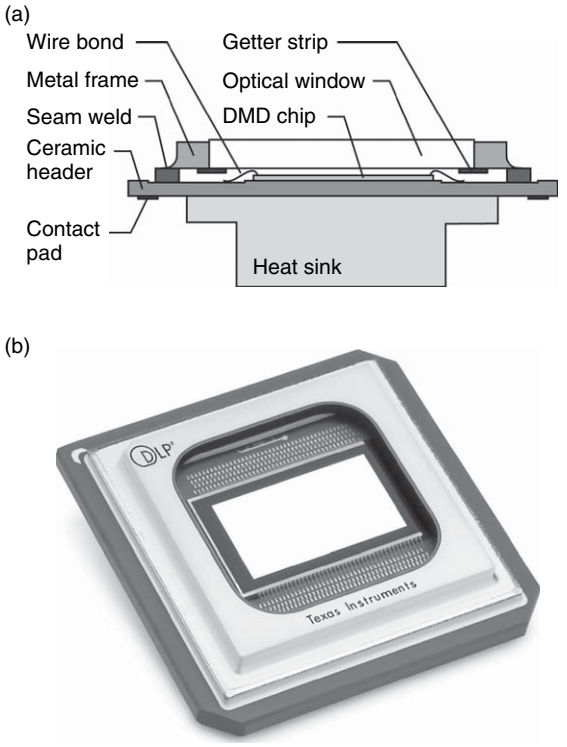


Figure 14 Packaging for Digital Micromirror Device (DMD™) by Texas Instruments. (a) Schematic cross section of hermetic ceramic package with optical window embedded into the lid. (Source: Faris C, Kocian T 1998 DMD™ packages – Evolution and strategy. *TI Tech. J.* **15(3)**, 87–94. Adapted with permission.); (b) Photograph of packaged 1920 × 1080 pixel DMD™ chip. (Photo courtesy: Texas Instruments, Dallas, TX, USA.)

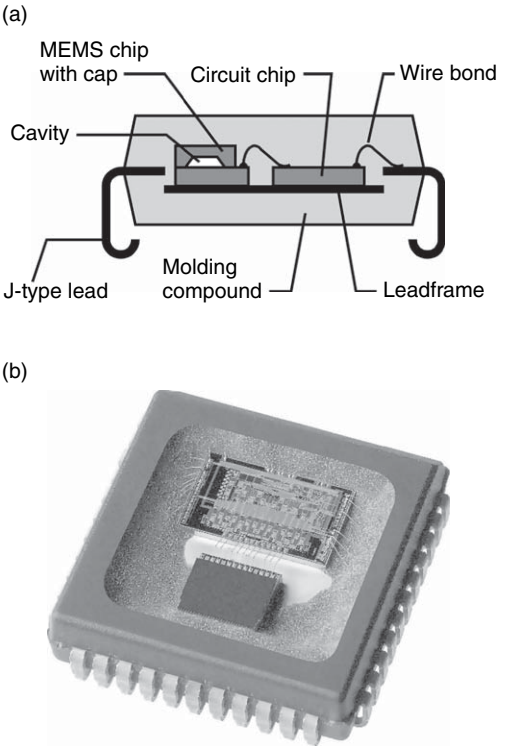


Figure 15 (a) Schematic cross-section through postmolded plastic package with J-type leads; the package houses a capped microelectromechanial systems (MEMS) chip and a circuitry chip on a common leadframe; (b) Photograph of angular rate sensor by Robert Bosch GmbH packaged in postmolded plastic package (44-pin Plastic Leaded Chip Carrier); the molding compound on top of the package is partly removed to show the location of the wafer-bonded sensor chip and the interface circuit. (Photo courtesy: Robert Bosch GmbH, Reutlingen, Germany.)

1.15.4.4 Molded Plastic Packages

We saw earlier that the number of different plastic packages for ICs is ever increasing and it can be challenging to keep up with all the designs and abbreviations. Luckily, most plastic packages used for MEMS have relatively low pin counts and are, thus, leadframe based. To package an IC using a leadframe-based package, the die is commonly die- and wire-bonded onto the leadframe, followed by molding a plastic compound, typically an epoxy, around the chip and wire bonds. Finally, the leads of the leadframe are formed to the desired shape. In case of packaging MEMS, we are typically facing two challenges:

(1) The released mechanical structures are possibly blocked during the molding process.
(2) Micromechanical devices are typically more sensitive to (packaging-induced) mechanical stresses than are ICs.

As a result, not only postmolded plastic packages, i.e., packages where the leadframe–chip assembly is encapsulated with the plastic mold after die attach, but also the so-called premolded plastic packages, which provide a cavity for the microsystem, are frequently used in microsystem packaging.

With a few exceptions (some of them are discussed below), postmolded plastic packages are used to package capped microsystems, i.e., microsystems where the fragile micromechanical structures have been protected by a zero-level packaging process, such as the wafer bonding processes described in Section 1.15.3. In this case, the capped microsystem can be handled just like any other IC and is die- and wire-bonded onto the leadframe and then inserted into a multicavity mold. In a transfer molding press, the hot, liquid molding compound is then transferred through so-called runners in the mold into the cavities surrounding the leadframe–chip assemblies. **Figure 15** shows a cross section

through a leadframe-based postmolded plastic package, as well as a photograph of a postmolded accelerometer from the Bosch Group (http://www.bosch.com). Freescale Semiconductor (formerly Motorola) uses a similar packaging approach for its surface-micromachined accelerometers (Frank *et al.* 1994). In both examples, the capped accelerometer sensor is mounted together with an interface circuit chip on a common leadframe. Electrical interconnects from sensor chip to circuit chip and from circuit chip to the package are made by wire bonding. **Figure 7(b)** shows the Freescale accelerometer after die and wire bonding to the leadframe but before the molding process. In contrast, **Figure 15(b)** shows the packaged Bosch accelerometer after transfer molding with part of the mold removed to illustrate the location of the sensor and circuit chip (Marek 1999). The encapsulation of the die with a transfer mold, paired with the different thermal expansion coefficients of the involved materials, generally results in considerable mechanical stress acting on the packaged silicon chips. To make things worse, the stress distribution is generally not uniform over the microsystem chip, potentially resulting in sensor offsets due to, e.g., piezoresistive effects. To obtain a more uniform stress distribution across the packaged chip, it can be encapsulated with an elastomer, e.g., a silicon gel or a polyimide resin, before the transfer molding process. Amkor Technology calls this process a Gel Dome (MEMS Technology Solution Sheet, Amkor Technology, 2005) and offers the possibility that the gel is in contact with the environment through an opening in the mold. The flexible elastomer inside the plastic package exerts a more or less uniform hydrostatic pressure on the microsystem chip. Microsystems relying on resistive sensing elements, such as Hall sensors or piezoresistive pressure sensors, are particularly sensitive to nonuniform packaging-induced stresses. The influence of a uniform hydrostatic stress on the output signal, i.e., the sensor offset, can be minimized using, e.g., Wheatstone bridge arrangements or dynamic offset compensation techniques.

While capped inertial sensors, i.e., accelerometers and gyroscopes, can be completely encapsulated within a molding compound and are still able to sense inertial forces, many other microsystems require nonelectrical feedthroughs to be able to sense the quantity of interest. This raises the question whether a postmolded plastic package can be outfitted with a pressure port or an optical window. The answer is actually Yes in both cases. As an example, SensoNor (now Infineon Technologies SensoNor AS, http://www.sensonor.com) has developed a packaging technology for its pressure sensors, which provides a pressure port in a postmolded plastic package (Grelland *et al.* 1999). In this approach, glass wafers are anodically bonded to both sides of the silicon sensor chip. One glass wafer features recesses to form the reference (vacuum) cavity on the topside of the micromachined silicon membrane. The glass wafer attached to the bottom features a through-hole that is aligned with an opening in the leadframe during the die attach. A pressure port is now formed during the transfer molding process by pressing a special molding tool directly onto the leadframe. **Figure 16** shows schematics

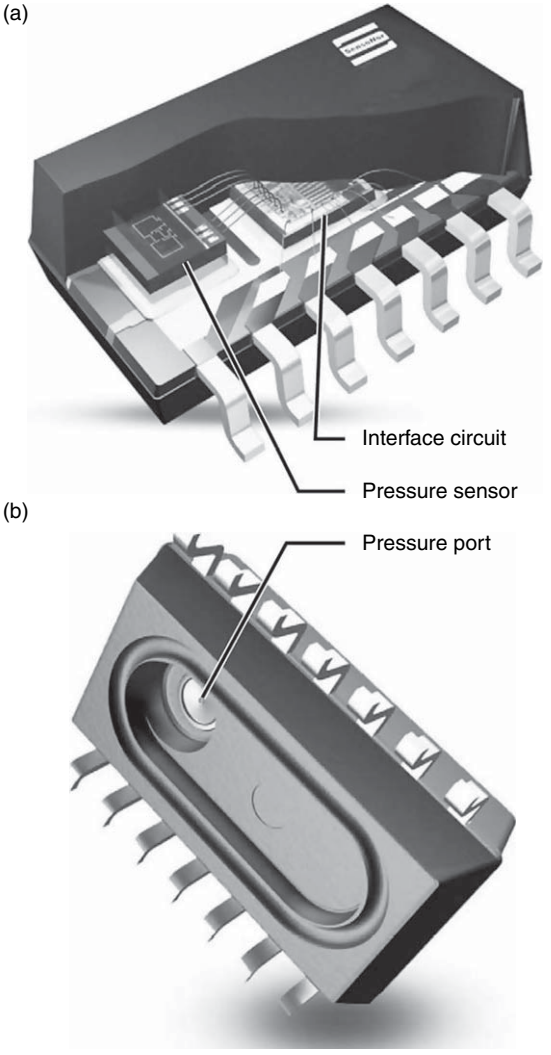


Figure 16 Pressure sensor by Infineon Technologies SensoNor AS packaged in postmolded small outline package. (a) Schematic of packaging concept with triplestack sensor chip and interface circuitry, die and wire bonded to leadframe; (b) Schematic of postmolded pressure sensor package with pressure access port. (Images courtesy: Infineon Technologies SensoNor AS, Horten, Norway.)

of the SP12 pressure sensor package by Infineon Technology SensoNor AS with the pressure port (depending on the actual sensor type located on either the bottom or the top of the surface-mountable SOIC package). Similar technologies have been described by Bossche *et al.* (1997) and are commercially offered by ELMOS Advanced Packaging (formerly Eurasem) (http://www.elmosap.nl) for packaging optical sensors (with an optional optical window), pressure sensors, and mass flow sensors.

The alternative to providing window options in postmolded plastic packages is the use of premolded plastic packages with a cavity for the sensor chip. In a premolded plastic package, the leadframe is partially surrounded by a molded plastic compound, providing a cavity for (uncapped) microsystem chips and access to the chip pad and bond pads for die attach and wire bonding. Premolded plastic packages are basically low-cost versions of the ceramic cavity packages discussed earlier for application that do not require a hermetic encapsulation. Compared to the postmolded plastic packages, microsystems assembled in premolded plastic packages benefit from a simpler, thus less expensive packaging process, are less sensitive to mechanical stress, but lack the possibility of protecting the wire bonds and parts of the chip with the molding compound. Typically, premolded plastic packages are used for pressure sensors, with the sensor chip and wire bonds protected by a silicone gel and/or a plastic lid closing the cavity. **Figure 17** shows a schematic cross section of a premolded plastic package and photographs of pressure sensors from Infineon Technologies (http://www.infineon.com), Silicon Microstructures (http://www.si-micro.com), and Freescale Semiconductor (http://www.freescale.com).

1.15.4.5 Chip on Board

Using COB technologies, bare silicon dies are directly attached to (multilayer) laminate polymer or ceramic circuit boards. Electrical interconnections are done either by wire bonding or by flip chip bonding. In case of ICs, a glob top material, typically an opaque epoxy, dispensed over the chip, finally protects the bare die and the wire bonds. Using COB, bare silicon chips can be mounted together with active and passive surface-mount components on a PCB, effectively skipping the conventional first-level package of the IC. Microelectronic components fabricated this way are nonhermetic, but very cost-effective. In recent years, COB techniques are also used more and more as the base of first-level packages. The main advantage is that not single

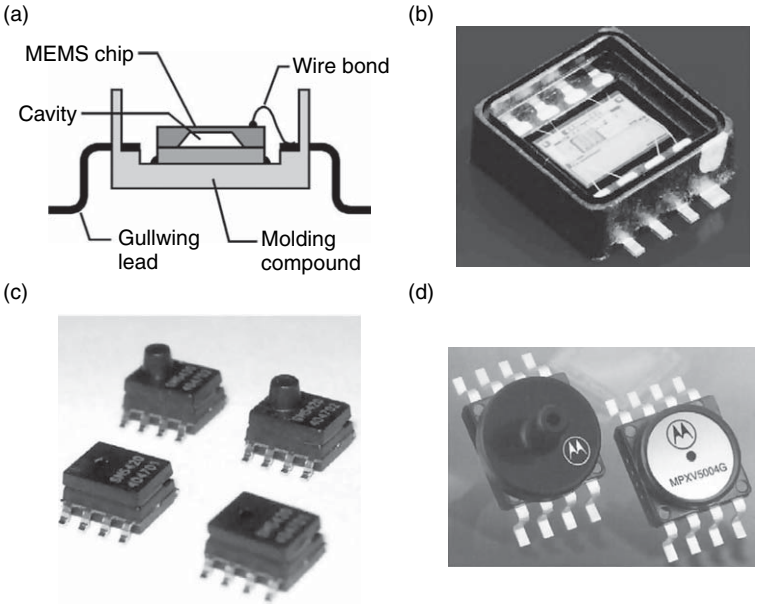


Figure 17 (a) Schematic cross section through premolded plastic package; photographs of pressure sensors packaged in premolded plastic packages: (b) KP100 side crash pressure sensor by Infineon Technologies. (Photo courtesy: Infineon Technologies, Neubiberg, Germany.); (c) SM5420 series pressure sensors by Silicon Microstructures. (Photo copyright: Silicon Microstructures, Milpitas, CA, USA.); (d) MPXV5000 series pressure sensors by Freescale Semiconductor. (Photo courtesy: Freescale Semiconductor, Austin, TX, USA.)

packages but a large-area polymer or ceramic sheet containing arrays of packages are processed. This way, quasi batch processing can be extended through part of the packaging process flow. It should be noted that the substrates are populated in pick-and-place machines that can handle both packaged and bare dies in a serial manner, and that processing a large substrate omits the time-consuming feeding and alignment of individual packages; this is not only true for package processing but also for device testing. Examples of microsensor packages employing COB technologies (**Figure 18**) include the MS5534 pressure sensor modules by Intersema (http://www.intersema.com), the SHT1x series humidity sensors by Sensirion (http://www.sensirion.com), and silicon-based microphones in SMT packages by Knowles (http://www.knowlesacoustics.com) and Infineon Technologies (http://www.infineon.com).

In the following, the package process flow for the MS5534 barometric pressure sensor module (**Figure 18(a)**) developed by Intersema (http://www.intersema.com) will be described in more detail (Datasheet MS5534B Barometer Module, Intersema Bevaix, Switzerland 2005). The module is a hybrid SMD comprising the silicon pressure sensor element and an interface IC on a common ceramic substrate. The interface chip also stores six calibration parameters for a software calibration performed in an external microcontroller. Communication between the sensor module and the microcontroller is achieved by a three-wire digital interface. The packaging process flow (World of Microsystems, multi-media CD-Rom, 1st edn., http://www.fsrm.ch 2000) starts with the preparation of the ceramic substrate by screen-printing conductive and dielectric thick-film patterns. Each ceramic substrate actually holds 81 sensor modules in a 9×9 array pattern. Next, the sensor die, a piezoresistive silicon pressure sensor, which is anodically bonded to a Pyrex wafer under vacuum, and the interface IC chip are die attached to opposite sides of the ceramic substrate. After wire bonding sensor and interface circuit chips, the application-specific integrated circuits (ASICs) are protected by dispensing and curing a glob top material. Next, metal rings surrounding the sensor chips are attached to the ceramic carrier and a soft gel is dispensed to protect the sensor chip from humidity. Then, the modules are calibrated in a pressure chamber at different temperatures and the calibration coefficients are stored in the interface ICs. Finally, the ceramic carriers are individualized.

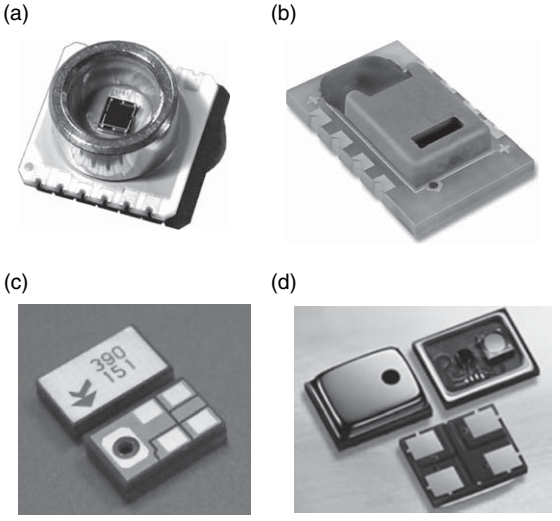


Figure 18 Examples of microsensor packages relying on chip-on-board technologies: (a) MS5534 pressure sensor by Intersema. (Photo courtesy: Intersema, Bevaix, Switzerland.); (b) SHT1x series humidity sensor by Sensirion (Photo courtesy: Sensirion AG, Staefa, Switzerland.); (c) SiSonic microphone by Knowles Acoustics. (Photo courtesy: Knowles Acoustic, Itasca, IL, USA.); (d) Silicon MEMS microphone SMM 310 by Infineon Technologies. (Photo courtesy: Infineon Technologies, Neubiberg, Germany.)

Sensirion's (http://www.sensirion.com) humidity/temperature sensor series SHT1x uses a similar packaging approach based on polymeric instead of ceramic substrates (Datasheet SHT1x/SHT7x, Sensirion, Staefa, Switzerland, 2005). The SHT1x is a single-chip sensor module comprising a polymer-coated capacitive humidity sensor, a bandgap temperature sensor, on-chip readout circuitry with a 14-bit analog-to-digital converter (ADC), and a serial interface. Calibration coefficients are programmed into one-time-programmable (OTP) on-chip memory. The CMOS chip is die and wire bonded to a standard FR4 polymer substrate and protected by a liquid crystal polymer (LCP) cap and a partial epoxy glob top. The SHT1x sensor module is, thus, supplied in a surface-mountable package, similar to a LCC (**Figure 18(b)**).

Figure 19 shows a packaging concept based on flip chip bonding developed at ETH Zurich, Switzerland, for CMOS-based chemical gas sensor arrays (Koll *et al.* 1998, Song *et al.* 2002). A similar concept has been implemented previously for the encapsulation of thermal flow sensors (Mayer *et al.* 1997). The chemical sensor array is based on a single-chip microsystem featuring three chemical sensor

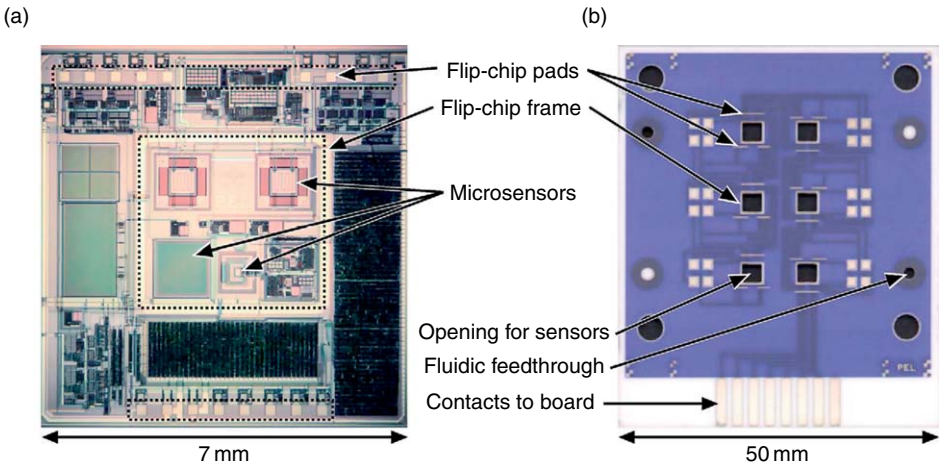


Figure 19 Flip chip packaging concept for chemical sensor microsystem developed at ETH Zurich, Switzerland. (a) Photograph of complementary metal oxide semiconductor (CMOS)-based chemical sensor array with flip chip seal ring and bond pads; (b) Ceramic substrate for flip chip attachment of six sensor chips. (Song *et al.* 2000)

types cointegrated with sensor driving circuitry, readout circuitry, analog-to-digital conversion, and a digital interface (Hagleitner *et al.* 2001). The three transducer elements, a capacitive sensor, a calorimetric sensor, and a mass-sensitive sensor, are each coated with a chemically sensitive polymer to detect volatile organic compounds (VOCs) in air samples. To improve the selectivity of the chemical sensing system, six CMOS chips with a total of 18 chemical sensors are combined on a ceramic substrate, with the sensors of each chip coated by a different polymer. The chip layout (**Figure 19(a)**) has been optimized for the flip chip packaging: the three sensing elements are located in the chip center and are surrounded by an aluminum frame. The on-chip electronics and most of the sensor references are located along the perimeter of the die. Due to the on-board digital interface, the number of I/Os could be substantially reduced and the flip chip pads are located along two edges of the chip. After processing the microsystems in a commercial CMOS process, the flip chip pads and the metal frame surrounding the sensing elements were coated with an electroless nickel/gold (5 µm Ni and 50 nm Au) UBM provided by the bumping service of PacTech, Nauen, Germany, to prepare for the flip chip bonding. After the UBM deposition, the sensor structures are mechanically released using bulk micromachining techniques (Hagleitner *et al.* 2002). Finally, the wafers are diced in a regular wafer saw operated with a reduced water pressure.

Figure 19(b) shows the 50 mm × 50 mm × 0.7 mm ceramic (Al_2O_3) substrate used for the packaging concept (Song *et al.* 2002). Palladium–silver conductor lines and a dielectric layer for electrical insulation were screen printed onto the substrate. The six openings cut into the substrate with a Nd:YAG laser mark the locations for the six microsensor chips. The openings have the size of the sensor area in the center of the chip and are used to expose the microsensors to the air sample. To this end, a fluidic network is attached to the back of the ceramic substrate, pumping the air sample across the polymer-coated sensor surfaces. At the same time, the ceramic substrate protects the circuitry components surrounding the sensor area. A hermetic seal can be achieved by soldering the metal frame surrounding the sensors to a mirror structure on the ceramic carrier during the flip chip bonding process. To prepare for the bonding process, $Sn_{62}Pb_{36}Ag_2$ solder paste is stencil printed onto the flip chip pads on the processed Al_2O_3 substrate. The microsensor chips are placed onto the substrate using a fine placer (or a pick-and-place machine in an industrial environment) and the solder is reflown in a computer-controlled furnace at a peak temperature of 215°C. Subsequently, an epoxy-based underfill is applied and cured to protect on-chip circuitry and solder interconnects and reduce thermal stresses acting on the interconnects. Because of missing capillary forces, the underfill does not flow onto the sensor structures within the openings in the ceramic substrate.

1.15.5 Conclusion: Automation and Cost Issues

After reading through this chapter, the reader might ask where the small outline package for the microsystems/MEMS field is, the first-level package that is good for one-third or even one-half of the microsystem chips. Obviously, such a package does not exist yet and it is highly questionable whether it will ever exist. The requirements for the package are just too different for the various microsystem applications. An RF-MEMS package requires materials and processes that are compatible with high-frequency applications, e.g., low-loss encapsulants with low dielectric constants and electrical interconnects based on flip chip bonding rather than wire bonding. A BioMEMS package might require, in the first place, biocompatible materials and processes that efficiently protect the microsystem from aggressive media. The focus for a microfluidic package might be on the side of standardized microfluidic interconnects and, finally, for an optical MEMS (MOEMS) package, the quality of the optical window or optical interconnect is decisive.

Having said that, the author still likes to acknowledge efforts toward modular packages for microsystems (Schuenemann *et al.* 2000, Ucok *et al.* 2005). System modularity can help reduce system cost, but modular packages will likely target certain application areas, such as microfluidics or optical MEMS. In such a given application area, customer-specific packages might only differ in a limited number of specifications, which can be accounted for in a modular approach. An example of a modular packaging concept is Match-X (http://www.match-x.org), which has been developed in Germany especially for microsystem applications with small- to medium-scale production volumes (Schuenemann *et al.* 2000). Match-X consists of a set of building blocks in the form of BGA packages, which are vertically integrated by standard SMT soldering processes. The building blocks are associated to different microsystem components, such as a pressure sensor, a microcontroller, a microfluidic component, or an MOEMS part. Besides electrical interfaces from level to level, optical and fluidic interconnects are also provided in the walls of the BGA packages. By stacking a set of building blocks on top of each other, a microsystem for a particular application can be built. Different customer needs now only might require the exchange of one particular building block. The individual building blocks can be built from LTCC carriers or from a combination of flexible polyimide-based layers and rigid FR4 layers (Schuenemann *et al.* 2000).

What are other ways to reduce packaging cost? First of all, think about the package at an early stage and design the package together with the MEMS/microsystem chip, making it an integral part of the microsystem. This might seem obvious, but it is more often not done than one might think. Just keep in mind that only about 15% of the microsystems/MEMS publications published over the past 20 years mention the word packaging or package in their searchable abstract. Considering the package needs at an early stage can result in small chip design changes that can reduce packaging cost substantially, because suddenly a standard rather than a custom package can be used. In some cases, parts of the microsystem functionality can be embedded into the package or the package might form the microsystem altogether. A good example along these lines are the recently developed medical pressure sensors that are implemented with standard packaging processes (Allen 2005, Fonseca *et al.* 2006).

Lastly, process automation and batch fabrication should be used wherever possible. WLP processes in the microsystems but also the microelectronics field illustrate this trend of extending batch processing into the packaging and testing process flow, dicing the wafer as late as possible in the packaging sequence. As mentioned earlier, not only wafers benefit from batch processing, but also large-area polymeric and ceramic panels holding arrays of microsystems. The need for automation does not end with first-level packaging, but also applies to the subsequent board assembly (second-level packaging). State-of-the-art board assembly is performed in pick-and-place machines that can handle single-chip packages, passive components, and bare dies for COB. If the microsystem is to be mounted on the board as well, it should be compatible with standard surface-mount assembly processes. A very good example in this respect are silicon-based capacitive microphones that have been commercialized by Knowles Acoustics (http://www.knowlesacoustics.com), Akustica (http://www.akustica.com), and Infineon Technologies (http://www.infineon.com) to replace conventional electret condenser microphones in high-volume applications, such as cell phones. A main selling point of this new generation of microphones is not necessarily a better performance compared to the electret

condenser microphones, but the fact that they are surface-mountable, thus withstanding the lead-free solder reflow cycles of standard board assembly processes (Loeppert and Lee 2006). This advantage leads to significant cost savings for the system integrator.

References

Agarwal A, Mohanraj S, Premachandra C S, Singh J 2005 Process technique for singulating fragile devices. *Proc. Electronic Packaging Technology Conf.* (*EPTC 2005*), Singapore, pp. 716–19

Alexe M, Gösele U 2004 *Wafer Bonding: Applications and Technology*, 1st edn. Springer, Berlin

Allen M G 2005 Micromachined endovascularly-implanted wireless aneurysm pressure sensors: From concept to clinic. *Proc. 13th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '05)*, Seoul, South Korea, pp. 275–8

Arnold D P, Das S, Cros F, Zana I, Allen M G, Lang J H 2006 Magnetic induction machines integrated into bulk-micromachined silicon. *IEEE J. Microelectromech. Syst.* **15**, pp. 406–14

Baldwin D F 2001 Fundamentals of IC assembly. In: Tummala R R (ed.) *Fundamentals of Microsystem Packaging.* McGraw-Hill, New York, Chap. 9, pp. 342–97

Baltes H, Brand O, Waelti M 2000 Packaging of CMOS MEMS. *Microelectron. Reliab.* **40**, 1255–1262

Berenschot J W, Gardeniers J G E, Lammerink T S J, Elwenspoek M 1994 New applications of r.f.-sputtered glass films as protection and bonding layers in silicon micromachining. *Sens. Actuators A* **41–2**, 338–43

Berthold A, Vellekoop M 1997 IC compatible silicon wafer-to-wafer bonding. *Sens. Actuators A* **60**, 208–11

Bossche A, Cotofana C V B, Kaldenberg P, Dommelen I V, Mollinger J 1997 Low-cost reliable transfer mold sensor packaging concept. *Proc. SPIE* **3224**, 153–60

Brand O 2006 Microsensor integration into systems-on-chip. *Proc. IEEE* **94**, 1160–76

Brand O, Baltes H 1998 Micromachined resonant sensors: An overview. In: Baltes H, Göpel W, and Hesse J (eds.) *Sensors Update*. Wiley-VCH, New York, Vol. 4, pp. 3–51

Brandl M, Kempe V 2001 High performance accelerometer based on CMOS technologies with low cost add-ons. *Proc. Int. Conf. Microelectromechanical Systems (MEMS)*, Interlaken, Switzerland, IEEE, pp. 6–9 , IEEE

Bright V M, Stoldt C R, Monk D J, Chapman M, Salian A 2004 Packaging of advanced micro- and nanosystems. In: Baltes H, Brand O, Fedder G K, Hierold C, Korvink J G and Tabata O (eds.) *Advanced Micro and Nanosystems: Enabling Technologies for MEMS and Nanodevices*, Wiley-CH, Weinheim, Germany, Vol. 1, Chap. 4, pp. 93–163

Burns D W, Zook J D, Horning R D, Herb W R, Guckel H 1995 Sealed-cavity resonant microbeam pressure sensor. *Sens. Actuators A* **48**, 179–186

Cai J, Wang Q, Li X, Kim W, Wang S, Hwang J, Moon C 2005 Microstructure of Au–Sn wafer bonding for RF MEMS packaging. *Proc. IEEE Int. Conf. Electronic Packaging Technology,* Shenzhen, China, IEEE, pp. 537–41

Chau K H L, Sulouff R E 1998 Technology for high-volume manufacturing of integrated surface-micromachined accelerometer products. *Microelectron. J.* **29**, 579–86

Despont M, Gross H, Arrouy F, Stebler C, Staufer U 1996 Fabrication of a silicon–Pyrex–silicon stack by a.c. anodic bonding. *Sens Actuators A* **55**, 219–24

Dragoi V, Farrens S, Lindner P 2005 Plasma activated wafer bonding for MEMS. *Proc. SPIE* **5836**, 179–87

Du M, Xu W, Luo L 2005 Explore of a new wafer level hermetic sealing method by Cu/Sn isothermal solidification technique for MEMS/NEMS devices. *Proc. SPIE* **5650**, 332–6

Faris C, Kocian T 1998 DMD™ packages – Evolution and strategy. *TI Tech. J.* **15(3)**, 87–94

Felton L E, Duffy M, Hablutzel N, Farrell P W, Webster W A 2003 Low-cost packaging of inertial MEMS devices. *Proc. SPIE* **5288**, 402–6

Fonseca M A, Allen M G, Kroh J, White J 2006 Flexible wireless passive pressure sensors for biomedical applications. *Proc. Solid-State Sensor, Actuator and Microsystems Workshop*, Hilton Head Island, SC, USA, pp. 37–42

Frank R, Kniffin M L, Ristic L 1994 Packaging for sensors. In: Ristic L (ed.) *Sensor Technology and Devices.* Artech House Inc. Norwood MA, pp. 203–38

Frechette L G, Jacobson S A, Breuer K S, Ehrich F F, Ghodssi R, Khanna R, Wong C W, Zhang X, Schmidt M A, Epstein A H 2005 High-speed microfabricated silicon turbomachinery and fluid film bearings. *IEEE J. Microelectromech. Syst.* **14**, 141–52

Garrou P, Tummala R R 2001 Fundamentals of wafer-level packaging. In: Tummala R R (ed.) *Fundamentals of Microsystem Packaging.* McGraw-Hill, New York, Chap. 10, pp. 398–418

Gilleo K 2005 *MEMS/MOEMS Packaging.* McGraw-Hill, New York

Gösele U, Tong Q-Y 1998 Semiconductor wafer bonding. *Annu. Rev. Mater. Sci.* **28**, 215–41

Gösele U, Tong Q-Y, Schumacher A, Kräuter G, Reiche M, Plössl A, Kopperschmidt P, Lee T-H, Kim W-J 1999 Wafer bonding for microsystems technologies. *Sens. Actuators A* **74**, 161–8

Goyal A, Tadigadapa S, Islam R 2003 Solder bonding for microelectromechanical systems (MEMS) applications. *Proc. SPIE* **4980**, 281–8

Goyal A, Cheong J, Tadigadapa S 2004 Tin-based solder bonding for MEMS fabrication and packaging applications. *J. Micromech. Microeng.* **14**, 819–25

Grelland R, Jakobsen H, Liverod B 1999 A low-cost, fully signal conditioned pressure sensor microsystem with excellent media compatiblity. *Proc. Advanced Microsystems for Automotive Applications*. Springer, Berlin, Germany, pp. 121–31

Hagleitner C, Hierlemann A, Lange D, Kummer A, Kerness N, Brand O, Baltes H 2001 Smart single-chip gas sensor microsystem. *Nature* **414**, 293–6

Hagleitner C, Lange D, Kerness N, Kummer A, Song W H, Hierlemann A, Brand O, Baltes H 2002 CMOS single-chip multisensor gas detection system. *Proc. Microelectromech. Syst. Conf. (MEMS 2002)*, Las Vegas, NV, USA, IEEE, pp. 244–7

Harper C A 2004 *Electronic Packaging and Interconnection Handbook*, 4th edn. McGraw-Hill, New York

Harpster T J, Najafi K 2003 Field-assisted bonding of glass to Si–Au eutectic solder for packaging applications. *Proc. Int. Conf. Microelectromechanical Systems (MEMS 2003)*, Kyoto, Japan, IEEE, pp. 630–3

He R, Kim C-J 2006 A low temperature vacuum package utilizing porous alumina thin film encapsulation. *Proc. Int. Conf. Microelectromechanical Systems (MEMS 2006)*, Istanbul, Turkey, IEEE, pp. 126–9

Hsu T-R 2003 *MEMS Packaging*, 1st edn. INSPEC, The Institution of Electrical Engineers, London

Ikeda K, Kuwayama H, Kobayashi T, Watanabe T, Nishikawa T, Yoshida T, Harada K 1990 Silicon pressure sensor integrates resonant strain gauge on diaphragm. *Sens. Actuators A* **21–3**, 146–50

Iyer S S, Auberton-Herve A J 2002 *Silicon Wafer Bonding Technology for VLSI and MEMS Applications*, 1st edn. INSPEC, The Institution of Electrical Engineers, London

Joseph P, Monajemi P, Ayazi F, Kohl P A 2007 Wafer-Level packaging of micromechanical resonators. *IEEE Trans. Adv. Packag.* **30**, 19–26

Jourdain A, Rottenberg X, Carchon G, Tilmans H A C 2003 Optimization of 0-level packaging for RF-MEMS devices. In: *Proc. 12th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '03)*, Boston, MA, USA, pp. 1915–18

Jourdain A, Vaesen K, Scheer J M, Weekamp J W, van Beek J T M, Tilmans H A C 2005 From zero to second-level packaging of RF MEMS devices. *Proc. Int. Conf. Microelectromechanical Systems (MEMS 2005)*, San Jose, CA, USA, IEEE, pp. 36–9

Judy M W 2004 Evolution of integrated inertial MEMS technology. *Proc. Solid-State Sensor, Actuator and Microsystems Workshop*, Hilton Head Island, SC, USA, pp. 27–32

Karpenkopf L, Frage N, Ripp A, Froumin N, Dariel M P 2005 Sealing technique for wafer-level integrated cavity using In–Ag multilayers. *Proc. SPIE* **5716**, 63–9

Kim H, Najafi K 2005 Characterization of parylene-assisted wafer bonding: Long-term stability and influence of process chemicals. In: *Proc. 13th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '05)*, Seoul, South Korea, pp. 2015–18

Kim K, Kim J-M, Kim J-M, Baek C-W, Kim Y-K 2005 Packaging method for RF MEMS devices using LTCC capping substrate and BCB adhesive bonding. *Proc. 13th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '05)*, Seoul, South Korea, pp. 1092–5

Knechtel R 2005 Glass frit bonding: An universal technology for wafer level encapsulation and packaging. *Microsyst. Tech.* **12**, 63–8

Knechtel R, Wiemer M, Froemel J 2006 Wafer level encapsulation of microsystems using glass frit bonding. *Microsyst. Tech.* **12**, 468–72

Ko W H, Suminto J T, Yeh G J 1985 Bonding techniques for microsensors. In: Fung C D, Cheung P W, Ko W H, and Fleming D G (eds.) *Sensor Technology and Devices.* Elsevier, Amsterdam, pp. 41–61

Koll A, Kawahito S, Mayer F, Hagleitner C, Scheiwiller D, Brand O, Baltes H 1998 A flip-chip packages CMOS chemical microsystem for detection of volatile organic compounds. *Proc. SPIE* **3328**, 223–32

Lau J H 1995 *Flip Chip Technologies*, 1st edn. McGraw-Hill, New York

Lau J H 2000 *Low Cost Flip Chip Technologies for DCA, WLCSP, and PBGA Assemblies*, 1st edn. McGraw-Hill, New York

Lau J H, Wong C P, Prince J L, Nakayama W 1998 *Electronic Packaging: Design, Materials, Process, and Reliability*, 1st edn. McGraw-Hill, New York

Legtenberg R, Tilmans H A C 1994 Electrostatically driven vacuum-encapsulated polysilicon resonators. *Sens. Actuators A* **45**, 57–66

Lin L 2000 MEMS post-packaging by localized heating and bonding. *IEEE Trans. Adv. Packag.* **23**, 608–16

Lin L, Howe R T, Pisano A P 1998 Microelectromechanical filters for signal processing. *IEEE J. Microelectromech. Syst.* **7**, 286–94

Loeppert P V, Lee S B 2006 SiSonic – The first commercialized MEMS microphone. *Proc. IEEE Solid-State Sensor Actuator Workshop*, Hilton Head, SC, USA, IEEE, pp. 27–30

Marek J 1999 Microsystems for automotive applications. *Proc. 13th European Conf. Solid-State Transducers (Eurosensors XIII)*, Den Hague, The Netherlands, pp. 1–8

Mayer F, Haeberli A, Jacobs H, Ofner G, Paul O, Baltes H 1997 Single chip CMOS anemometer. *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Washington, DC, USA, pp. 895–8

Mei Y, Lahiji R, Najafi K 2002 A robust gold–silicon eutectic wafer bonding technology for vacuum packaging. *Proc. Solid-State Sensor, Actuator and Microsystems Workshop*, Hilton Head Island, SC, USA, pp. 86–9

Mignardi M A 1998 From ICs to DMDs. *TI Tech. J.* **15(3)**, 56–63

Min B-W, Entesari K, Rebeiz G M 2004 DC-50 GHz low-loss wafer-scale package for RF MEMS. *Proc. 34th European Microwave Conference*, Amsterdam, The Netherlands, pp. 1289–91

Monajemi P, Joseph P J, Kohl P A, Ayazi F 2005 A low cost wafer-level MEMS packaging technology. *Proc. Int. Conf. Microelectromechanical Systems (MEMS 2005)*, San Jose, CA, USA, IEEE, pp. 634–7

Monajemi P, Joseph P, Kohl P A, Ayazi F 2006 Wafer-level packaging of MEMS via thermally released metal-organic membranes. *J. Micromech. Microeng.* **16**, 742–50

Monajemi P, Ayazi F, Sparks D 2007 MEMS packaging. In: Tummala R R (ed.) *Introduction to System-on-Package (SOP)*. McGraw-Hill, New York

Moraja M, Amiotti M 2003 Getters films at waver level for wafer to wafer bonded MEMS. *Proc. 2003 Design, Test, Integration and Packaging of MEMS/MOEMS Conf.*, Cannes, France, pp. 346–9

Najafi K 2003 Micropackaging technologies for integrated microsystems: Applications to MEMS and MOEMS. *Proc. SPIE* **4979**, 1–19

Oberhammer J, Stemme G 2004 Contact printing for improved bond-strength of patterned adhesive full-wafer bonded 0-level packages. *Proc. Int. Conf. Microelectromechanical Systems (MEMS 2004)*, Maastricht, Holland, IEEE, pp. 713–16

Obermeier E 1995 Anodic wafer bonding. *Proc. 3rd Int. Symp. Semiconductor Wafer Bonding: Science, Technology and Applications*, Reno, NV, USA, pp. 212–20

Oppermann H H, Kallmayer C, Klein M, Aschenbrenner R, Reichl H 2000 Advanced flipchip technologies in RF, microwave and MEMS applications. *Proc. SPIE* **4019**, 308–14

Park W-T, Candler R N, Kronmueller S, Lutz M, Partridge A, Yama G, Kenny T W 2003 Wafer-scale film encapsulation of micromachined accelerometers. *Proc. 12th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '03)*, pp. 1903–6

Partridge A, Lutz M, Kim B, Hopcroft M, Candler R N, Kenny T W, Petersen K, Esashi M 2005 MEMS resonators: Getting the packaging right. *Proc. 9th SEMI Microsystem/MEMS Seminar*. Semicon, Chiba, Japan

Pascariu G, Cronin P, Crowley D 2003, Next generation electronics packaging utilizing flip chip technology. *IEEE/CPMT/SEMI Int. Electronics Manufacturing Technology Symposium*, San Jose, CA, USA, pp. 423–6

Pelzer R, Dragoi V, Kettner P, Lee D 2004 Advanced low temperature bonding technologies. *Proc. Int. Conf. Semiconductor Electronics (ICSE)*, Kuala Lampur, Malaysia, IEEE, pp. 34–8

Petersen K, Barth P, Poydock J, Brown J, Mallon J Jr., Bryzek J 1988 Silicon fusion bonding for pressure sensors. *Proc. IEEE Solid-State Sensor Actuator Workshop*, Hilton Head, SC, USA, IEEE, pp. 144–7

Popovic R S, Drljaca P M, Kejik P 2006 CMOS magnetic sensors with integrated ferromagnetic parts. *Sens. Actuators A* **129**, 94–99

Rogge B, Moser D, Oppermann H, Paul O, Baltes H 1998 Solder-bonded micromachined capacitive pressure sensors. *Proc. SPIE 3514,* pp. 307–15

Schaufelbuehl A, Schneeberger N, Münch U, Waelti M, Paul O, Brand O, Baltes H, Menolfi C, Huang Q, Loepfe M, Doering E

2001 Uncooled low-cost thermal imager based on micromachined CMOS integrated sensor array. *IEEE J. Microelectromech. Syst.* **10**, 503–10

Schmidt M A 1998 Wafer-to-wafer bonding for microstructure formation. *Proc. IEEE* **86**, 1575–85

Schuenemann M, Jam K A, Grosser V, Leutenbauer R, Bauer G, Schaefer W, Reichl H 2000 MEMS modular packaging and interfaces. *Proc. IEEE Electronic Components and Technology Conference (ECTC)*, Las Vegas, NV, USA, pp. 681–8

Shiv L, Heschel M, Korth H, Weichel S, Hauffe R, Kilian A, Semak B, Houlberg M, Egginton P, Hase A, Kuhman J 2006 Ultra thin hermetic wafer level chip scale package. *Proc. IEEE Electronics Components and Technology Conference (ECTC)*, Las Vegas, NV, USA, IEEE, pp. 1122–8 , IEEE

Song W H, Hagleitner C, Hierlemann A, Brand O, Baltes H 2002 Multi flip-chip module for chemical microsystems. In: *Proc. IEEE 4th Int. Workshop Area Array Packaging Technologies*, Berlin, Germany, pp. 4.5.1–4.5.5

Spangler L 2007 *The Art and Science of Microsystem Packaging,* 1st edn Springer, Berlin

Sparks D R, Massoud-Ansari S, Najafi N 2003 Chip-level vacuum packaging of micromachines using NanoGetters. *IEEE Trans. Adv. Packag.* **26**, 277–82

Sparks D R, Massoud-Ansari S, Najafi N 2004 Reliable vacuum packaging using NanoGetters™ and glass frit bonding. *Proc. SPIE* **5343**, 70–8

Stark B H, Najafi K 2004 A low-temperature thin-film electroplated metal vacuum package. *IEEE J. Microelectromech. Syst.* **13**, 147–57

Suni T, Henttinen K, Lipsanen A, Dekker J, Luoto H, Kulawski M 2006 Wafer scale packaging of MEMS by using plasma-activated wafer bonding. *J. Electrochem. Soc.* **153**, G78–82

Taklo M M V, Storas P, Schjolberg-Henriksen K, Hasting H K 2004 Strong, high-yield and low-temperature thermocompression silicon wafer-level bonding with gold. *J. Micromech. Microeng.* **14**, 884–90

Tong Q-Y, Gan Q, Hudson G, Fountain G, Enquist P, Scholz R, Gösele U 2003 Low temperature hydrophobic silicon wafer bonding. *Appl. Phys. Lett.* **83**, 4767–9

Tsau C H, Spearing S M, Schmidt M A 2002 Fabrication of wafer-level thermocompression bonds. *IEEE J. Microelectromech. Syst.* **11**, 641–7

Tsau C H, Spearing S M, Schmidt M A 2004 Characterization of wafer-level thermocompression bonds. *IEEE J. Microelectromech. Syst.* **13**, 963–71

Tummala R R 2001 *Fundamentals of Microsystems Packaging*. McGraw-Hill, New York

Tummala R R, Rymaszewski E J 1989 *Microelectronics Packaging Handbook,* Van Nostrand Reinhold, New York

Tummala R R, Rymaszewski E J, Klopfenstein A G 1997 *Microelectronics Packaging Handbook*, 2nd edn. Chapman & Hall, New York

Ucok A B, Giachino J M, Najafi K 2005 Modular assembly/packaging of multi-substrate microsystems (WIMS CUBE) using thermo-magnetically actuated cables. *Proc. Int. Conf. Microelectromechanical Systems (MEMS 2005)*, Miami Beach, FL, USA, IEEE, pp. 536–9

Ulrich R K, Brown W D 2006 *Advanced Electronic Packaging*, 2nd edn. Wiley-IEEE Press, Hoboken, NJ, Piscataway, NJ, USA

Wallis G, Pomerantz D I 1969 Field assisted glass–metal sealing. *J Appl. Phys.* **40**, 3946–9

Waelti M, Schneeberger K, Brand O, Baltes H 2000 Fluxless microjoining by Au–In–Ni isothermal solidification. *Proc. Mater. Res. Soc. Meet.* **605**, 183–8

Waelti M, Schaufelbühl A, Brand O, Baltes H 2001 Plastic BGA package and direct filter attachment for CMOS thermal imagers. *Transducers 2001, Digest of Technical Papers*, Munich, Germany, pp. 198–201

Wei J, Nai M L, Wong C K, Sun Z, Lee L C 2003 Low temperature glass-to-glass bonding. *IEEE Trans. Adv. Packag.* **26**, 289–94

Weichel S, de Reus R, Lindahl M 1998 Silicon-to-silicon wafer bonding using evaporated glass. *Sens. Actuators A* **70**, 179–84

Welch W C III, Chae J, Lee S-H, Yazdi N, Najafi K 2005a Transient liquid phase (TLP) bonding for microsystem packaging applications. *Transducers 2005, Digest of Technical Papers*, Seoul, South Korea, pp. 1350–3

Welch W C III, Chae J, Najafi K 2005b Transfer of metal MEMS packages using a wafer-level solder transfer technique. *IEEE Trans. Adv. Packag.* **28**, 643–9

Wolffenbuttel R F 1997 Low-temperature intermediate Au–Si wafer bonding: Eutectic or silicide bond. *Sens. Actuators A* **62**, 680–6

Yang H-A, Wu M, Fang W 2005 Localized induction heating solder bonding for wafer level MEMS packaging. *J. Micromech. Microeng* **15**, 394–9

Yun C H, Brosnihan T J, Webster W A, Villarreal J 2005 Wafer-level packaging of MEMS accelerometers with through-wafer interconnects. *Proc. IEEE Electron. Components and Technology Conference (EPTC 2005)*, Lake Buena Vista, FL, USA, pp. 320–3

Zhang W-Y, Labukas J P, Tatic-Lucic S, Larson L, Bannuru T, Vinci R P, Ferguson G S 2005 Novel room-temperature first-level packaging process for microscale devices. *Sens. Actuators A* **123–4**, 646–54

Ziaie B, von Arx J A, Dokmeci M R, Najafi K 1996 A hermetic glass–silicon micropackage with high-density on-chip feedthroughs for sensors and actuators. *IEEE J. Microelectromech. Syst.* **5**, 166–79

Biography

Dr. Brand is an Associate Professor in the School of Electrical and Computer Engineering at the Georgia Institute of Technology. He received his diploma degree in Physics from Technical University Karlsruhe, Germany in 1990 and his Ph.D. degree from ETH Zurich, Switzerland in 1994. From 1995 to 1997, he worked as a postdoctoral fellow at the Georgia Institute of Technology. From 1997 to 2002, he was a lecturer at ETH Zurich in Zurich, Switzerland and deputy director of the Physical Electronics Laboratory (PEL).

Dr. Brand has co-authored more than 120 publications in scientific journals and conference proceedings. He is a co-editor of the Wiley-VCH book series *Advanced Micro and Nanosystems*, a member of the editorial board of *Sensors and Materials*, a member of the technical program and steering committees of the *IEEE International Conference on Microelectromechanical Systems*, and has been elected General Co-Chair of the 2008 *IEEE International Conference on Micro Electro Mechanical Systems (MEMS 2008)*. Dr. Brand is a senior member of the *IEEE* and a co-recipient of the *2005 IEEE Donald G. Fink Prize Paper Award*. His research interests are in the areas of CMOS-based microsystems, microsensors, MEMS fabrication technologies, and microsystem packaging.

1.16 Interface Circuits

M. Ortmanns, A. Buhmann, and Y. Manoli, Department of Microsystems Engineering – IMTEK, University of Freiburg, Freiburg, Germany

Glossary

AC Alternating Current
ADC Analog-to-Digital Converter
ASIC Application-Specific Integrated Circuit
CDS Correlated Double Sampling
CHS Chopper Stabilization
CMOS Complementary Metal Oxide Semiconductor
CMRR Common-Mode Rejection Ratio
DC Direct Current
FES Functional Electrical Stimulation
FET Field-Effect Transistor

HV High Voltage
LV Low Voltage
MEM Microelectromechanical
PSD Power Spectral Density
RF Radio Frequency
SC Switched Capacitor

1.16.1 Introduction

While the electronic properties of silicon have been widely utilized in high-volume integrated circuit (IC) manufacturing for many decades, the use of its mechanical and other properties is catching up in many applications. IC-compatible micromachining techniques form the technological basis for the construction of miniature, high-precision mechanical structures in and on a silicon substrate and thus allow fabrication of various microsensing elements (Petersen 1982).

These developments in silicon technology also allow the implementation of nonelectrical functions alongside sophisticated microelectronic circuitry on the same silicon substrate (Manoli and Mokwa 1999, Wise and Najafi 1991). Thus, microsensors and microactuators can be merged together with microelectronics to form complex microsystems such as system-on-a-chip (SoC) or are realized as two-chip solution resulting in a system in package (SiP). Such systems may contain sensors fully compatible to a standard complementary metal oxide semiconductor (CMOS) process technology, as well as sensors requiring additional processing steps, or sensors on separate substrates. The accompanying microelectronics makes it possible to realize not only sensor readout but also signal conditioning, interfacing, and complex signal processing.

CMOS technologies are the preferred choice for electronic interface circuits, since CMOS is the mainstream technology and is distinguished by its ability to combine complex digital functions with analog electronics. Its low-power capabilities make it suitable also for energy autonomous sensor acquisition or portable and implantable systems. The silicon integration of both sensors and electronics or their combination in a package, besides lowering manufacturing costs for high-volume production, offers several significant advantages, such as small size and low weight (e.g., important for medical implants or portable systems), close physical proximity of sensors and readout electronics (important for low-noise pickup), and the possibility of integrating multisensor arrays in combination with highly complex signal processing.

This chapter discusses some of the challenges met in designing such silicon microsystems from the interface electronics point of view.

1.16.2 Electronic Interface Systems

The challenges faced when designing sensing or actuating microsystems span the whole spectrum of microelectronics. The wide range of applications as well as the demanding and often changing operating conditions for such microsystems makes the system design a challenging task. Illustratively, in **Figure 1** a typical architecture of a signal acquisition and processing system is given. Since the circuit designer must deal with all parts of it, the following discussions will focus on this overall system concept. In Sections 1.16.3 and 1.16.4, only the critical parts of sensor interfacing and analog signal processing as well as analog actuator interfacing are discussed in more detail.

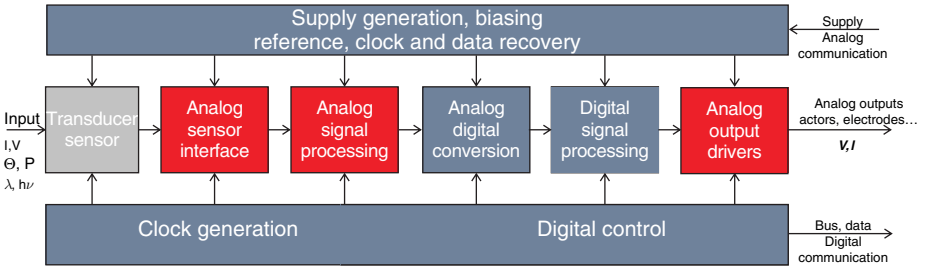


Figure 1 Illustration of an integrated sensor acquisition channel.

1.16.2.1 Integration Concepts

In the case of monolithically integrated systems, sensor and process engineers have to combine CMOS standard processes with additional steps for the implementation of the nonelectrical functions without influencing the electrical parameters of the standard devices. On the other hand, integrating multichip implementations makes life challenging for the package engineer as well. Whereas a usual package is designed to protect silicon from its aggressive environment, sensors need an approximate exposure that renders them sensitive to – and only to – the quantity to be detected. Circuit designers have to detect and amplify the usually minuscule signals delivered by the miniaturized sensors. This has to be accomplished when the circuit is being exposed to conditions (temperature, light, humidity, etc.) from which an IC is usually very well protected. In order to function as a sensor, silicon (as well as any other material) has, of course, to be exposed to its environment.

1.16.2.2 Sensing Devices

Although silicon is a promising sensor material, it should be noted that not all types of sensors are amenable to the silicon integration. Thus, while several physical effects in silicon, such as thermal conductivity, photoconductivity, or magnetoresistivity, can be used directly to build sensors, some effects (e.g., piezo- and magnetoresistivity) may require deposition of additional materials. In this case, the silicon chip can serve as a common carrier substrate for the deposition and for the integration of electronic circuits. In some cases, separate substrates for the sensors and the electronics are the solution of choice.

As already mentioned, for the implementation of silicon microsystems, those sensors and actuators that can be fabricated compatible to a standard CMOS process are usually preferred. One group of such elements is directly based on electronic devices. There is no need for additional masks or process steps and the sensor can be made part of the electronic circuit, i.e., a differential stage. Different devices, for example, can be designed so that they are sensitive to magnetic fields that are parallel or perpendicular to the silicon surface (Gottfried-Gottfried *et al.* 1991, Nathan and Baltes 1989). Photosensitive MOS transistors and complex CMOS camera systems have also been reported (Smith *et al.* 1998). Different electron devices can be used for temperature sensing (Kordas *et al.* 1990).

A further group of CMOS compatible sensors is also based on standard process techniques but additional photomasks and process steps (e.g., deposition and liftoff) are necessary for the fabrication of these devices. One example in this group are the chemically sensitive MOS devices (e.g., transistor, capacitors). In contrast to the standard polysilicon transistor gate or capacitor plate, a sensor-specific material is used. By applying the appropriate material, devices can be obtained that are sensitive to ion concentrations in aqueous solutions (ion-sensitive field-effect transistors, ISFETs). They can also serve as the basic structure for biosensors (Eversmann 2003, Kakerow *et al.* 1993).

In the 1990s, a major development focused on the field of surface micromachining. This technique also needs extra deposited materials such as polysilicon or metal but processes them so that moveable parts are produced. One example is the digital mirror device utilized in the digital light processing technology (van Kessel 2001). **Figure 2** shows an example of a pressure sensor fabricated with planar etching techniques, which also fits into this group of devices. This sensor is formed by a capacitor consisting of a polysilicon membrane and a silicon substrate. The membrane is obtained by laterally etching a sacrificial oxide layer, and the spacer, via etch channels. These channels are sealed by depositing silicon dioxide on the membrane. The pressure sensors are very small, having diameters of up to 150 µm. To get larger signals, sensor arrays can be easily designed (Dudaicevs *et al.* 1994). The capacitive readout techniques compatible with this sensor are discussed in Section 1.16.3.2.

Using similar techniques, accelerometers can be produced, which, nowadays are widely used in air bags and stabilization systems in automobiles. In this

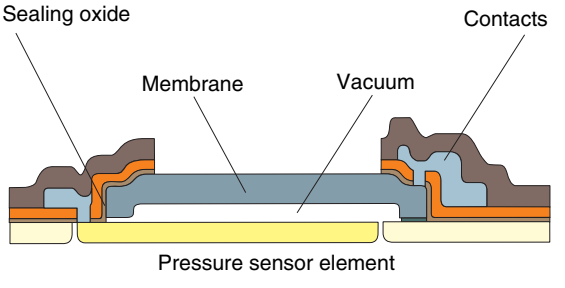


Figure 2 Cross section of a surface-micromachined capacitive pressure sensor.

case, etching the polysilicon layer and the sacrificial oxide forms moveable finger structures that are intertwined with matching counterparts that are fixed. Acceleration forces the moveable parts toward the fixed parts resulting in a higher capacitance, which is detected and evaluated by the on-board electronics (Payne *et al.* 1995). A more complex and much more sensitive device is the angular rate sensor or gyro (Link *et al.* 2005), as shown exemplarily in **Figure 3**. Together with the accelerometers they constitute complete inertial systems. The small capacitive signals achieved with these sensors present a challenge to the circuit designer and require low-noise techniques as discussed in Sections 1.16.3.1–1.16.3.3.

The widely sold silicon sensor is still the old-fashioned bulk-micromachined pressure sensor. Here the whole silicon wafer of about 500 µm is used to produce the moveable parts. Through backside etching a cavity is formed defining a thin membrane on the order of 5 to 10 µm at the surface of the silicon wafer. Piezoresistors implanted in the membrane are sensitive to deformations caused by exposing the membrane to a higher pressure (Bryzek 1993). The same process steps can be used to define a membrane for a flow sensor (Kersjes *et al.* 1996) or a large moving mass for high-precision accelerometers. The sensing elements in these cases are usually piezosensitive resistors arranged in a Wheatstone bridge configuration. This reduces parasitic effects and directly provides a voltage output.

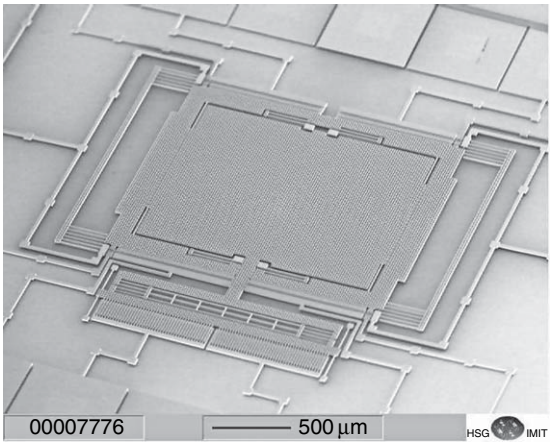


Figure 3 Microphotograph of a surface-micromachined capacitive gyroscope. (Courtesy: HSG-IMIT. Source: Link T, Simin I, Trachtler M, Gaisser A, Braxmaier M, Manoli Y, Sandmaier H 2005 A new self-test and self-calibration concept for micro-machined gyroscopes. *Transducers, Digest of Technical Papers*, pp. 401–4.)

1.16.2.3 Analog Electronic Interfacing

A sensor or a transducer converts a physical or a chemical effect into an electrical quantity. This can be a voltage, a current, a change in a resistance or capacitance or similar. For the purpose of further processing, these quantities have to be converted into an electric signal, usually a voltage. Due to very small sensor signals, an amplification and filtering must take place as well. Compensation, linearization, and calibration are then necessary to correct nonidealities that the signal might present. Ideally, this signal processing should be performed before transmission, thus avoiding signal distortion.

MOS analog integrated circuit design techniques make use of device-matching properties rather than relying on absolute values of devices. Whereas absolute values of resistors, capacitors, etc., are subject to process variations on the order of 10–30%, the matching errors of these elements are on the order of 0.1–0.5%. In addition, the voltage and temperature coefficients also match very well. This property of MOS processes has led to the wide application of the switched capacitor (SC) circuit technique that makes use of capacitor ratios as precise circuit elements. The charge collected on a capacitor can be nondestructively sensed due to the almost infinite input resistance of the MOS transistor gate and converted into a voltage. The corresponding circuit techniques are discussed in Section 1.16.3.2.

1.16.2.4 Digital Postprocessing for Electronic Interfaces

After analog signal processing, an analog-to-digital (A/D) conversion provides a digital output that is more appropriate for transmitting the data over longer distances. Further, a bus interface can be implemented, simplifying the interconnection of the sensor or a number of sensors to a microcontroller or a network. Although analog calibration is possible, it might be easier from the system point of view to do the calibration in the digital domain. Calibration is, apart from packaging, one of the less obvious cost-determining factors for a sensor. It has to be considered thus as a part of the system design and has to be dealt with early in the design cycle. Calibration is necessary to correct any offset, gain, linearity, or other errors such as temperature dependence of the sensor. The process requires subjecting the sensor to a number of well-known conditions and extracting the necessary parameters.

These parameters can be used to adjust the sensor signal or the readout electronics to eliminate or reduce the nonidealities.

Doing the correction in the digital domain makes it possible to use polynomial linearization functions or a table-look-up. Applications with extensive signal processing on the digital side or complicated control sequences would require a large chip area, if done in a standard cell application-specific integrated circuit (ASIC) manner. To avoid this and the long design time that such implementations would require, a configurable microcontroller that can be easily adapted to the application represents the most economic solution. Not only the size of the nonchip RAM and ROM is variable but the configuration and the kind of peripheral units are flexible as well (Lerch *et al.* 1995).

1.16.2.5 Peripheral System Components

The overall system design of sensing or actuating microdevices covers such a wide range of applications and demands operation over changing environmental conditions that the system design becomes a challenging task; microsystem as well as circuit designers can contribute to alleviate this task.

For portable or very small devices, for example, battery operation may be required. Considering the volume and the lifetime of batteries, low-voltage (LV) operation and low power dissipation are very desirable. In order to minimize overall power consumption, sensor and circuit concepts, system operating speed as well as analog and digital signal processing and partitioning should be selected properly. A telemetric data transmission might be desirable or necessary if a direct connection to the microsystem is not possible. If battery operation is not applicable, not only the information data but also the power for the device operation has to be transmitted via nongalvanic links. For power transmission, an inductive coupling is recommended for high power efficiency. Energy harvesting techniques might be applicable, which can then convert, for example, vibration energy into usable electrical power (Spreemann *et al.* 2006). For data transmission, besides radio frequency (RF) and carrier modulation, other media can also be used, e.g., infrared. For a reliable transmission, digital coding of the information is desirable since it is less susceptible to noise pickup and allows easy error detection.

1.16.3 Electronic Interfaces for Micromechanical Devices

1.16.3.1 Basic Limitations for Sensor Readout Electronics

Various noise mechanisms limit the performance of sensor readout electronics. In this section, two main noise sources, i.e., the mechanical and electrical noise, are considered and their influence on the performance of the readout electronics is investigated. Whereas the mechanical noise cannot be influenced by a circuit designer, the electronic noise can be improved by a good design choice of the interface circuit. An understanding of the sensor noise source is required in order to appropriately design the interface electronics, which should be smaller than the mechanical noise to have the full performance of the sensor system. Thus, some fundamentals of mechanical noise are discussed before covering the electrical noise sources.

1.16.3.1.1 Mechanical noise

Noise generated by temperature fluctuation occurs in any mechanical system. Commonly in large-scale sensors, this kind of noise is usually neglected. But for microelectromechanical (MEM) sensors having very small masses, these effects dominate the sensor performance. The mechanical noise sources can be distinguished into external and internal noise source, e.g., Brownian noise and phonon–phonon noise, respectively (Djuric *et al.* 1998). The most frequent noise is the Brownian motion that originates from the interaction of molecules with flexible parts of the sensing elements immersed in a gaseous or a liquid medium. This behavior can be described by the equation of a harmonic oscillator which is as follows:

$$m\frac{\mathrm{d}x^2}{\mathrm{d}t^2}+\xi\frac{\mathrm{d}x}{\mathrm{d}t}+kx=\mathbf{F}(t)+F(t) \qquad [1]$$

where $\mathbf{F}(t)$ is the deterministic excitation force, $\mathrm{F}(t)$ is the stochastic force excited from the surrounding molecules on the flexible part of the sensor, m is the mass, ξ is the damping coefficient, and k is the mechanical spring constant. Assuming that the mean of the stochastic force is zero, $\mathrm{F}(t) = 0$, and no external force $\mathbf{F}(t)$ is applied, the mean velocity $\bar{v} = \mathrm{d}\bar{x}/\mathrm{d}t$ vanishes. On the other side, the instantaneous velocity $v(t) = \mathrm{d}x/\mathrm{d}t$ is not zero as the sensor is excited constantly by the surrounding molecules. The Brownian force (Gabrielson 1993,

Gabrielson *et al.* 1998a) is given by the following equation:

$$F_B = \sqrt{4k_B T\xi} \quad [2]$$

which causes Brownian motion of the harmonic oscillator x_B and can be solved by combining eqns [1] and [2]. Solving the Laplace domain results in

$$x_B = \frac{\sqrt{4k_B T\xi}/(\omega_n^2 m)}{\left(\frac{s}{\omega_n}\right)^2 + \frac{1}{Q}\frac{s}{\omega_n} + 1} \quad [3]$$

where Q is the quality factor, k_B is the Boltzmann constant, and w_n is the natural resonance frequency. Recalculating the equivalent acceleration originating from the Brownian force gives the following equation:

$$a_B = \frac{\sqrt{4k_B T\xi}}{m} = \sqrt{\frac{4k_B T\omega_n}{mQ}} \quad [4]$$

Eqn [4] shows that a high Q factor and a huge mass are desirable, if the sensor is used as an accelerometer. A huge mass can be achieved by bulk micromachining being carved out of a full wafer. To further reduce the influence from the Brownian noise, the Q factor has to be increased. This can be achieved by suspending the proof mass in vacuum. But for high-performance accelerometers this is sometimes not enough. Therefore, other noise sources can become dominant and the different contributions have to be investigated and design rules for the sensor design need to be derived. These additional contributions are the following (Hosaka *et al.* 1995):

- Internal friction, which is often negligible with single-crystal materials.
- Change of entropy caused by heat flow.
- Phonon interaction with phonons from the supporting structures.

1.16.3.1.2 Electronic noise

As mentioned in Section 1.16.1, analog signals processed by integrated circuits are also corrupted by noise. These can be distinguished into two types of noise: device electronic noise and environmental noise sources. The first includes the offset as a direct current (DC) noise, the thermal noise, and the flicker or the so-called $1/f$ noise. The environmental noise comprises random disturbances that enter the circuit through the supply, ground lines, or through the substrate (Razavi 2001). In the following, the consideration of noise is restricted to thermal, $1/f$, and kT/C noise, with kT/C noise being the most dominant.

1.16.3.1.2.(i) Thermal noise

Similar to the mechanical system and its Brownian motion, thermal noise occurs in resistors by the random motion of electrons in a conductor. The average of these electron fluctuations can be measured as a voltage across the conductor being proportional to the absolute temperature. Commonly, the power spectrum density of the thermal noise can be assumed to be white. White noise follows from the definition of white light meaning a constant amplitude or power density for all frequencies. In both cases, this cannot be true as the power of such a signal would be infinite. In the case of thermal noise, the cutoff frequency is in the range of 70 THz for room temperature and thus the assumption of white noise holds for most practical applications.

Very often it is necessary to estimate the noise power at the output or the input of a circuit. Therefore, the thermal noise has to be modeled within the circuit. This can be done easily by using a voltage source in series with the resistor having a spectral density of

$$\overline{V_{nth}^2} = 4k_B TR \quad [5]$$

From eqn [5] it can be seen that a high temperature T and a resistance R excites a large voltage noise. Consequently, the resistance has to be small and the temperature reduced if a low thermal noise is desired. Alternatively, the noise can be modeled by a current source in parallel having the spectral density as

$$\overline{I_{nth}^2} = \frac{4k_B T}{R} \quad [6]$$

Depending on the circuit topology, one model may result in a simpler expression than the other. The calculation itself is done by using Kirchoff's law, as it is also valid for noise signals. But as a rule of thumb, the voltages and currents of uncorrelated noise sources are estimated by superposition and afterward added quadratically.

Not only resistors exhibit thermal noise, but also MOS transistors constitute thermal noise due to the resistive channel of the transistor in the active region. For a long-channel MOS device operating in saturation the thermal noise can be modeled by a current source connected between the drain and the source. The value of this output current noise is

$$\overline{I_n^2} = 4k_B T\gamma g_m \quad [7]$$

where γ is approximately $2/3$ depending on the technology used. Eqn [7] suggests that to reduce the

current noise of a transistor its transconductance g_m should be as small as possible. This is true if a transistor operates as a constant current source, whereas for an input transistor in a differential stage g_m should be high. This is because the input referred voltage noise reduces with large g_m.

To generate the drain current, a corresponding voltage with the value I_D/g_m has to be applied to the gate of the transistor. Dividing eqn [7] by g_m leads to

$$\overline{V_n^2} = \frac{4k_B T\gamma}{g_m} \quad [8]$$

1.16.3.1.2.(ii) Flicker noise For semiconductors, flicker noise is mostly generated between the gate oxide and the silicon substrate (van der Ziel 1978). Since the silicon crystal lattice ends at this interface, dangling bonds occur, giving rise to extra energy states. Charge carriers that move across this interface can be trapped in these energy states and later randomly released, generating flicker noise. In circuits, the flicker noise is modeled as a voltage source in series with the gate and is given by the following equation:

$$\overline{V_n^2}(f) = \frac{K}{C_{ox} WL}\frac{1}{f} \quad [9]$$

where K is a process-dependent parameter and can vary widely for different devices in the same process (Johns and Martin 1997). The variables W and L are the width and the length of the transistors, and C_{ox} represents the gate capacitance per unit area. To achieve the equivalent noise current, eqn [9] has to be multiplied by g_m^2, leading to

$$\overline{I_n^2}(f) = \frac{K}{C_{ox} WL}\frac{1}{f} g_m^2 \quad [10]$$

The trap-and-release phenomenon, as eqn [9] reveals, occurs more often at low frequencies, and therefore flicker noise is also called $1/f$ noise. An important attribute of flicker noise is its inverse dependence on the area of the transistor, which suggests that to decrease the $1/f$ noise, the $W \times L$ of the transistor must be maximized. In addition, P-channel-metal-oxide-semiconductor (PMOS) exhibit less flicker noise than their N-channel-metal-oxide-semiconductor (NMOS) counterparts as there are less electrons in the conduction band and are thus less likely to be trapped. In **Figure 4**, both spectral densities of $1/f$ and thermal noise are plotted in one graph. The intersection point is referred to as the $1/f$ noise corner frequency and

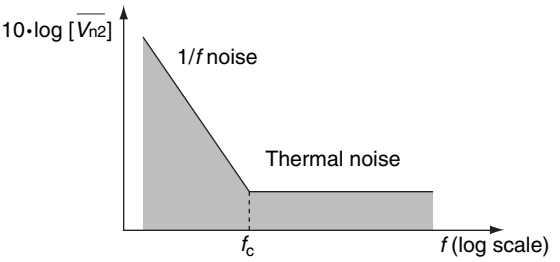


Figure 4 Spectral densities of both flicker noise and thermal noise are drawn on the same axis. The intersection point is referred to as the $1/f$ noise corner frequency, f_c, and indicates what part of the band is corrupted dominantly by flicker noise or thermal noise.

serves as a measure of which part of the band is dominantly affected by flicker noise. Eqn [9] with eqn [10] and solving for the $1/f$ noise corner frequency, f_c, leads to

$$f_c = \frac{K}{C_{ox}}\frac{g_m}{WL}\frac{1}{4k_B T\gamma} \quad [11]$$

This result indicates that the corner frequency is a function of the device dimensions and the bias current.

1.16.3.1.2.(iii) kT/C noise The kT/C noise occurs only in SC circuits and is not a fundamental noise source like those mentioned previously. Commonly, a switch in an electrical circuit is realized through a transistor. This transistor has a high resistance if it is switched off and a low resistance, $R_{on,}$ if it is switched on. While switching the capacitor, the thermal noise of the on resistor R_{on} is stored on the capacitor along with the instantaneous value of the input voltage. The on resistor and the capacitor together form a low-pass filter if they are shorted to the ground. As it is possible to estimate the noise power stored on the capacitor, the spectral density over the noise bandwidth has to be integrated. Using eqn [5] for the spectral noise of the on resistor, the total output noise is

$$P_{kT/C} = \int_0^\infty \left|\frac{1}{R_{on}C + 1}\right|^2 \overline{V_{nth}^2}\, df \quad [12]$$

Eqn [12] reveals that a sufficiently large capacitor C reduces the kT/C noise. Further it is independent on the used sampling frequency f_s and thus it describes a fundamental limitation for SC circuits. This restriction can be reduced by increasing the capacitance of C. But this also requires more current to charge the capacitance, if the same bandwidth is

desired. Thus, the kT/C noise can be reduced only by accepting a higher power consumption.

1.16.3.2 Capacitive Readout Electronics

As the dimensions of micromachined inertial sensors are small, the signals from these devices are also very small. As an example, the displacement of a proof mass for an accelerometer with 10 kHz natural resonant frequency and a 5 mg input acceleration, where 1g means the gravity acceleration, is only one-tenth of an angstrom. For gyroscopes, the displacement, which has to be detected, is even smaller (Clark and Howe 1996). In the literature, many different techniques are shown for sensing such small displacements, e.g., tunneling tips (Zavracky *et al.* 1993), magnetic, capacitive, or optical transducers (Garcia-Valenzuela and Tabib-Azar 1995).

While each approach has specific advantages, none of them outperforms the other. So, the choice for a specific technique is guided primarily by other factors such as available experience or the number of fabrication steps required. From the last point of view, capacitive interfaces have several appealing features as in most micromachining technologies where no or minimal additional processing is needed (Boser 1997). Additionally, capacitors can be used as sensors and actuators having high sensitivity. Further, the capacitors are designed through lithography having well-defined properties and are insensitive to temperature. On the other hand, the dual operation of capacitors both as sensors and as actuators increases the requirements on the interface design. As the readout of the capacitor itself excites a force on the capacitor plates, active shielding has to be considered. Especially for two-chip hybrid realizations consisting of one chip for the sensor and one chip for the electronics, wiring and stray capacitances can reduce the signal strength and thus the performance of the entire sensor device. Hence, alternative transducers (e.g., piezoresistive) are frequently attractive in situations where parasitics due to wiring are dominating. But, this comes with a higher temperature sensitivity and power consumption of such transducers (Boser 1997).

1.16.3.2.1 Capacitive-to-voltage converters

When using capacitive sensors, it is necessary to convert a change in capacitance ΔC to a corresponding voltage. Charge integration is the method of choice due to its lower susceptibility to parasitics (Boser 1997). **Figure 5** shows two different architectures of capacitive sensing.

In **Figure 5(a)** a differential capacitive-to-voltage converter based on charge integration is presented. The functional concept can be described as follows. An alternating current (AC) or DC signal V_m is applied to the sense capacitances. The resulting charge will be transferred from the sensor to the feedback capacitors C_i of the amplifier by integration. At the output a differential voltage V_0 will appear depending on the variation of ΔC (Baltes *et al.* 2004):

$$V_0 = \frac{2\Delta C}{C_i} V_m \qquad [13]$$

The advantage of this design is that the signal V_m is easy to realize. It is important to note that the differential amplifier must be able to handle large input common-mode signals. In case the common-mode range of the amplifier cannot accomplish the voltage step caused by V_m, an input common-mode feedback (ICMFB) is absolutely essential. Reducing the voltage step V_m to a lower value is not desirable, because to improve the signal-to-noise ratio a large voltage step is required.

In **Figure 5(b)** a single-ended capacitive-to-voltage converter is shown. The charge integrator is connected to the middle node being the sensor proof-mass. The

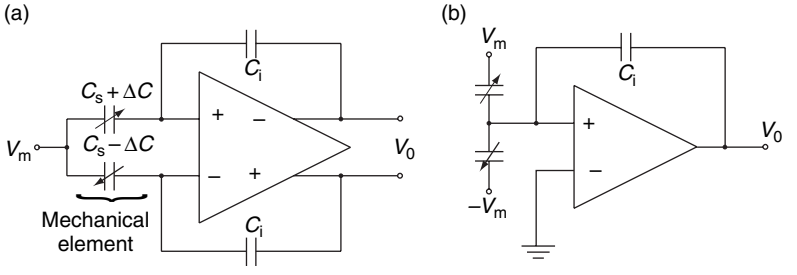


Figure 5 Draft of a differential capacitance-to-voltage converter based on (a) charge integration and (b) single-ended capacitive-to-voltage converter.

AC or the DC signal is applied to the ends of the capacitive bridge with two out-of-phase waveforms. In this case the input of the amplifier is at the virtual ground and ICMFB is not needed (Baltes *et al.* 2004). Unfortunately, this architecture requires a precise generation of two AC drive signals to ensure high output stability (Lemkin and Boser 1999).

Besides the easier implementation of the carrier signal V_m, the differential capacitive-to-voltage converter has some other important advantages such as suppression of even-order harmonic distributions in the output signals (Razavi 2001) or improved power supply rejection ratio (Lemkin and Boser 1999).

1.16.3.2.2 Noise in capacitive-to-voltage converters

Capacitive interfaces have the attributes to measure very small capacitance changes and hence very small displacements. The following amplifier can boost these signals to manageable amplitudes. Unfortunately, not only the signals but also the undesired noise are amplified. Since all physical systems suffer from low-frequency flicker noise as mentioned in Section 1.16.3.1.2.(i), two different techniques have been developed over the years. The first one is called correlated double sampling (CDS) and is a discrete-time technique. The advantage is that its output can be easily digitized by using an analog-to-digital converter (ADC). An additional pre-filtering in the adjacent stage is therefore not needed. Further, no additional ICMFB is necessary reducing the circuit complexity. The second one is referred to as chopper stabilization (CHS), which is a continuous-time technique. To digitize its output, a prefilter in the adjacent stage is required. Both techniques can reduce low-frequency noise and the offset of the amplifier. The functional concept of both techniques are discussed in the following section.

1.16.3.2.2.(i) Correlated double sampling As already mentioned above, CDS is a discrete-time technique and has the advantage that its output signal can be digitized with an ADC without the need for an additional antialiasing prefilter (Baltes *et al.* 2004). The basic idea of this technique is to subtract two sampled signals. The first containing only the error, such as noise and offset, and the second containing both the error and the desired signal (Enz and Temes 1996). If the error has strongly correlated values in both samples, the error can be eliminated by subtracting these two samples. The condition that the two error samples have to be strongly correlated indicates that an effective cancellation is possible only if the frequencies of the error components are much lower than the sampling frequency f_s.

In **Figure 6** the functional concept of CDS is described. The process requires three phases. In phase 1 the reset takes places. All the nodes are connected to the ground except for the node M of the mechanical element. At node M a voltage V_1 is applied and the sense capacitors C_{s+} and C_{s-} are charged to the difference between V_1 and the ground. During this first phase the two integration capacitors C_{i+} and C_{i-} are discharged. In the second phase the nodes are taken from the ground and the amplifier is active. Error components such as offset, flicker noise, switch kT/C noise, and charge injection are now stored on the capacitors C_{H+} and C_{H-}. At the beginning of phase 3, a signal with magnitude of V_1-V_2 is applied to the node M. The errors stored in C_{H+} and C_{H-} of phase 2 are now in series with the output signal of phase 3 and are subtracted. This results in an output signal containing the signal and the kT/C error generated by the last sampling step between phases 2 and 3.

When compared to the CHS technique being outlined in Section 1.16.4.2.2.(ii), no antialiasing filter is necessary if an A/D conversion is required. Unfortunately, this technique suffers from several drawbacks. Like any sampled system, CDS suffers from noise folding, which limits the resolution. The noise density can be reduced by increasing the load capacitance, which results in a higher power

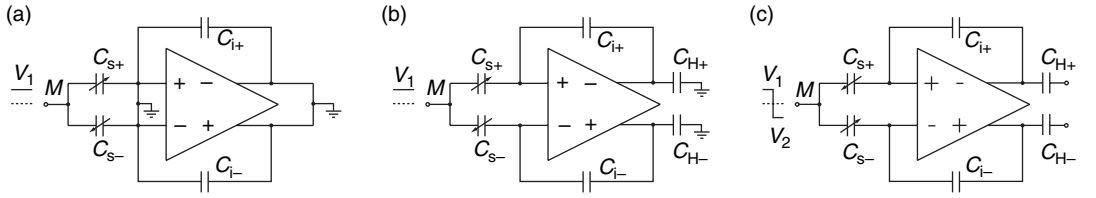


Figure 6 Correlated double sampling (CDS) has three phases. In phase 1 (a) the reset takes place. In phase 2 (b) the offset and low-frequency noise are sampled. In phase 3 (c) the input is applied and previously sampled errors are subtracted (Baltes *et al*. 2004).

consumption in order to preserve the bandwidth (Baltes *et al.* 2004).

1.16.3.2.2.(ii) Chopper stabilization Like CDS, the CHS technique can be employed to reduce low-frequency noise and amplifier offset. Unlike the CDS process, the CHS does not use sampling and is therefore a continuous-time technique. **Figure 7** shows the operation in case of a capacitive sensor readout (Baltes *et al.* 2004). A sinusoidal voltage V_m with frequency f_m is applied to the center node and modulates the signal to a higher frequency where there is no $1/f$ noise. A displacement of the mass will change ΔC and produces a current difference Δi that flows into the integrating capacitor C_i. As a result, a corresponding voltage proportional to ΔC appears at the output with a frequency of f_m. Assuming that the signal of the sensor is band limited at least to half of the modulation frequency f_m, no signal aliasing occurs. If the signal is modulated with a frequency f_m, much higher than the $1/f$ corner frequency, flicker noise and offset will not disrupt the signal. After the signal is amplified it can be demodulated back to the original frequencies, thereby modulating the low-frequency noise to higher frequencies in a band out of interest. To recover the original signal in an amplified form, a low-pass filter must follow with a cutoff frequency slightly above the signal bandwidth of the sensor and with a high slope to eliminate the noise that is shifted to higher frequencies and higher harmonics.

To ease the generation of the modulation signal, a square wave instead of a sine wave can be used for modulation. In this case the signal power also appears at multiples of the modulation frequency f_m. Further, this increases the requirements of the charge amplifier, as it has to provide bandwidth and slew rate.

1.16.3.2.3 Slope converter

Another very simple conversion method is the single-slope converter. This converter works in two phases. In the first phase, a reference capacitor C_L is charged by the unknown reference current source. As long as the voltage over the reference capacitor C_L is above a reference voltage U_{ref} the output of the comparator is 1. The output of this converter is a pulse-width-modulated signal of the unknown current signal. This method, which is well-known for ADCs, can also be used with slight modifications as a capacitive interface. The reference current source charges the sensing capacitance cycle. Thereafter, the sensor capacitance is shortened which starts the next conversion cycle. This is also shown in **Figure 8**. One major advantage of this approach is its simplicity and that the output signal can be easily digitized by a counter. But the resolution is limited by the accuracy of the reference current source and it is feasible only for slowly varying sensor signals.

1.16.3.2.4 Example: Integrated interface circuit for a capacitive pressure sensor

The capacitive pressure sensor consists of an array of single circular pressure-sensitive elements connected in parallel and of similarly built reference elements. An exemplary circuit design and a chip photo are

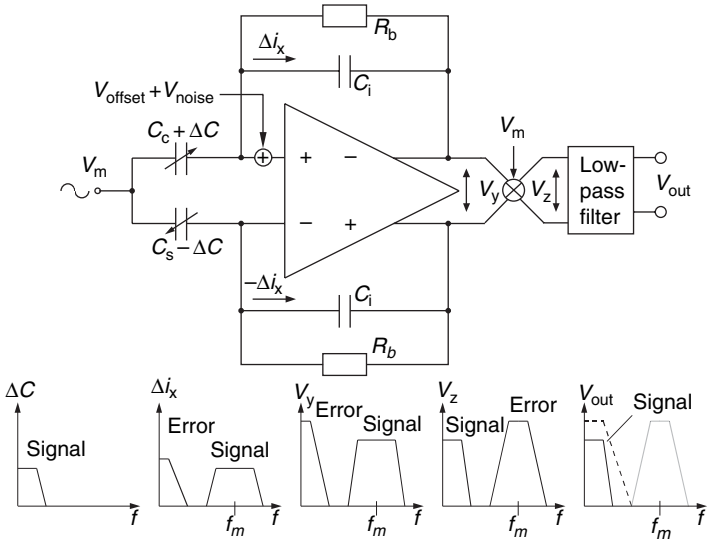


Figure 7 The functional operation of chopper stabilization (CHS) for capacitive sensor readout (Baltes *et al.* 2004).

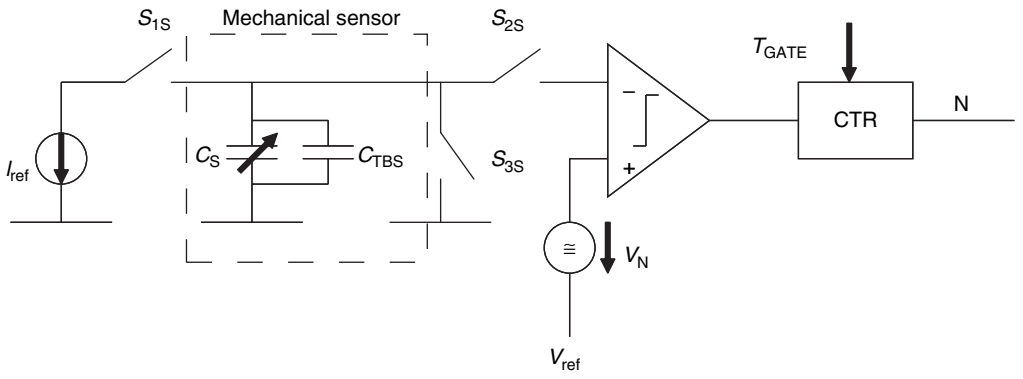


Figure 8 Single-slope converter with current source, comparator, and counter used for digital output.

shown in **Figure 9**. The capacitor is formed by a fixed electrode in the substrate generated by implantation and a moveable membrane of polycrystalline silicon above. The cavity under the membrane is obtained by etching a sacrificial layer of silicon dioxide via etch channels and later by vacuum sealing. The height of the cavity is less than 1 µm while the membrane diameter is in the range of 100 µm, depending on the pressure range. These reference elements are constructed in an identical manner with the only difference being the thickness of the membrane. The output voltage signal is proportional to the quotient of the capacitance of the sensor and the reference element. Thus, parasitic effects are suppressed.

The signal conditioning is based on the SC technique (Hosticka *et al.* 1977). The capacitor of the pressure sensor C_S and its reference sensor C_R is implemented in the first stage of the readout electronics. In this stage a signal conversion from the capacitance to voltage occurs as discussed in Section 1.16.3.2.2.(i). Because of the negative influence in the case of clock feedthrough and current leakage, the differential path technique is used, so the pressure sensor is divided into two parts.

In the second stage a subtraction of he zero pressure signal, a differential to single-ended conversion, and an amplification are performed. An integrated sample and hold function supplies a continuous output signal. The excitation voltage is generated by a

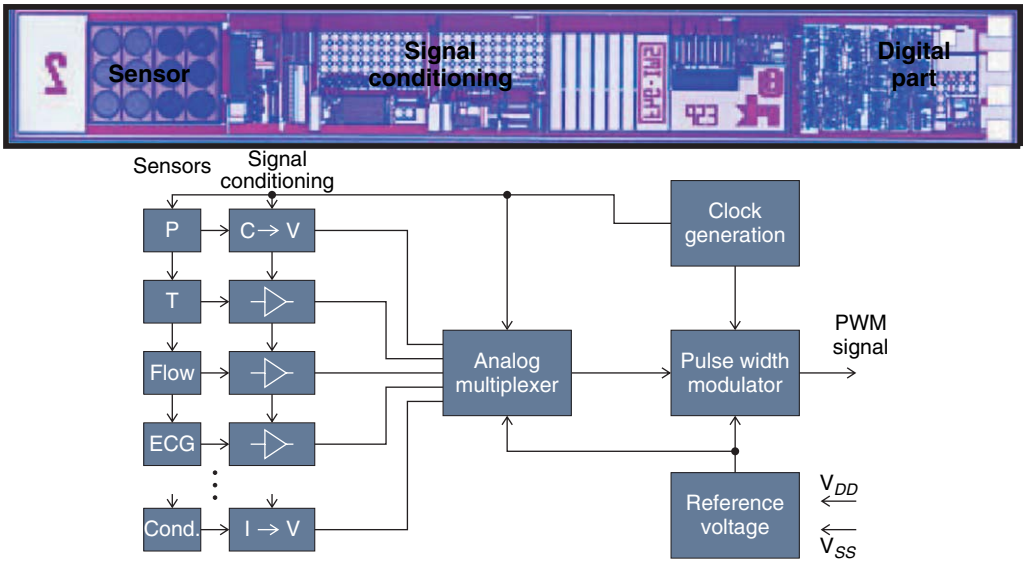


Figure 9 Pressure sensor chip and its block diagram consisting of the pressure sensors, the signal conditioning, and the digital electronics.

matched resistor network that leads to a full ratiometric signal conditioning. A central bias circuit and an oscillator that generates a four-phase clock for the SC amplifiers complete the system to a three-terminal device (Schmidt *et al.* 1995).

1.16.3.3 Sigma–Delta Interface

As the quality of the mechanical systems improves, the performance of the interface electronics becomes the limiting factors on the performance of the whole system. Being able to improving the performance of microelectromechanical systems (MEMS) even further, it is not possible to separately examine the mechanical and the electrical part of the system. In circuit design, the application of feedback is well-known for improving, e.g., the linearity or the temperature stability of an amplifier. Due to the increasing demand for sensors with digital output, the development of novel interface circuits focuses on this new aspect. The sigma–delta conversion is well-known in the field of electrical A/D conversion. It offers many advantages compared to Nyquist rate converters being relaxed circuit complexity, such as precision requirements and often lower power consumptions by using oversampling and its noise shaping behaviour. This approach is carried over to the sensor readout in the following section.

1.16.3.3.1 Fundamentals of sigma–delta modulator

An A/D conversion of the sensor signals is desirable so as to transit them over long distances and to perform complex signal processing. Depending on the signal bandwidth, different ADC concepts can be implemented. Although successive approximation ADCs are usually a best fit for sensor applications, a sigma–delta modulator (**Figure 10**) has significant advantages making it also a very good concept (Geerts *et al.* 2002).

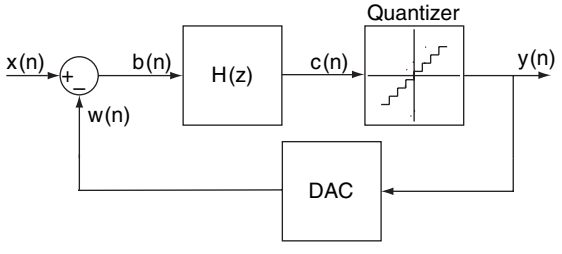


Figure 10 Block diagram of a sigma–delta modulator.

1.16.3.3.1.(i) Oversampling Oversampling a signal, which means sampling the signal with a rate higher compared to the Nyquist rate, reduces the quantization noise. If a signal is sampled at the Nyquist frequency, all of its quantization noise power is folded in the frequency range between $-f_s/2$ and $f_s/2$, which is equal to $-f_b$ and f_b, respectively. Since the quantization noise power σ_e^2 depends only on the number of quantizer bites, an increase of the sampling rate beyond the Nyquist rate spreads the noise power over a wider band and hence decreases the power in the band of interests, as eqn [14] reveals. The quantization error in sigma–delta modulators can be modeled as an additive noise source. From the quantization noise power σ_e^2, which is $\Delta^2/12$ (Jarman 1995), where Δ is the size of the least significant bit of the internal quantizer, the power spectral density (PSD) b_e can be calculated as follows:

$$b_e = \frac{\Delta}{\sqrt{12 f_s}}. \qquad [14]$$

1.16.3.3.1.(ii) Noise shaping As already mentioned, the main property of a sigma–delta modulator is its noise shaping and can be explained as follows. By feeding the output signal back to the input via a feedback loop, it is possible to suppress the quantization error. In **Figure 11** a block diagram of a sigma–delta modulator is shown. Compared to **Figure 10** the quantizer is replaced by an additional noise source e. Under the assumption that the quantization error is not correlated with the input signal, the error can be modeled as white noise.

From **Figure 11**, the following expression can be found for $y(n)$:

$$y(n) = x(n)\frac{1}{1+\frac{1}{H(z)}} + e(n)\frac{1}{1+H(z)} \qquad [15]$$

$$y(n) = \mathrm{STF}(z)x(n) + \mathrm{NTF}(z)e(n) \qquad [16]$$

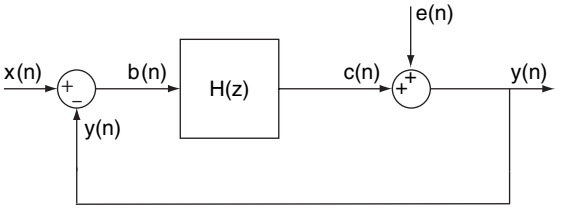


Figure 11 Block diagram of a sigma–delta modulator, where the quantizer is replaced by the white noise model.

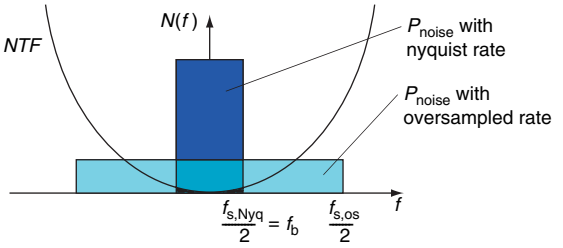


Figure 12 The noise transfer function (NTF) shifts the undesirable noise out of the band of interest to higher frequencies.

As eqn [16] reveals, there are two transfer functions. One for the signal called signal transfer function (STF) and one for the undesirable noise called noise transfer function (NTF). To obtain the best performance, $H(z)$ must have a large amplification at low frequencies and a low amplification at high frequencies and hence shifting the noise to higher frequencies in a band out of interest. Usually $H(z)$ is realized as a chain of integrators (Aboushady and Louerat 2002).

The noise-shaping attribute of a sigma–delta modulator can be illustrated graphically as in Figure 12. Oversampling a signal spreads the quantization noise over a wider band. Since the total noise power is still the same, as if sampled with the Nyquist rate, the noise spectral density in the band of interest is greatly reduced. The additional noise-shaping feature of a sigma–delta modulator shifts the undesirable noise to higher frequencies.

1.16.3.3.2 Electromechanical sigma–delta modulator

Due to the continuous-time nature of the mechanical element, the obtained signal has to be quantized in time and amplitude. Whereas the quantization is performed by the internal quantizer in the sigma–delta loop, the time quantization (sampling) can be done in the sensor front end instead. Figure 13 shows a block diagram of a single-bit sigma–delta modulator. In the forward path, a mechanical element, a capacitance-to-voltage converter, a gain, a lead compensator, and the binary quantizer are seen. The feedback circuit consists of a voltage-to-force converter that generates a corresponding force to keep the proof mass in its nominal position. Thus, the mechanical stress of the springs are reduced, which improves their linearity, as the influence of the cubic spring term is reduced. Lower stress can also increase the lifetime of the sensor. Alternatively, a mechanical resonator with a softer spring or larger mass can be applied, yielding a higher static resolution.

The feedback architecture also provides a linearization of the sensor's STF and increases the usable bandwidth. These advantages and the attributes of sigma–delta modulators, in general, supplement each other and a high-performance capacitive interface with direct digital output, noise-shaping feature, low chip area and with it low cost and low power consumption can be derived. In contrast, a more complex system design has to be considered due to potential instability of the system. This is because typical mechanical sensor can be modeled as resonators, which feature two dominant conjugate complex poles. Thus, the system in Figure 13 shows internal state variables that are difficult to control. One solution is the lead compensator, which gives a positive phase shift and thus stabilizes the system concerning the two poles of the sensor (Petkov and Boser 2005).

1.16.3.3.3 Sigma–delta interface conclusion

As mentioned previously, electromechanical sigma–delta modulators have several advantages compared to the common open-loop sensor evaluation. These are the increased linearity, bandwidth, and reduced circuits requirements, among others. But on the other hand, the circuit designer has to perform a complex circuit design due to the possible

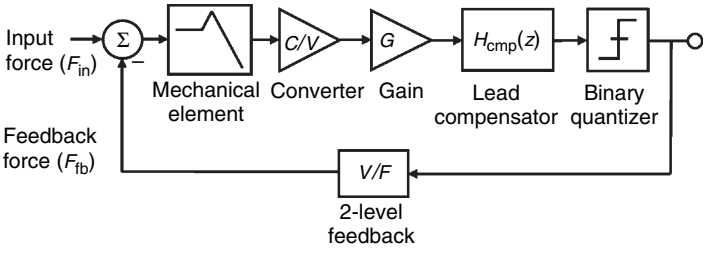


Figure 13 Schematic of an electromechanical sigma–delta modulator.

instability of the electromechanical sigma–delta modulators, which reduces the attractiveness of this novel approach.

1.16.4 Electronic Interface for Biomedical Applications

Over the past few decades, the development of implantable, functional electrical stimulation (FES) or functional neural stimulation (FNS) have emerged as promising approaches in the treatment of neural or muscular disabilities. Among others, the cochlear implant (McDermott 1991), cardiac pacemaker, visual prosthesis (Zrenner 2002), deep brain stimulators (Kuchta 2004), implants for bladder or sphincter control, or stimulators for hand or extremity control are some of the prominent examples. Today's neural prosthesis are systems that try to restore the functionality of disabled parts of the biological nervous system through electrical stimulation. Therefore, they selectively generate action potentials in nerve cells by electrical stimulation, which yield, e.g., to the contraction of a muscle or the excitation of the optic nerve. The stimulation is executed in a coordinated manner, so as to generate functional movement, visual, or acoustic reception.

Beyond the purely stimulating systems, the interest for general analysis of biosystems has increased over the last few years. Exemplary applications come from neurosciences, e.g., the understanding of neural communication on cellular-level chemical processes amongst neurons, rapid, diagnosis, or cell-based pharma screening (Aziz and Genov 2006, Eversmann 2003, Harrison *et al.* 2006). These sensor arrays for noninvasive neural monitoring from living cells are becoming a key tool in neurosciences for studying and understanding biological systems and neural networks. This is because they allow the study of an enormous amount of difficult to investigate neurophysiological phenomena, e.g., in sensory–motor systems, hippocampus, auditory nerve, and visual cortex. Thereby, the availability of the vast amount of data recorded simultaneously from many distributed sites is necessary in studying and analyzing heterogeneity across neural populations.

1.16.4.1 Electronic Nerve Signal Recording

Over the last few years, scientists from neurology and also clinicians have started to use implantable, MEMS by means of arrays with a large number of electrodes to simultaneously observe the activity of many neurons in the brain; knowledge gained from these studies increasingly yields to clinical and neuroprosthetic applications.

By observing, for example, the potential, or spikes, of many neurons in a certain region of the brain, it is now possible to obtain sufficient information to electronically predict hand trajectories in real time for reaching tasks. Experiment have shown that the development of neuroprosthetic devices becomes possible, which are controlled directly by thoughts, i.e., the brain–computer interface is becoming reality.

Thereby, the microintegration of both the electrodes and the readout electronics is a key objective for the successful evolution of such systems; currently, data are often recorded from implanted multielectrode arrays using bundles of thin wires and connectors that are mounted onto the patient's head; all electronics for amplification and recording are externally realized, i.e., outside of the body. This presents three major barriers to the development of neuroprosthetic devices with practical use:

(1) The transcutaneous connectors give rise to infection.
(2) External noise and interfering signals can couple more easily to the wires that convey the weak neural signals ($<500\,\mu V$) from high-impedance electrodes ($>100\,k\Omega$).
(3) The cable, connector, and external electronics are typically large and inconvenient to handle, when compared to the small ($<5\,mm^2$) electrode arrays.

In order to eliminate these problems, data from the numerous implanted electrodes must be preprocessed and wirelessly transmitted out of the body. This requires integrated microelectronics at (each of) the recording sites, which amplifies, preprocesses, and digitizes the neural signals from all electrodes. Such circuitry must also be powered wirelessly, since rechargeable batteries are large and have limited lifetimes, and an exchange of batteries is not possible as it would require surgery. Consequently, low-power operation ($<<100\,mW$) is essential, as for any implantable electronics, because only a limited power budget can be provided wirelessly, and even more critical, a temperature rise of 1°C may harm the tissue (Harrison and Charles 2003).

1.16.4.1.1 General challenges for nerve signal recording

A neuron that is not sending a signal, is at rest; then the inside of the neuron is negatively charged relative to

the outside, which is achieved by different ion concentrations. The balancing of the concentrations to an equilibrium is prevented by ion channels in the cell membrane, which can be closed or opened. The resting membrane potential of a neuron is approximately −70 mV. The action potential occurs when a neuron sends information via an axon, away from the nerve cell body. This spike firing is caused by an exchange of ions across the neuron membrane; therefore, an external stimulus causes sodium channels to open. Since the inside of the neuron is negative relative to the outside, sodium ions (Na^+) flow into the neuron. Consequently, the neuron becomes depolarized. When the opening of the potassium (K^+) channels are delayed, they reverse the depolarization. Almost at this time, the sodium channels start to close. This causes the action potential to return toward −70 mV (repolarization). Since the potassium channels stay open slightly longer, the action potential reaches −70 mV (hyperpolarization). Then, the ion concentrations gradually go back to resting levels and the cell returns to −70 mV. The external stimulus for firing the spike is critical, since a neuron will fire an action potential only if the depolarization reaches about −55 mV. This is called the threshold. If the neuron does not reach this threshold level, consequently no action potential will fire. This scheme for action potentials is indicated by **Figure 14**. These elementary neural signals are thus temporal peaks of the intracellular voltage relative to its environment. When recorded extracellularly, neural action potentials typically show amplitudes up to 500 μV, with a spectral content between 100 Hz and 7 kHz (Najafi and Wise 1986); in contrast, the so-called low-frequency local field potentials (LFPs) can be as large 5 mV, while containing signal energy even below 1 Hz (Harrison and Charles 2003). Even more challenging, signals typically recorded from electroencephalogram (EEG), electrocardiogram (ECG), and electromyogram (EMG) waves outside the body may be in the lower microvolt range. A major hindrance for direct measurement of these biopotentials are huge DC signals on top of the small signals: due to electrochemical effects at the electrode–tissue interface, offsets in the range of several hundred microvolts to several volts can arise across differentially recorded electrodes (Ferris 1978, Najafi and Wise 1986, Wise *et al.* 2004).

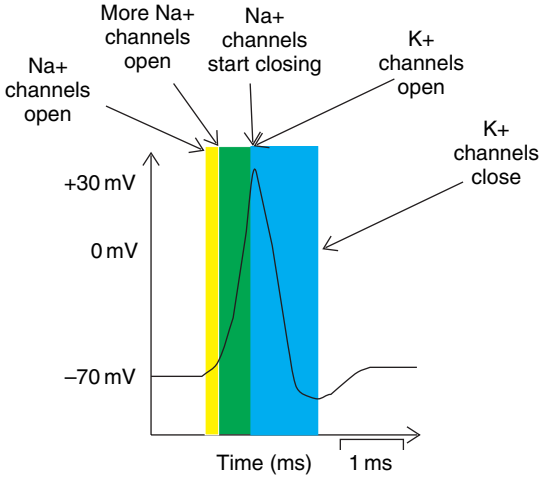


Figure 14 Illustration of neural action potential.

In conclusion, the common challenge of all biopotential readout circuits is the detection of very small, low-frequency signals, combined with high common-mode rejection ratio (CMRR) and sufficient blocking (high-pass filter characteristic) for the DC offset potentials, which are orders of magnitudes larger.

1.16.4.1.2 Circuit concepts for nerve signal recording

As mentioned above, the design of a readout circuit for nerve signal recording requires most of all high sensitivity, i.e., low noise and DC signal blocking combined with high CMRR in order to avoid the influence of the large DC offsets at the electrode–electrolyte interface. A high-pass filter with very low cutoff frequency is the obvious solution, but low cutoff in the range of tens of hertz requires very large capacitor–resistor values. Most early and even some recent integrated bioamplifier designs made use of large off-chip capacitors in the nanofaraday range to achieve this low corner frequency (Martins *et al.* 1998, Steyaert *et al.* 1987). But obviously, this approach is not feasible for large numbers of sites nor for implanted systems.

Generally, there are two dominant methods for neural recording: current mode readout and voltage mode readout (Aziz and Genov 2006). The former performs chemical neural activity recording, while the latter monitors electrical neural activity. Current-acquisition systems use potentiostats, which record (the change of) chemical concentrations by measuring reduction–oxidation currents between the electrolyte and the electrodes. Thus, the current-acquisition potentiostat is able to simultaneously transduce neurotransmitter activity at multiple locations; the neural interface acquires oxidation–reduction currents generated at the surface of each sensing electrode. The voltage-acquisition neural

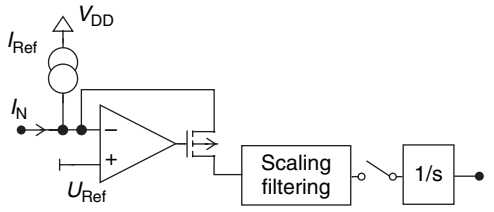


Figure 15 Current-acquisition neural interface. (Source: Aziz J N Y, Genov R 2006 Electro-chemical multi-channel integrated neural interface technologies. *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 2201–4.)

interface monitors a 2D action potential field, e.g., within a brain tissue area under observation.

Recently, an increasing number of integrated potentiostats have been presented (Kakerow *et al.* 1995, Murari *et al.* 2005), which achieved impressive low-power operation and high dynamic range in the lower picoampere range. An exemplary current-acquisition neural interface is schematically shown in **Figure 15** (Aziz and Genov 2006). Usually, for amperometric sensing over electrodes, potentiostatic instrumentation is required (which allows current measurement at regulated potential). Each potentiostat recording channel features a working electrode at the virtual ground, and a reference electrode, which applies the redox potential. The potentiostat provides the voltage for driving the redox reactions at the sensor with respect to the reference in the bath and a counter electrode on the electrode array. Since the expected currents are in the range of nanoamperes to picoamperes, they do not severely affect the reference electrode. Thus, the principally required three-electrode system is most often reduced to a two-electrode system (Aziz and Genov 2006) and the large counter electrode is used instead of a currentless reference. To provide some flexibility for the potentiostat readout, the reference voltage should be programmable, while the detectable currents need a large dynamic range. Also, the filtering behavior must be adjustable in order to acquire the signals under different working conditions (Aziz and Genov 2006, Eversmann 2003).

The circuit in **Figure 15** illustratively works as follows: in order to convert the bipolar input signal into a unipolar form, it is shifted by summing with a constant reference current I_{ref}. As voltage measurement needs high impedance measurement, so does current measurement require low input impedance of the measurement system. This is achieved by the transconductance amplifier that drives the PMOS transistor in **Figure 15**. The acquired signal current is then conducted into normalization circuit, which scales the different signals into a unified range of 0.1 μA. This normalized unipolar signal is fed into an antialiasing low-pass filter. The capacitive feedback amplifier at the end of the channel is used as current-to-voltage converter. Several of these signals can subsequently be multiplexed into an output word (Aziz and Genov 2006) or further processed in parallel.

In opposition to the current measurement, the more commonly used voltage-acquisition microsystems record neural action potentials by measuring the extracellular potential field generated by action potentials as explained in Section 1.16.4.1.1 (Aziz Genov 2006, Harrison *et al.* 2006, Wise *et al.* 2004). Thereby, most existing designs trade between acquisition channel circuit area and thus the number of parallel channels against intrinsic circuit noise. One of the most often used interface circuits for neural spike detection is shown as a simplified circuit in **Figure 16**. It features a preamplifier, which is capacitively AC-coupled in order to filter the DC offset through polarization at the electrode–electrolyte interface (Harrison and Charles 2003, Olsson *et al.* 2002, Wise *et al.* 2004). This amplifier ideally shows a high-pass transfer function:

$$A_{cc}(i\omega) = -\frac{R_{fb}C_{in}}{\frac{1}{i\omega} + R_{fb}C_{fb}} \qquad [17]$$

Thus, a zero gain at DC is obtained and the midband gain is set by the capacitors to:

$$A_{mb} = -\frac{C_{in}}{C_{fb}} \qquad [18]$$

and a low cutoff frequency at:

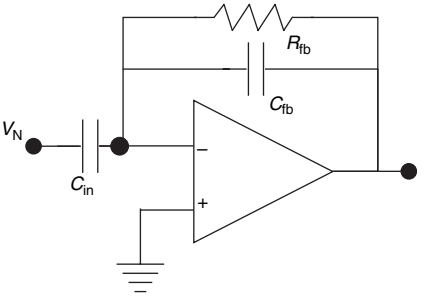


Figure 16 Capacitively coupled neural readout amplifier. (Source: Olsson R H, Gulari M N, Wise K D 2003 A fully-integrated bandpass amplifier for extracellular neural recording. *Proceedings of the IEEE Engineering in Medicine and Biology Symposium*, pp. 165–8.)

$$f_{\text{lowcut}} = \frac{1}{2\pi R_{\text{fb}} C_{\text{fb}}} \quad [19]$$

For realistic integrated capacitances in the picofaraday range, in order to obtain very low cutoff frequency below $\leq$100 Hz, the feedback resistor must be in the gigaohms range. Therefore, the feedback resistors are usually realized by specially biased transistors, e.g., MOS in the subthreshold range (Olsson *et al.* 2002). This capacitive feedback scheme in its various implementations (Harrison and Charles 2003, Olsson *et al.* 2002, Wise *et al.* 2004) allows a precise amplifier, since it is set by the capacitor ratio in eqn [18]; these can be very precisely achieved in an integrated CMOS process. The large feedback resistors also form a DC path, which supports a lower output offset voltage. Thus, the output offset approximately equals the input offset of the amplifier. This is important, first in order not to saturate the preamplifier by too large offsets, and second this is important in time-multiplexed systems, where the multiplexer otherwise could see large switching transients and thus might saturate.

Finally, a recently reported approach to overcome the limitations of this usually employed voltage readout in **Figure 16** is discussed. Among others, Harrison and Charles (2003) has shown that the feedback capacitor may degrade the input referred noise, and furthermore lowers the input impedance of the amplifier. High input impedance is, on the other hand, mandatory for voltage readout. Consequently, Gosselin *et al.* (2006) proposed a preamplifier with active rather than passive DC rejection, thus avoiding the need for passive high-pass-filter (HPF) with very low cutoff frequency, large resistors, etc. It consists of a low-noise operational transconductance amplifier with an active feedback loop that rejects any DC offset. An illustrative schematic is shown in **Figure 17** (Gosselin *et al.* 2006). The integrator time constant is set by a small capacitor C_i and a very large MOS resistor, implemented as in Harrison and Charles (2003). Beyond its advantages, the proposed amplifier is obviously more prone to noise, since more active devices are seen at the electrode input.

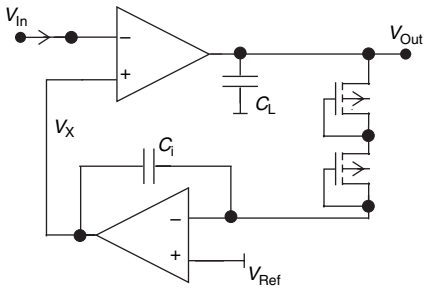


Figure 17 Illustrative schematic of a bioamplifier with active direct current (DC) suppression. (Source: Gosselin *et al*. 2006.)

In conclusion, a third way for neural recording is, based on the ISFET principle, by connecting the electrode to the gate of a metal oxide semiconductor field-effect transistor (MOSFET). Within an electrically grounded electrolyte, a neuron can be brought in direct contact with an extracellular electrode, covered by a dielectric layer; as a result, a cleft of the order of 50 nm between the cell membrane and the dielectric is obtained. Membrane currents that flow through the cleft lead to a potential drop due to the resistance of the cleft. This voltage signal is capacitively coupled to the electrode below. An illustrative model is presented, for example, in Eversmann (2003). Thus, the neuron's action potential induces changes in the sensing electrode potential, which again results in a modulation of the transistor's drain current. This method was used, for example, as early as 1986 (Najafi and Wise 1986) or as recently as 2003 (Eversmann 2003), to achieve a 128×128 electrode sensor with a readout rate of 2000 frames per second; note that this results in a data rate of approximately 32 million samples per second, which must be processed and subsequently transmitted.

1.16.4.1.3 Advanced circuit techniques

Beyond the basic ideas of nerve signal recording, various circuit techniques are quite commonly adopted in order to improve the signal-to-noise ratio and therewith increase the resolution. Moreover, the high-pass filter behavior should not be degraded, but even improved. Some of these techniques are the following:

- Differential recording of each electrode with respect to a dedicated, reference electrode, which improves the common-mode rejection. This technique, usually adapted in analog signal processing, also improves the signal-to-noise ratio, the dynamic range, suppression of even order harmonics, and more.
- CDS, which is usually adopted in subsequent analog signal processing in order to remove offsets that are caused by device mismatch (Aziz and Genov 2006).
- The use of voltage-dependent resistive devices in the feedback of the readout amplifier in **Figure 16**.

It can, for example, be realized by a pseudoresistor, i.e., two antiserial diodes (Olsson *et al.* 2003) or two serial MOS diodes (Harrison and Charles 2003), as used in **Figure 17** (Gosselin *et al.* 2006). This combination has a very large resistance (>1 GΩ) for low AC signals (<200 mV), but decreases its value for larger signals, thus lowering large signal distortion. Additionally, the devices used do not need a biasing circuit, as, for example, subthreshold transistors would need (Wise *et al.* 2004).

- The use of analog signal chopping, which is a commonly used technique for suppressing unwanted low-frequency signals, thus implementing high-CMRR and low-noise front ends (Enz and Temes 1996). This is achieved by mixing the signal to an intermediate frequency (IF), and after low-frequency noise and offset appearance, this deteriorated signal is mixed again, bringing the disturbance to the IF and the signal back to DC. Since chopping amplifiers are by principle DC-coupled devices, a modified concept was proposed for an AC-coupled instrumentation amplifier in Yazicioglu *et al.* (2006).

1.16.4.1.4 Parallel recording

Nowadays, development of neural signal recording is driven mainly by one demand: the possibility to massively record in parallel at a vast amount of different electrode sites (Wise *et al.* 2004). Many previously published integrated neural interfaces with large number of electrodes were realized with time-multiplexed signal acquisition, i.e., multiple sites are connected to a smaller number of signal-recording channels. Nonetheless, time multiplexing prevents the accurate capture of heterogeneity or communication across biological ensembles because the recording on different channels is performed consecutively rather than parallelly and simultaneously (Aziz and Genov 2006).

Thus, the development of very small, low-power readout amplifiers with subsequent parallel analog signal processing is a must, and many recent publications focus on this topic (Aziz and Genov 2006, Eversmann 2003, Wise *et al.* 2004, Yazicioglu *et al.* 2006). Furthermore, the integration and packaging of the system also becomes a major focus, since huge amounts of circuitry, wiring, and electrode tips must be assembled in a small, reliable housing. Outstanding advances have been achieved at the University of Michigan in this field over the past decades, who integrated active circuits with electrodes in a 3D system, nowadays also combining recording and stimulation electronics (Wise *et al.* 2004).

Finally, a common, very important topic is the data transmission: if, for example, a number of 100 electrodes is assumed, each of which is sampled at a data rate of only 5 kb s^{-1}, and if each of the sampled data points provide only 5 bits of resolution then a huge amount of data of 2.5 Mbits s^{-1} is available from this recording system. If such a system is wired and used for *in vitro* experiments, then the major task is to handle the data on chip (Eversmann 2003). A completely different challenge appears, if the recording system is implanted into the body: a wireless link with a capability of data rate of several megabits per second is a nontrivial task. This becomes especially true, because power consumption for an implantable device is a very critical issue. Therefore, either multiplexing is employed again or signal preprocessing is performed before the data is sent over the wireless link. This can be as easy as a simple spike detector (Harrison *et al.* 2006, Watkins *et al.* 2004) or a more complex digital signal processing (O'Driscoll *et al.* 2006).

1.16.4.2 Electronic Stimulators for Functional Electrical Simulation

As outlined in Section 1.16.4.1.1, the nerve cell potential can be depolarized from its resting potential at −70 mV toward ≈ −55 mV by an external stimulus; when this threshold is crossed the nerve cell fires a spike that runs along an axon toward the nervous system, e.g., the cell could be one out of millions in the human retina, stimulated by the incoming light whose action potential is transmitted via the optic nerve toward the brain. Nonetheless, the external stimulus can also be electrically generated by means of a neural stimulator implant.

1.16.4.2.1 Basics of FES

During the past decade, the development of implantable FES and FNS has made great progress in the treatment of neural and muscular disabilities. Among others, the cochlear implant, cardiac pacemaker, deep brain stimulators, implants for bladder control, or stimulators for hand or extremity control are some of the prominent examples. One of the most recent advances is the retinal implant for visual restoration (Chow and Chow 1997, Eckmiller *et al.* 2005, Liu 2004, Ortmanns *et al.* 2006b, Stieglitz 2006, Wyatt and Rizzio 1996, Zrenner 2002). Electrical stimulation of a neural response is achieved by transferring

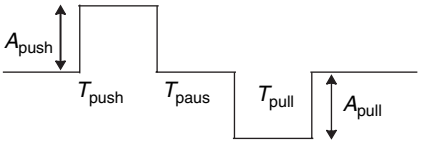


Figure 18 Biphasic, push–pull stimulation current.

the charge over a conducting electrode into the neural tissue. The amount of charge needed and the time for this transfer heavily depend on many factors: the stimulation circuitry, the stimulation waveform, the frequency, the excitation threshold of the cells (see Section 1.16.4.1.1) as well as on the stimulating electrode's location, geometry, surface coating, and configuration (Merrill *et al.* 2005). Commonly, the charge transfer is done using biphasic current pulses as indicated in **Figure 18**, with amplitudes ranging from as low as 10 μA for stimulation in the cortex to as high as several hundred microamperes for the cochlear or retinal implant (Humayun *et al.* 1999, Liu *et al.* 2000, Prutchi and Norris 2004, Weiland *et al.* 2003, Wise *et al.* 2004). Constant voltage stimulation has been used especially in commercially available, external stimulators, but the control on the delivered charge is low and therefore the safety is doubtful, unless external measures can be taken. Other approaches, for example, SC charge delivery, are still under research (Ghovanloo 2006), since it is unknown how the neural cells respond on spiking excitation.

If electrical current is forced through an electrolyte via electrodes, electrochemical effects must be taken into account. By bringing a conducting material in contact with a saline solution, a molecular layer of water is adsorbed at the interface, as indicated in **Figure 19**. As a consequence, when charging the electrode, the opposite charge in the solution is separated from the electrode by the so-called Helmholtz double layer acting as a capacitor (Dymond 1976). A feasible electrical model of the electrode–tissue interface will therefore incorporate the double layer capacitance and in parallel an impedance that models the electrode–electrolyte charge movement (Franks *et al.* 2005). This current results from both reversible and irreversible electrochemical faradaic reactions. In addition, the solution itself represents a resistance for the current, and thus a series resistor is taken into account.

Figure 20 shows this simplified electrode model for two stimulation electrodes and one counter electrode; C_E represents the double layer interface

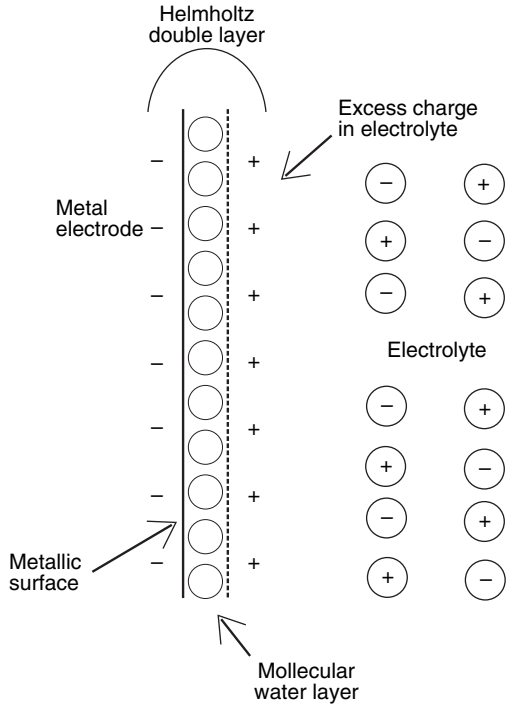


Figure 19 Simplified model for an electrode–tissue interface.

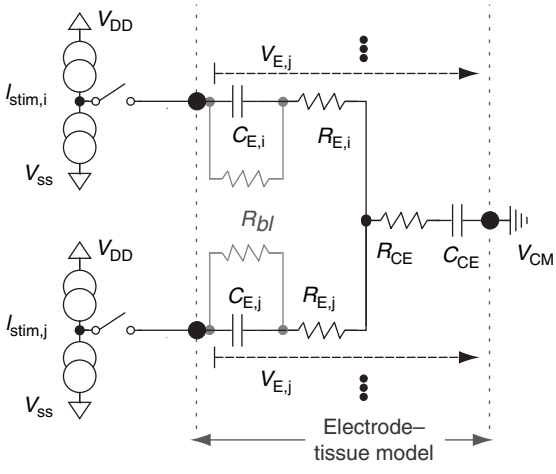


Figure 20 Simplified model for an electrode–tissue interface.

capacitor, R_E the solution–tissue impedance, while R_{Bl} represents the charge transfer resistance, which models reduction and oxidation currents flowing in equilibrium. In other models (Ortmanns *et al.* 2006b), diodes with series resistors R_D in parallel to R_E model strong faradaic currents that flow if the potential over the electrode–tissue interface permanently increases over a large oxidation–reduction potential, also

known as the water window. Even more complex models can be found in Franks *et al.* (2005), Najafi and Wise (1986), and Olsson *et al.* (2003); but commonly, for safe operation only R_E and C_E are of interest. In the exemplary case of a retinal implant, experimental results with the geometrically small electrodes yields $R_E \approx 10\,k\Omega$ and $C_E \approx 100\,nF$; the common ground electrode has $C_{CE} \approx 10\,\mu F$, whose influence can thus be neglected.

Generally, in multichannel microstimulators, the stimulator circuitry selects an active electrode and controls a stimulus current I through it. Especially, if many electrodes are spaced very closely to each other, a possible cross talk between neighboring sites can be significant (Lovell *et al.* 2005), which is not considered in the simple model in **Figure 20**. Constant current stimulation with biphasic current pluses is also schematically indicated in **Figure 20**; in bipolar stimulation, current is stimulated from one electrode to another, while in monopolar stimulation a physically large ground electrode serves as a common counter electrode.

1.16.4.2.2 Basic requirements and challenges

Besides the commonly accepted requirements of an implantable integrated circuit, i.e., low power consumption, small physical size, only few external components (Wise *et al.* 2004), modern stimulation systems, e.g., a visual prosthesis, mainly exceed these by one more critical feature: stimulation currents have to be independently delivered into a very large number of small electrodes, all of which need to fulfill the same safety issues as before (McDermott 1991, Scheiner and Mortimer 1990). This results in three major tasks that have to be fulfilled: high voltage compliance, high parallelism, and safety and test. These are discussed in more detail in the following sections.

1.16.4.2.2.(i) High voltage compliance Constant current stimulators can work only as long as the maximum possible output voltage swing ($\leq V_{DD}$ in **Figure 20**) is larger than the stimulation current times the effective electrode impedance. For example, for a biphasic stimulation current of ± 1 mA and a tissue impedance in the range of 10 kΩ (Humayun *et al.* 1999, Ortmanns *et al.* 2006b, Sivaprakasam *et al.* 2005), the stimulator voltage compliance must be larger than ± 10 V. With some voltage overhead for the current sources and the stored charge on the interface capacitor C_E, this yields in a supply voltage of 20–25 V. Several approaches have been presented to achieved this voltage compliance:

- Using old CMOS technologies with larger breakdown voltage (Sivaprakasam *et al.* 2005, Suaning and Lovell 2001).
- Limiting the maximum output current to a value feasible for the chosen technology.
- Using electrode reversal, i.e., a monopolar current source is adapted, but the polarity of the electrodes is turned around (Liu *et al.* 2000, McDermott 1991).
- Employing a high-voltage (HV) CMOS technology (Doguet *et al.* 2000, Ortmanns *et al.* 2006b).

Note that an old CMOS requires much higher die area, especially if a large digital overhead is needed for communication, signal processing, or control, since digital area is quadratically reduced with CMOS scaling. Second, a flexible stimulator should cover a large range of output currents, thus its limitation limits the functionality and the flexibility of the stimulator. Finally, electrode reversal requires voltage switches that are realized in CMOS by transmission gates; these need the highest switching voltage at their gates and even in HV technologies, such transistors are not available (Ortmanns *et al.* 2006a). Furthermore, electrode reversal allows only one electrode being stimulated at a time, if monopolar systems with one large counter electrode are used.

1.16.4.2.2.(ii) Parallelism For many modern FES systems, the primary requirements is a large number of stimulation electrodes (Liu 2004, Zrenner 2002), e.g., for a visual prosthesis in order to provide basic pattern recognition. In addition to the pure number of stimulation sites, the frequency of their independent stimulation is also important. In order to achieve flicker-free vision in a retinal prosthesis, stimulation frames should be applied at a rate of 50–60 Hz. Imagine as the little as 16 bits of addressing and programming data for each electrode, 60 Hz frame repetition means approximately $100\,kb\,s^{-1}$ data per 100 electrodes. This amount of data must be received with low power consumption and thereafter effectively distributed and stimulated on the ASIC. Recent approaches with large numbers of electrodes provided only a limited number of independent current sources, which were then multiplexed to a large number of output electrodes (Kim and Wise 1997, Liu *et al.* 2000, Sivaprakasam *et al.*

2005, Suaning and Lovell 2001). Nonetheless, this limits the flexibility and, most importantly, multiplexing again requires serial switches, which are not available in a HV stimulator (see Section 1.16.4.2.2.1).

1.16.4.2.2.(iii) Safety and test In order to be feasible for long-term implantation, besides the basic functionality of delivering stimulation currents into a large amount of electrodes, test and safety features must also be provided. Typically, the electrode impedance may vary significantly depending on the patient, long-term effects as well as attachment to the tissue (Weiland *et al.* 2003). Therefore, a possibility for an electrode impedance measurement is commonly provided (Ortmanns *et al.* 2006b, Wise *et al.* 2004).

The most critical safety risk is the prevention of DC stimulation currents into the electrodes in order to protect the tissue from destruction electrolysis with electrode dissolution and pH change (Rattay 1990, Scheiner and Mortimer 1990). Ideally, biphasic stimulation pulses as in **Figure 18** do not show net charge transfer; but through device mismatch, a finite DC current appears (Liu *et al.* 2000). Previous measurements for a retinal stimulator showed that DC currents I_{dc} must be <10 nA for the used electrodes (Ortmanns *et al.* 2006b). (This corresponds to a large charge transfer resistor R_{BI} in **Figure 20**.) Then, charge accumulation on the electrodes is prevented and strong faradaic currents do not flow. On the other hand, stimulation currents of more than 1 mA are commonly seen (Liu 2004, Ortmanns *et al.* 2006b); thus, biphasic current pulse matching should be in the range of:

$$\frac{\Delta I}{I} \leq \frac{10\,\text{nA}}{1\,\text{mA}} = 0.001\,\% \qquad [20]$$

which is not a realistically achievable value. The corresponding countermeasures are explained subsequently.

1.16.4.2.3 DC current prevention through charge balance concepts

Besides the biphasic current in **Figure 18**, which achieve first-order charge-balanced stimulation (Scheiner and Mortimer 1990), several methods exist to guarantee safe, DC-free current stimulation.

1.16.4.2.3.(i) Blocking Capacitors In most designs since 1970s, serial blocking capacitors between the stimulation sources and the tissue interface capacitors (C_E in **Figure 20**) have been the predominant technique (Huang *et al.* 1999, Sauning and Lovell 2001). By their physical property, they cannot conduct DC currents, which is the wanted characteristic, and furthermore this is the only safety feature in the case of semiconductor failure. But blocking capacitor have two major disadvantages: first, a serial capacitor integrates the total stimulation current, while it must not saturate the output current source by too high-voltage requirements; thus usually much larger values than the interface capacitor C_E are required, which cannot be realized on chip. Additionally, one capacitor is needed per electrode if parallel stimulation is required. For a large number of sites (e.g., the retinal prosthesis), the technique is thus not feasible (Liu 2004, Ortmanns *et al.* 2006b).

1.16.4.2.3.(ii) Passive and switched charge balancing Recent integrated FNS implementations used charge-balancing techniques, which made use of shorting all electrodes (Liu *et al.* 2000) after stimulation or of discharge resistors between the stimulation electrodes and/or the counter electrode (Sivaprakasam *et al.* 2005): therewith, accumulated charge and thus excess electrode voltage should be neutralized. But there arise two contradictory requirements: small discharge resistors are efficient for fast and reliable charge balancing at all electrodes, but they carry a substantial amount of stimulation current if they are in the range of the electrolyte resistance R_E as in **Figure 20**. On the other hand, large discharge resistors are efficient during stimulation, but the time constant may become too large for sufficient charge balancing, e.g., for the retinal implant in Ortmanns *et al.* (2006b), with $C_E \approx 100\,\text{nF}$ and $R_E \approx 10\,\text{k}\Omega$, a discharge resistor of $100\,\text{k}\Omega$ already carries 10% of the stimulation current, but the discharge time constant is 10 ms; if the charge imbalance is 1–10 μA over time, the voltage over the electrode increases to several hundred millivolts, which may finally yield strong faradaic currents.

Consequently, recent implementations used switched electrode shortening with low discharge resistors, which are connected only after a stimulation; this is efficient during stimulation and sufficient for reliable charge balancing. But, for HV applications, there arises again the question of a HV switch, and additionally, common to all of these techniques the charge balancing is not a closed-loop controlled process, but rather uncontrolled.

1.16.4.2.3.(iii) Active charge balancing In order to solve these challenges, active charge balancers were studied; until recently, mostly all active approaches relied on controlling a copied version of the stimulation current and performing electrode charge balancing until the copy was also balanced (Gnudason 1999). The disadvantage of such techniques is the need for an exact copy, and as indicated in eqn [20] the required mismatch is unrealistically high. Consequently, in Ortmanns *et al.* (2006b) a direct closed-loop control of charge-balanced stimulation has been presented. This system actively steers the excess electrode potential after a mismatched biphasic stimulation into a safe window. The idea is illustrated in **Figure 21** as an extension of the two stimulation electrodes in **Figure 20**: after stimulation at either electrode with a biphasic current I_{Stim}, due to mismatch an excess electrode voltage will be left on C_{E}. This voltage is not directly accessible, since without the additional sensing electrode, only the counter electrode or the stimulation electrodes can be electrically connected. As long as the current flows, mainly the large resistive voltage drop over the tissue and the electrolyte impedance R_{E} is seen, but not the voltage over the interface capacitance. Without current, $V_{\mathrm{CE}} \approx \Delta V_{\mathrm{E}}$ can be measured. (The voltage over the counter electrode interface capacitance is negligible because $C_{\mathrm{CE}} \gg C_{\mathrm{E}}$.) Note that the exact value is important only if the excess voltage exceeds a safe voltage window around the common-mode potential. This check is done subsequent to each stimulation by using a window comparator, for which a feasible implementation has been presented in Ortmanns *et al.* (2006b). In case $V_{\mathrm{E}} > V_{\mathrm{CM}} \pm \Delta V_{\mathrm{E,safe}}$, after the measurement the stimulation current sources are used to counteract

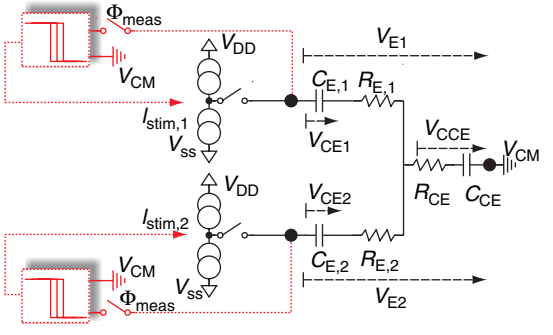


Figure 21 System of closed-loop, active charge balancer. (Source: Ortmanns M 2007 Charge balancing in functional electrical stimulators: a comparative study *Proc. IEEE Int. Symp. Circuits and Systems,* New Orleans, LA, USA, pp. 573–6).

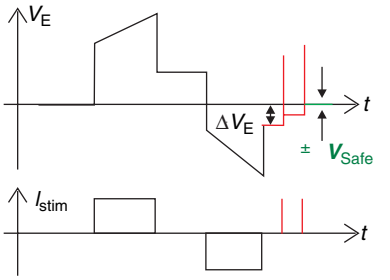


Figure 22 Excess electrode potential elimination.

the excess electrode voltage by applying a small charge packet of feasible polarity. This can be done with current spikes, as indicated in **Figure 22** (Ortmanns *et al.* 2006b), but other implementations are also feasible. This check-and-charge operation is repeated until the electrode potential lies within a safe window. Therewith, an exact control of the safe operating condition is given to the system and the operator.

1.16.4.2.4 A retina stimulator ASIC

As an example for electronic interfaces to neural tissue, a recently presented visual prosthesis will be shortly reviewed (Ortmanns *et al.* 2007). About 1.5 million people worldwide suffer from a chronic degeneration of the retina (retinitis pigmentosa or age-related macular degeneration), which disables their ability of vision. Therefore, basic restoration of vision through functional electrical nerve stimulation has become a focus for biomedical implants over the past few years (Ohta *et al.* 2005, Ortmanns *et al.* 2006a, Liu 2004, Zrenner 2002). Several techniques are used, while the retinal approach relies on electrically stimulated neurons in the retina that bypass the damaged photoreceptors and directly create visual excitation.

Most systems of such stimulators consist of external image reception by means of a camera and an external image preprocessing. A system, which has been implemented by IIP Technologies GmbH and sci-worx GmbH, has been published in Ortmanns *et al.* (2006a, b, 2007) and is illustrated in **Figure 23**. The data are sent to the implanted stimulator via an optical link at $1\,\mathrm{Mb\,s^{-1}}$ Manchester code. The energy is transferred with a telemetric RF link at 13.56 MHz. Internal rectification and buffering provide supply voltage for the stimulator. Uniquely, the system of Ortmanns *et al.* (2006a) uses two internal coils that provide three voltage levels: a high voltage is used for the stimulator output ($V_{\mathrm{Stim}} > 20\,\mathrm{V}$); a middle tap gives the counter electrode potential

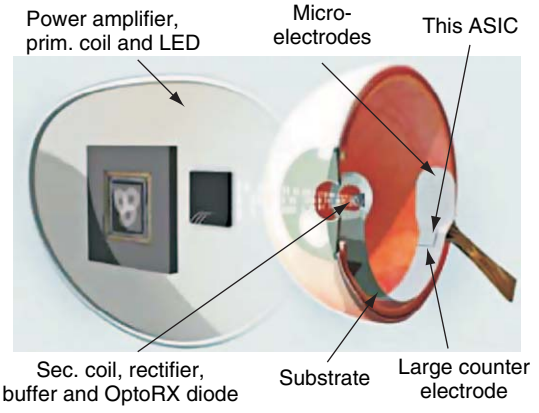


Figure 23 System overview of en epiretinal implant. (Courtesy: IEEE and IIP Technologies GmbH. Source: Ortmanns M, Unger N, Rocke A, Rackow S, Gehrke M, Tiedtke H J 2007 A 232 channel visual prosthesis ASIC with production compliant safety and testability. *Proceedings of the IEEE International Solid-State Circuits Conference*.)

$V_{cm} = V_{Stim}/2$, which is a large area connected to the intraocular tissue; and a second coil provides the unregulated low supply voltage (4–5 V) for the LV analog and digital part of the implant. In order to achieve high flexibility and efficient execution of the stimulation commands, the implant itself has to be separated into shared, global functions and an array of digitally programmable stimulation pad cells with dedicated functionality.

This concept of row/column stimulation array has been used as early as in 1973 (Donaldson 1973) for addressing that different electrodes has been used; but here, it is extended to a digitally programmable array of dedicated stimulation cells in Ortmanns *et al.* (2006b).

1.16.4.2.4.(i) Shared functions The architecture of the global functions is illustrated in **Figure 24(a)**. From the unregulated low supply voltage a constant $V_{DD} = 3.3$ V supply is provided by a series regulator. The high supply voltage is safely limited to approximately $V_{Stim} = 25$ V with a shunt regulator. A bandgap circuit provides a reference voltage and a current. An optical receiver (OptoRX) recovers the incoming optical data stream at 1 Mb s^{-1} Manchester coded data and clock recovery is directly performed on the sinusoidal RF input signal. A global stimulation control unit (GSU) serves as a communication interface. It decodes the 1 Mb s^{-1} incoming serial data stream and restores the information of the next stimulation sequence. Using load modulation on the secondary coil, data are back communicated sent outside with 20 kb s^{-1}. An attached 32-bit PROM block stores the implant ID and also contains tuning bits. The data transfer from the GSU to the local pad cells is done over a digital bus and a row column decoder. Finally, a dual-slope ADC is used for the measurement of electrode voltage.

1.16.4.2.4.(ii) Local stimulation functions The architecture of the local functions is shown in **Figure 24(b)** (Ortmanns *et al.* 2006b). Since not all pads are stimulated in parallel due to power issues,

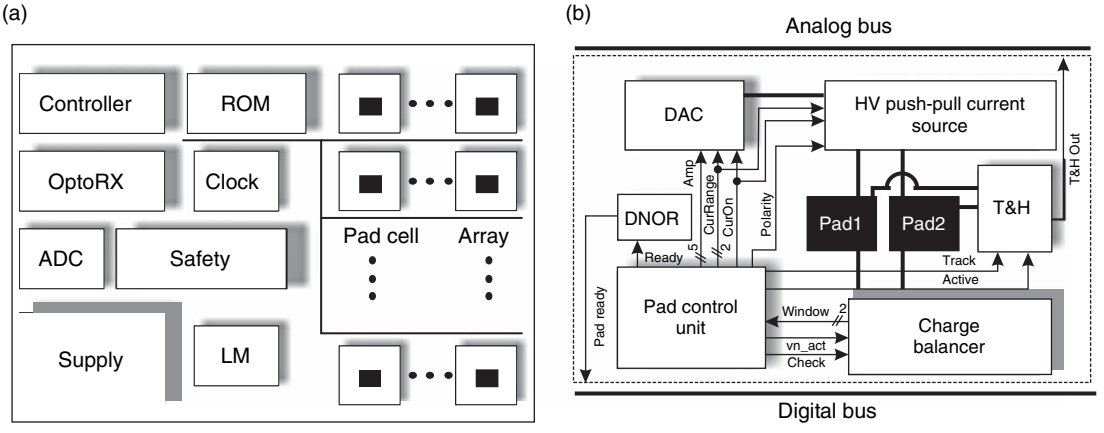


Figure 24 Illustration of retinal stimulator. (a) Global architecture with distributed pad cells and (b) pad cell architecture. (Courtesy: IEEE and IIP Technologies GmbH. Source: Ortmanns M, Unger N, Rocke A, Gehrke M, Tiedtke H J 2006 A tetina stimulator ASIC with 232 electrodes, custom ESD protection and active charge balancing. *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 652–5; Ortmanns M, Unger N, Rocke A, Gehrke M, Tiedtke H J 2006 A 0.1 mm^2 digitally programmable nerve stimulation pad cell with high-voltage capability for a retinal implant. *Proceedings of the IEEE International Solid-State Circuits Conference*.)

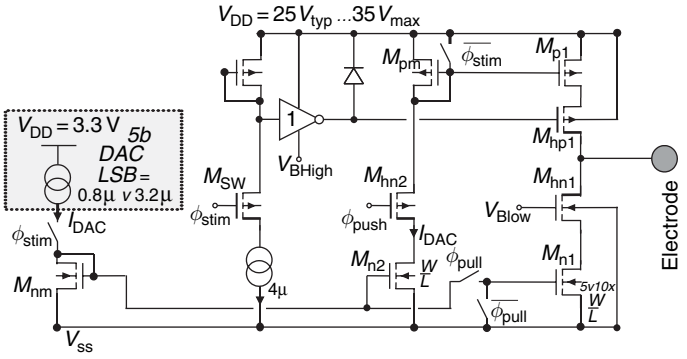


Figure 25 Output stimulation current source. (Courtesy: IIP Technologies GmbH. Source: Ortmanns M, Unger N, Rocke A, Gehrke M, Tiedtke H J 2006 A tetina stimulator ASIC with 232 electrodes, custom ESD protection and active charge balancing. *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 652–5; Ortmanns M, Unger N, Rocke A, Gehrke M, Tiedtke H J 2006 A 0.1 mm^2 digitally programmable nerve stimulation pad cell with high-voltage capability for a retinal implant. *Proceedings of the IEEE International Solid-State Circuits Conference*.)

each local stimulation pad cell serves as a pair of stimulation electrodes. This way, the local functions are used twice, die area is saved, and full flexibility is still achieved. The distributed pad cells are programmed via the digital bus. A digital pad control unit (PCU) accepts and stores data from the bus if addressed. Then, the local amplitude data are provided to a 5-bit current steering digital-to-analog converter (DAC), which features two bias ranges (1LSB = 0.8/3.2 µA; LSB = least significant bit). The DAC output is mirrored into a HV output current source, which connects the stimulation electrode (**Figure 25**). The output mirror gain is either 5 or 10. Thus, the output current can be adjusted between 4 and 992 µA with 5-bit resolution and dynamic range (DR) ≈ 50 dB. Finally, the active charge balancer circuit explained in **Figure 21** is implemented at each stimulation electrode (Ortmanns *et al.* 2006b).

Figure 26 shows the used circuit for the charge balancer explained in Section 1.16.4.2.3.3. Differential voltage measurement is performed without stimulation current, thus the residual charge on the electrode capacitance C_E is obtained. The voltage measurement at a high common-mode electrode potential ($V_{cm} = V_{Stim}/2 \approx 10$ V) is achieved with a HV differential pair with source degeneration: it translates the HV difference $V_E - V_{CM}$ into an equivalent current difference, mirrors it into the LV domain, where it is compared to a reference window. If the corresponding voltage exceeds a safe operation window of ±50 mV, a charge packet of correct polarity is applied. In **Figure 25** the HV push-pull current source is also shown. For good matching of the biphasic pulse and thus good first-order charge

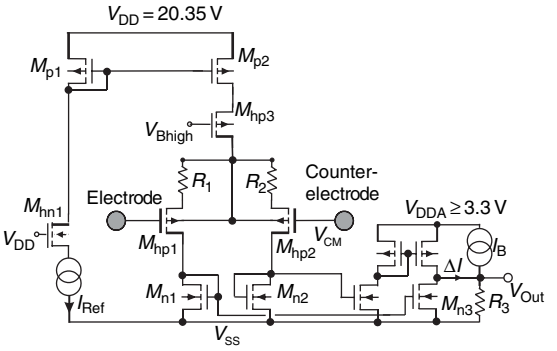


Figure 26 Exemplary implementation of the active change balancer. (Courtesy: IIP Technologies GmbH. Source: Ortmanns M, Unger N, Rocke A, Gehrke M, Tiedtke H J 2006 A tetina stimulator ASIC with 232 electrodes, custom ESD protection and active charge balancing. *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 652–5; Ortmanns M, Unger N, Rocke A, Gehrke M, Tiedtke H J 2006 A 0.1 mm^2 digitally programmable nerve stimulation pad cell with high-voltage capability for a retinal implant. *Proceedings of the IEEE International Solid-State Circuits Conference*.)

balancing, the current mirror matching is critical; therefore, LV transistors are used for the mirrors in the HV domain, while HV transistors shield the LV domain from overvoltage. Electrostatic discharge (ESD) protection is a must, even with the large number of 232 electrodes; standard ESD-protected HV pads are about the target size of the stimulation pad cell. Thus, in this design the parasitic drain-bulk diodes of the HV output transistors were used as ESD diodes to the supply rails (Ortmanns *et al.* 2006a, b).

1.16.4.2.4.(iii) Chip design This presented circuit concept for a visual prosthesis stimulator has

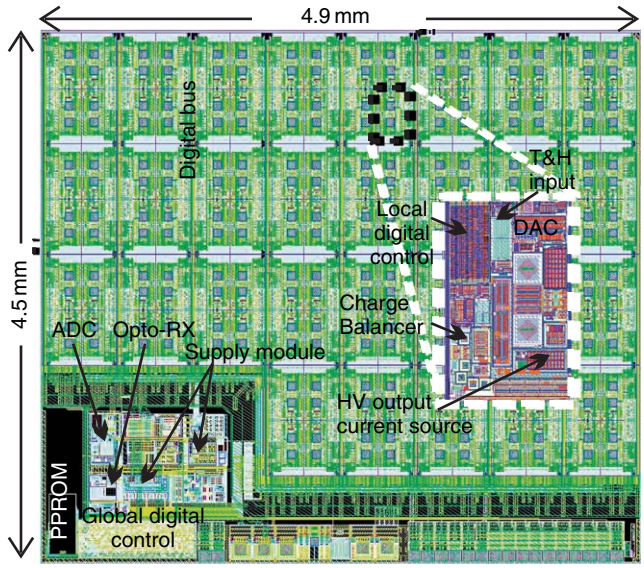


Figure 27 Layout of the 232 electrodes retina stimulator application-specific integrated circuit (ASIC). (Courtesy: IEEE and IIP Technologies GmbH. Source: Ortmanns M, Unger N, Rocke A, Gehrke M, Tiedtke H J 2006 A tetina stimulator ASIC with 232 electrodes, custom ESD protection and active charge balancing. *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 652–5; Ortmanns M, Unger N, Rocke A, Rackow S, Gehrke M, Tiedtke H J 2007 A 232 channel visual prosthesis ASIC with production compliant safety and testability. *Proceedings of the IEEE International Solid-State Circuits Conference*.)

been realized in a 0.35-μV HV CMOS technology, which provides up to 50 V voltage capability, while being feasible for a small digital circuit area. In total 116 stimulation pad cells were placed, each serving two electrodes; thus, this retinal stimulator shows the largest number of stimulation sites presented up to now (Ortmanns *et al.* 2007). Each pad cell consumes as little as 100,000 μm^{22} of die area, making the overall chip $\approx 20\,mm^2$ large. The layout of the retina stimulator ASIC is shown in **Figure 27**.

1.16.5 Conclusion

A number of different microelectronic techniques employed in the integration of microsystems have been presented in the previous sections. The use of microelectronics has brought major breakthroughs to a wide range of fields of application and the possibility of combining various individual microtechniques is seen as a major economic factor. This technology delivers microsensors with integrated electronics for local signal conditioning. The output signals are less susceptible to noise, thus enabling a safe data acquisition, transmission, and further processing. Not only is the size, weight, and power consumption of such systems reduced, but their functionality is also drastically enhanced. This miniaturization opens new application fields – some of which we are just beginning to envision. The most challenging and at the same time exciting part of this development is the fact that a number of engineers from different disciplines join together to make this possible.

References

Aboushady H, Louerat H-M 2002 Systematic approach for discrete-time to continuous-time transformation of sigma delta modulators. *IEEE Int. Symp. Circuits and Systems (ISCAS'2002),* Phoenix AZ, USA**4(4)**, 229–32

Aziz J N Y, Genov R 2006 Electro-chemical multi-channel integrated neural interface technologies. *Proc. IEEE Int. Symp. on Circuits and Systems*, Kos, Greece, May 2006, pp. 2201–4

Baltes H, Brand O, Fedder G K, Hierold C, Korvink J G, Tabata O 2004 *Enabling Technology for MEMS and Nanodevices.* WILEY-VCH Verlag, Weinheim, Germany, pp. 49–92

Bernstein J, Miller R, Kelley R, Ward P 1998a Low noise MEMS vibration sensor for geophysical applications. *Proceedings of the 1998 Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, pp. 55–8

Boser B E 1997 Electronics for micromachined inertial sensors. *Transducers '97*, Chicago, IL, USA, pp. 1169–72

Bryzek J 1993 Evolution of smart transducer design. *Sensor Expo West*, San Jose, CA, USA

Chow A Y, Chow V Y 1997 Subretinal electrical stimulation of the rabbit retina. *Neurosci. Lett.* **225**, 13–16

Clark W A, Howe R T Surface micromachined *Z*-axis vibratory rate gyroscope. *IEEE Solid-State Sensor and*

Actuator Workshop, Hilton Head Island, SC, USA, pp. 283–7

Djuric Z, Matovic J, Randjelovic D, Jakic Z, Dankovic T 1998 Noise in micromechnical structures caused by temperature fluctuations. *Proc. 5th NEXUSPAN Workshop Thermal Aspects in Microsystem Technology*, Budapest, Hungary

Doguet P, Mevel H, Verleysen M, Troostersz M, Trullemans C 2000 An integrated driver for the electrical stimulation of the optic nerve. *Proc. 5th Annu. Conf. Int. Funct. Elect. Stimul. Soc.*, Aalborg, Denmark, pp. 309–12

Donaldson P E K 1973 Experimental visual prosthesis. *Proc. IEE* **120**, 281–98

Dudaicevs H, Kandler M, Manoli Y, Mokwa W, Spiegel E 1994 Surface micromachined pressure sensors with integrated CMOS read-out electronics. *Sens. Actuators* **A43**, 157–63

Dymond A M 1976 Characteristics of the metal–tissue interface of stimulation electrodes *IEEE Trans. Biomed. Eng.* **23**, 274–80

Eckmiller R, Neumann D, Baruth O 2005 Tunable retina encoders for retina implants: Why and how. *J. Neural Eng.* **2**, S91–104

Enz C C, Temes G C 1996 Circuit techniques for reducing the effects of op-amp imperfections. *Proc. IEEE* **84(11)**, 1548–614

Eversmann B, Jenkner M, Paulus C, Hofmann F, Brederlow R, Holzapfl B, Fromherz P, Brenner M, Schreiter M, Gabl R, Plehnert K, Steinhauser M, Eckstein G, Schmitt-Landsiedel D, Thewes R 2003 A 128 × 128 CMOS biosensor array for extracellular recording of neural activity. *IEEE J. Solid-State Circuits* **38(12)**, 2306–17

Ferris C D 1978 *Introduction to Bioinstrumentation*. Humana Press, Clifton, NJ

Franks W, Schenker I, Schmutz P, Hierlemann A 2005 Impedance characterization and modeling of electrodes for biomedical applications. *IEEE Trans. Biomed. Eng,* **52(7)**, 1295–302

Gabrielson T B Mechanical-thermal noise in micromachined acoustic and vibration sensors. *IEEE Trans. Electron. Devices* **40**, 903–9

Garcia-Valenzuela A, Tabib-Azar M 1995 Comparative study of piezoelectric, piezoresistive, electrostatic, magnetic, and optical sensors. *Integr. Opt. Microstruct. II* **SPIE-2291**, 125–42

Geerts Y, Steyaert M, Sansen W 2002 *Design of Multi-Bit Delta–Sigma a/d Converters*. Kluwer Academic Publishers, Boston, MA

Ghovanloo M 2006 Switched-capacitor based implantable low-power wireless microstimulating systems. *Proc. IEEE Int. Symp. Circuits and Systems*, Kos, Greece, pp. 2197–200

Gnudason G 1999 An implantable mixed analog/digital neural stimulator circuit. *Proc. IEEE Int. Symp. Circuits and Systems,* Orlando, FL, USA Vol. 5, pp. 375–8

Gosselin B, Ayoub A E, Sawan M 2006 A low-power bioamplifier with a new active dc rejection scheme. *Proc. IEEE Int. Symp. Circuits and System*, Kos, Greece, pp. 2185–8

Gottfried-Gottfried R, Zimmer G, Mokwa W 1991 CMOS compatible magnetic field sensors fabricated in standard and in silicon-on-insulator technologies. *Sens. Actuators* **A25**, 753–7

Harrison R R, Charles C 2003 A low-power, low-noise CMOS amplifier for neural recording applications. *IEEE J. Solid-State Circuits* **38**, 958–65

Harrison R, Watkins P, Kier R, Lovejoy R, Black D, Normann R, Solzbacher F 2006 A low-power integrated circuit for a wireless 100-electrode neural recording system. *Proceedings of the IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA February 2006

Hosaka H, Itao K, Kuruda S 1995 Damping characteristics of beam-shaped micro-oscillators. *Sens. Actuator A* **49**, 87–95

Hosticka B J, Brodersen R W, Gray P R 1977 MOS sampled data recursive filters using switched capacitor integrators. *IEEE J. Solid-State Circuits* **12(6)**, 600–800

Huang C Q, Shepherd R K, Carter P M, Seligman P M, Tabor B 1999 Electrical stimulation of the auditory nerve. *IEEE Trans. Biomed. Eng.* **46(4)**, 461–9

Humayun M S, de Juan E, Weiland J D, Dagnelie G, Katona S, Greenberg R, Suzuki S 1999 Pattern electrical stimulation of the human retina. *Vis. Res.* **39**, 2569–76

Jarman D 1995 A brief introduction to sigma delta conversion. *Intersil*, Milpitas, CA

Johns D A, Martin K 1997 *Analog Integrated Circuit Design*. John Wiley & Sons, Inc., New York

Kakerow R G, Manoli Y, Mokwa W, Rospert M, Cammann K, Krause J, Meyer H 1993 A monolithic sensor array for measuring chemical and biochemical parameters. *Transducers, Digest of Technical Papers*, Yokohama, Japan pp. 1074–7

Kakerow R G, Kappert H, Spiegel E, Manoli Y 1995 Low-power single-chip CMOS Potentiostat. *Int. Conf. Solid-State Sensors and Actuators, Eurosensors*, Stockholn, Sweden pp. 142–5

Kersjes R, Liebscher F, Spiegel E, Manoli Y, Mokwa W 1996 An invasive catheter flow of sensor with on-chip CMOS read-out electronics for the on-line determination of blood flow. *Sens. Actuators* **A54**, 5637

van Kessel P F 2001 Electronics for DLPTM technology based projection systems. *Symp. VLSI Circuits, Digest of Technical Papers*, Kyoto, Japan, pp. 91–4

Kim C, Wise K D 1997 Low-voltage electronics for the stimulation of biological neural networks using fully complementary BICMOS circuits. *IEEE J. Solid-State Circuits* **33(10)**, 1483–90

Kordas N, Eichholz J, Langerbein A, Manoli Y, Mokwa W 1990 A digital-output monolithic temperature sensor for invasive applications. *Int. Conf. Micro System Technologies*, Berlin, Germany

Kuchta J 2004 Neuroprosthetic hearing with auditory brainstem implants. *Biomed. Tech. (Berl).* **49(4)**, 83–7

Lemkin M, Boser B E 1999 A three-axis micromachined accelerometer with a CMOS position–sense interface and digital offset-trim electronics. *IEEE J. Solid-State Circuits* **34(4)**, 456–68

Lerch R G, Spiegel E, Kakerow R, Hakenes R, Kappert H, Kohlhaas H, Kordas N, Buchmann M, Franke T, Manoli Y, Müller J 1995 A programmable mixed-signal ASIC for data-acquisition systems in medical implants. *Proceedings of the IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA pp. 160–1

Link T, Simin I, Trachtler M, Gaisser A, Braxmaier M, Manoli Y, Sandmaier H 2005 A new self-test and self-calibration concept for micro-machined gyroscopes. *Transducers, Digest of Technical Papers*, Coex, Seoul, Korea pp. 401–4

Liu W 2004 Retinal implants. *ISSCC Dig. of Tech. Papers*, San Francisco, CA, USA p. 12.1

Liu W, Vichienchom K, Clements M, DeMarco S C, Hughes C, McGucken E, Humayun M S, de Juan E, Weiland J D, Greenberg R 2000 A neuro-stimulus chip with telemetry unit for retinal prosthetic device. *IEEE J. Solid-State Circuits* **35(10)**, 1487–97

Lovell N H, Dokos S, Preston P, Lehmann T, Dommel N, Lin A, Wong Y T, Opie N, Hallum L E, Chen S, Suaning G J 2005 A retinal neuroprosthesis design base on simulataneous current injection. *Proc. IEEE Int. Conf. Microtechnolgies in Medical Biology*, Kahuku, Oahu, USA pp. 98–101

Manoli Y, Mokwa W 1999 Silicon microsystems – Merging sensors, circuits and systems. *Int. Conf. Computer Design*, Austin, TX, USA

Martins R, Selberherr S, Vaz F A 1998 A CMOS IC for portable EEG acquisition systems. *IEEE Trans. Instrum. Meas.* **47**, 1191–6

McDermott H 1991 A custom-designed receiver-stimulator chip for an advanced multiple-channel hearing prosthesis. *IEEE J. Solid-State Circuits* **26(8)**, 1161–4

Merrill D R, Bikson M, Jefferys J G R 2005 Electrical stimulation of excitable tissue: Design of efficacious and safe protocols. *J. Neurosci. Meth.* **141**, 171–98

Murari K, Stanacevic M, Cauwenberghs G, Thakor N 2005 Integrated potentiostat for neurotransmitter sensing. *IEEE Eng. Med. Biol. Mag.* **24(6)**, 23–9

Najafi K, Wise K D 1986 An implantable multielectrode array with on-chip signal processing. *IEEE J. Solid-State Circuits* **SC-21**, 1035–44

Nathan A, Baltes H P 1989 A high-resolution integrated Si magnetic sensor. *Transducers, Digest of Technical Papers*, Montreux, Switzerland, pp. 230–1

O'Driscoll S B, Meng T, Shenoy K, Kemere C 2006 Neurons to silicon: Implantable prosthetic processor. *Proceedings of the IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA

Ohta J, Kagawa K, Tokuda T, Nunoshita M 2005 Retinal prosthesis device based on pule-frequency-modulation vision chip. *Proc. IEEE Int. Symp. Circuits and Systems*, Kobe, Japan, pp. 2923–6

Olsson R H, Gulari M N, Wise K D 2002 A fully-integrated bandpass amplifier for extracellular neural recording. *Proceedings of the IEEE Engineering in Medicine and Biology Special Topic Conference*, pp. 237–40

Olsson R H, Gulari M N, Wise K D 2003 A fully-integrated bandpass amplifier for extracellular neural recording. *Proceedings of the IEEE Engineering in Medicine and Biology Symposium*, Copri Island, Italy, pp. 165–8

Ortmanns M, Rocke A, Unger N 2004 Method for control of electrical charge on stimulation electrodes, DE102004059973A1

Ortmanns M, Unger N, Rocke A, Gehrke M, Tiedtke H J 2006a A retina stimulator ASIC with 232 electrodes, custom ESD protection and active charge balancing. *Proc. IEEE Int. Symp. Circuits and Systems*, Kos, Greece, pp. 652–5

Ortmanns M, Unger N, Rocke A, Gehrke M, Tiedtke H J 2006b A 0.1 mm^2 digitally programmable nerve stimulation pad cell with high-voltage capability for a retinal implant. *Proceedings of the IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA

Ortmanns M, Unger N, Rocke A, Rackow S, Gehrke M, Tiedtke H J 2007 A 232 channel visual prosthesis ASIC with production compliant safety and testability. *Proceedings of the IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA

Payne R S, Sherman S, Lewis S, Howe R T 1995 Surface micromachining: From vision to reality to vision. *Proceedings of the IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA pp. 164–5

Petersen K E 1982 Silicon as a mechanical material. *Proc. IEEE* **70(5)**, 420–57

Petkov V P, Boser B E 2005 A fourth-order $\Sigma\Delta$ interface for micromachined inertial sensors. *IEEE J. Solid-State Circuits* **40(8)**, 1602–9

Prutchi D, Norris M 2004 *Design and Development of Medical Electronic Instrumentation*. Wiley, Hoboken, NJ, ISBN 0471676233

Rattay F 1990 *Electrical Nerve Stimulation*. Springer-Verlag, Wien

Razavi B 2001 *Design of Analog CMOS Integrated Circuits*. The McGraw-Hill Companies, New York

Scheiner A, Mortimer J T 1990 Imbalance biphasic electrical stimulation: Muscle tissue damage. *Ann. Biomed. Eng.* **18**, 407–25

Schmidt M, Spiegel E, Dudaicevs H, Manoli Y, Mokwa W 1995 Read-out electronics for wide temperature range with integrated surface micromachined capacitive pressure sensor. *Proceedings of the European Solid-State Circuits Conference*, Lille, France pp. 230–3

Sivaprakasam M, Liu W, Humayun M S, Weiland J D 2005 A variable range bi-phasic current stimulus driver circuitry for an implantable retinal prosthetic device. *IEEE J. Solid-State Circuits* **40(3)**, 763–71

Smith S, Hurwitz J, Torrie M, Baxter D, Holmes A, Panaghiston M, Henderson R, Murray A, Anderson S, Denyer P 1998 A single-chip 306 $\times$ 244-pixel CMOS NTSC video camera. *Proceedings of the IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA pp. 170–1

Spreemann D, Manoli Y, Folkmer B, Mintenbeck D 2006 Non-resonant vibration conversion. *J. Micromech. Microeng.* **16**, S169–73

Steyaert M S J, Sansen W M C, Zhongyuan C 1987 A micropower low-noise monolithic instrumentation amplifier for medical purposes. *IEEE J. Solid-State Circuits* **SC-22**, 1163–8

Stieglitz T 2006 Biomedical microimplants for sensory and motor neuroprostheses. *Proc. IEEE Int. Symp. Circuits and Systems*, Kos, Greece, pp. 2189–92

Suaning G J, Lovell N H 2001b CMOS neurostimulation ASIC with 100 channels, scalable output, and bi-directional radio-frequency telemetry. *IEEE Trans. Biomed. Eng.* **48(2)**, 248–60

Watkins P T, Santhanam G, Shenoy K V, Harrison R R 2004 Validation of adaptive threshold spike detector for neural recording. *Proceedings of the IEEE Engineering in Medicine and Biology Symposium*, San Francisco, CA, USA

Weiland J D, Yanai D, Mahadevappa M, Williamson R, Mech B V, Fujii G Y, Little J, Greenberg R J, deJuan E, Humayun M S 2003 Electrical stimulation of retina in blind humans. *Proceedings of the. IEEE International Conference on Engineering in Medicine and Biology*, Cancun, Mexico, pp. 2021–4

Wise K D, Najafi N 1991 The coming opportunities in microsensor systems. *Transducers, Digest of Technical Papers*, San Francisco, CA, USA pp. 2–7

Wise K D, Anderson D J, Hetke J F, Kipke D R, Najafi K 2004 Wireless implantable microsystems: High-density electronic interfaces to the nervous system. *Proc. IEEE* **92(1)**, 76–97

Wyatt J, Rizzio J 1996 Ocular implants for the blind. *IEEE Spectrum,* 47–53

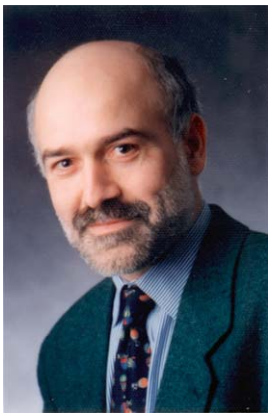
Yazicioglu R, Merken P, Puers R, Van Hoof C 2006 A 60 μW 60 nV/$\sqrt{}$Hz readout front-end for portable biopotential acquisition systems. *Proceedings of the IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA February 2006

Zavracky P M, Hertley F, Sherman N, Hansen T, Warner K 1993 A new force balanced accelerometer using tunneling tip position sensing. *Dig. Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, pp. 50–1

van der Ziel A 1978 Flicker noise in semiconductors: Not a true bulk effect. *Appl. Phys. Lett.* **33(10)**, 883–4

Zrenner E 2002 Will retinal implants restore vision? *Science* **295(5557)**, 1022–5

Biographies



Yiannos Manoli was born in Famagusta, Cyprus, in 1954. As a Fulbright scholar, he received the B.A. degree (*summa cum laude*) in Physics and Mathematics from Lawrence University in Appleton, Wisconsin, in 1978 and the M.S. degree in Electrical Engineering and Computer Science from the University of California, Berkeley, in 1980. He obtained the Dr.-Ing. Degree in Electrical Engineering from the Gerhard Mercator University in Duisburg, Germany, in 1987.

From 1980 to 1984 he worked at the University of Dortmund, Germany, in the field of digital and analog MOS integrated circuit design with an emphasis on A/D and D/A converters. For his patented work on a self-calibration method he received the best paper award of the European Solid-State Circuits Conference (ESSCIRC 1988).

In 1985 he joined the newly founded Fraunhofer Institute of Microelectronic Circuits and Systems in Duisburg where he established a design group for microsystem and microcontroller integrated circuits. His work there concentrated on mixed-signal CMOS circuits especially for monolithic integrated sensors such as surface micromachined pressure sensors, flow sensors, biosensors etc. Another highlight of his work was the design of application specific microcontrollers as well as novel architectures and development tools for such microcontrollers. This work has resulted in several patents and over 60 technical publications.

In September 1996 he joined the Department of Electrical Engineering at the University of Saarland in Saarbruecken, Germany where he held the Chair of Microelectronics. His research interests were focused on the design of low-voltage/low-power mixed-signal CMOS circuits and VLSI embedded system design and design methodology. As department head (1998–2000) he initiated a number of reforms that boosted the research activities and enhanced the teaching curriculum of the department. In 2000 he had the opportunity to spend half a year on a research project with Motorola in Phoenix, Arizona.

In July 2001 he was appointed Chair of Microelectronics at the Department of Microsystems Engineering (IMTEK) of the Albert-Ludwig-University in Freiburg, Germany. His current research interests are the design of low-voltage/low-power mixed-signal CMOS circuits, wireless sensor network electronics, sensor read-out circuits as well as A/D- and D/A-converters. In 2006 he spent his sabbatical semester on a research project with Intel in Santa Clara, California.

Since May 2005 he and two other IMTEK colleagues additionally serve as directors at the Institut für Mikro- und Informationstechnologie of the Hahn-Schickard Gesellschaft (HSG-IMIT) in Villingen-Schwenningen.

Professor Manoli has served on the committees of a number of conferences such as ISSCC, ESSCIRC, ICCD and DATE and was Program Chair (2001) and General Chair (2002) of the IEEE International Conference on Computer Design (ICCD). He is a member of Mortar Board, Phi Beta Kappa, IEEE, VDE and VDI.

Maurits Ortmanns received the Dipl.-Ing. in electrical engineering from the Saarland University, Saarbruecken, Germany, in 1999 and the Dr.-Ing. degree from the Albert-Ludwigs-University, Freiburg, Germany, in 2004, both with highest honors.

In 1997 and 1998, he was with the Research Center Karlsruhe, Germany, and with EXAR, Inc., Fremont, CA, as an engineering intern working in the field of Microsensors. In 1999, he joined the Institute of Microelectronics of the Saarland University, where he was working as a research assistant. In 2001, he moved to the Institute of Microsystem Technology, Albert-Ludwigs-University, where he was working toward the Ph.D. degree in the field of continuous-time sigma-delta modulator design until March 2004.

From 2004–2005, he has been with sci-worx GmbH, Hannover, Germany, where he was working in the field of mixed-signal and analog circuits for biomedical implants, low voltage, low-power applications, and high-speed circuits. In the beginning 2006,

Mr. Ortmanns was with the HSG-IMIT, Villingen-Schwenningen and additionally a guest lecturer at the Institute for Microsystems Engineering at the University of Freiburg, Germany.

Since May 2006, Mr. Ortmanns is an assistant professor for Integrated Interface Circuits at the same university. His main research interests include mixed-signal circuit design, smart microsensors, and biomedical applications.

Mr. Ortmanns received the VDI award for the Best Masters Degree and the VDE award for the Best Masters Thesis in electrical engineering from the University of Saarland in 1999. He is a member of the German National Academic Foundation, the IEEE and VDE.

Alexander Buhmann received the Dipl.-Ing. degree from the University of Freiburg, Germany, in 2003, with a thesis on the optimization of MEMS based micro resonators. He is currently pursuing the Ph.D. degree at the Chair of Microelectronics, Institute for Microsystems Engineering (IMTEK), University of Freiburg. His current research interests include sensors and systems, control theory, and algorithm development.

1.17 Testing, Calibration and Compensation

Jürgen Wilde,[1] Leonhard Reindl,[1] and Craig Stewart[2], [1]Department of Microsystems Engineering – IMTEK, University of Freiburg, Freiburg, Germany, [2]Candim Consulting Ltd, Bloxham, UK

Nomenclature

e Charge on an electron (1.60×10^{-19} C)
k Boltzmann constant (1.38×10^{-23} J K^{-1})

Glossary

BIST Built-in Self-test
BMM Bulk Micromachining
BPF Band Pass Filter
CMOS Complementary Metal Oxide Semiconductor
DC Directional Coupler
DUT Device Under Test
EMC Electromagnetic Compatibility
EIA Electronics Industry Association
ESD Electrostatic Discharge
FIB Focused Ion Beam
GS Ground–Signal
GSG Ground–Signal–Ground
HBM Human Body Model
IC Integrated Circuits
IF Intermediate Frequency
IR Infrared
ISS Impedance Standard Substrate
JEDEC (Formerly Joint Electron Devices Engineering Council) now Solid State Technology Association
L Load
LCL Lower Control Limit
LL Lower Limit
LRM Line-Reflect-Match
LRM+ Extended Line-Reflect-Match
LRRM Line-Reflect-Reflect-Match
LWL Lower Warning Limit
MEMS Microelectromechanical Systems
MM Machine Model
MTTF Mean Time to Failure
NDT Nondestructive Testing
NIST US National Institute of Standards and Technology
O Open
OEM Original Equipment Manufacturer
RF Radio Frequency
S Short cut
SEM Scanning Electron Microscope
SMM Surface Micromachining
SOLT Short-Open-Load-Thru
SOLR Short-Open-Load-Reciprocal
SPC Statistical Process Control
TCE Thermal Coefficient of Expansion
TCR Thermal Coefficient of Resistance
TRL Thru-Reflect-Line
TSOL Thru-Short-Open-Load
UCL Upper Control Limits
UL Upper Limits
UWL Upper Warning Limits
VLSI Very Large-scale Integration
VSWR Voltage Standing Wave Ratio
VNA Vector Network Analyzer
WL Wafer-level

1.17.1 Introduction

In microelectromechanical systems (MEMS) technology, testing is a verification technique for the performance, quality, or reliability of the devices. Furthermore, extensive testing is used for the control of manufacturing steps as well as for field monitoring and self-test of MEMS. A distinction must be made between the phases of design and development, and production. This chapter describes and discusses the various tests that are applied for MEMS and their specific features. Furthermore, it demonstrates how these methods can be utilized as part of a comprehensive quality strategy.

1.17.2 Qualification Procedures and Test Strategies

This section discusses systematically how tests are utilized in the life cycle of development and production of microsystems. Hence, it is helpful to consider the various domains of the process of product generation. The principal ones are the chain of supply and the timescale.

Typically, microsystems are developed and produced by a supplier of components for an original equipment manufacturer (OEM) like a car manufacturer who defines requirements and specifications. The three basic elements of a specification are functionality, reliability, and costs. A typical development model is the V-model (**Figure 1**).

As a first step, the concept of architecture for the microsystem and the specification must be validated. The validated specification must be fault-free, i.e., correct, complete, free of contradictions, and applicable. As a result of the formal validation, the specification will be a binding requirement for all further activities.

Based on the specification of the system, subspecifications are derived for the individual blocks of the system. In this case, such blocks might be MEMS hardware, software, and assembly and packaging technology. A specification for the system comprises an environment temperature of 150°C. As the MEMS device exhibits a heater, the hardware is specified for operation at 175°C. Also such a requirement must be validated formally. In a similar manner, requirements for the utilized materials are generated and formally accepted. All these activities make up the top-down branch of the V-model.

On the bottom-up path, the materials are first checked for their suitability, which requires materials testing. In the next step, they are used to build up hardware, which is subsequently tested to verify whether it meets the specification. Finally, the microsystem is integrated from its building blocks, and its functions and its reliability are verified. A large portion of this verification is done experimentally, i.e., by testing. Only when the system meets all the specifications, can it be qualified formally (**Figure 1**).

The V-model is an idealized structure. Also it does not make clear whether the number of subsystem elements increases like for a pyramid. All activities on one layer must be coordinated and synchronized. For that purpose, it is helpful to partition microsystems into three columns, as it has been implemented successfully for the design of mechatronic systems. In this case, separately designing software, MEMS hardware, and assembly and packaging technology appears to be an efficient partitioning. Also, it reflects the individual worlds of engineers working in different organizations and using different methods and tools (**Figure 2**).

By partitioning, parallelization is also achieved. In each branch, a standard sequence of analysis, design, prototype generation, and test is applied. After the individual tests and verifications, the system is integrated and tested as a whole. Design-for-testability is an important issue as testing costs make up a large portion of the development costs.

The considered approach is quite efficient when the microsystem development is performed in a steady-state process where subsequently similar systems are generated. Under the present conditions

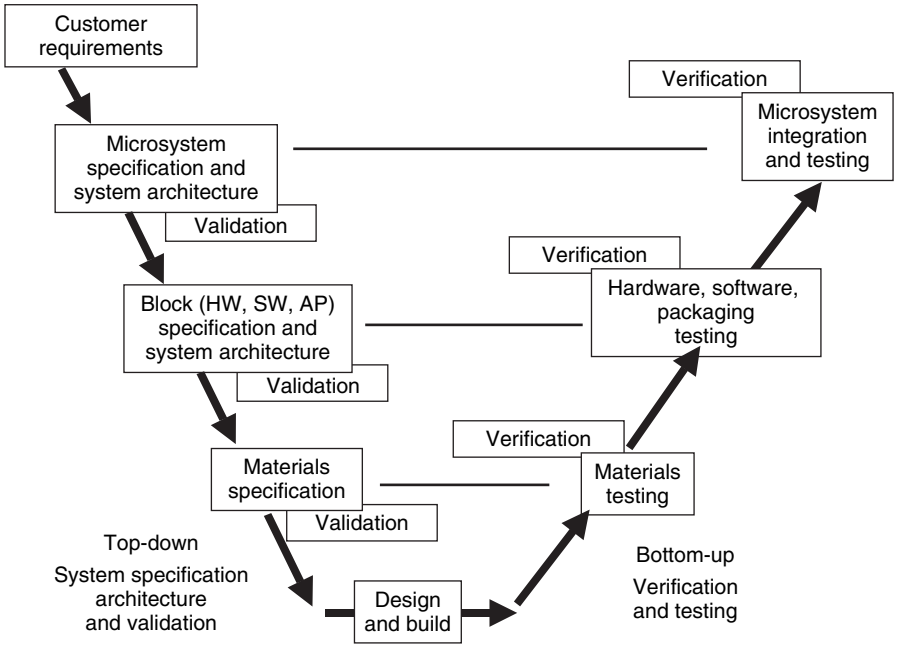


Figure 1 The V-model of top-down specification and bottom-up verification of microsystems.

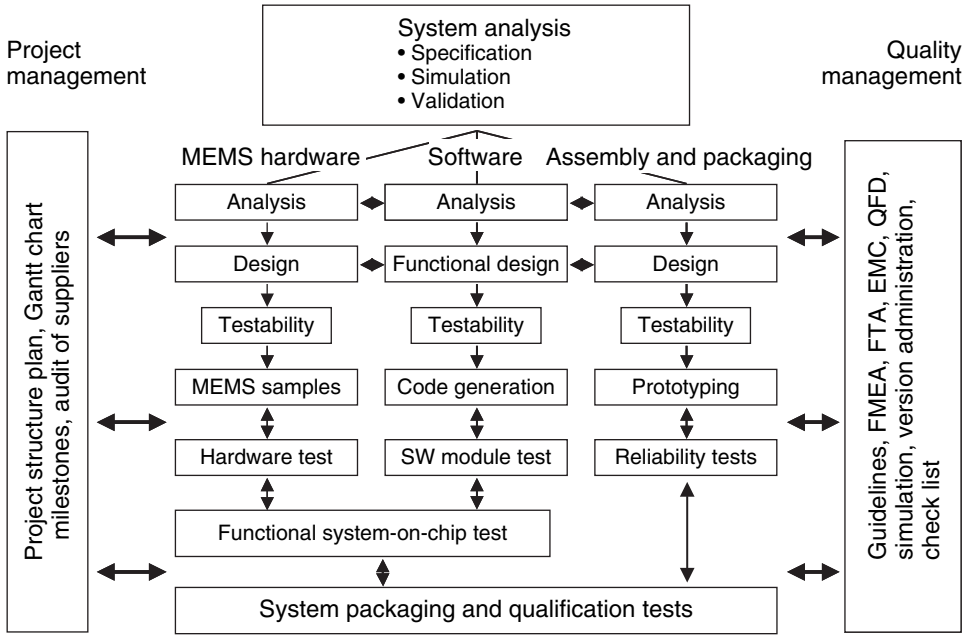


Figure 2 Partitioned system development of micromechatronic systems. EMC, electromagnetic compatibility; FMEA, failure mode and effects analysis; FTA, fault tree analysis; QFD, quality function deployment.

of accelerating development processes, the introduction of new technologies must also become more efficient. Also it will significantly affect the testing strategy.

Sometimes a microsystem producer plans to introduce a new technology. Examples include a transition from bulk micromachining (BMM) to surface micromachining (SMM) or a change from hermetic packages to wafer-bonded chips in plastic housings. In order to generate series products and to keep these qualified, several stages of qualification are necessary (**Figure 3**).

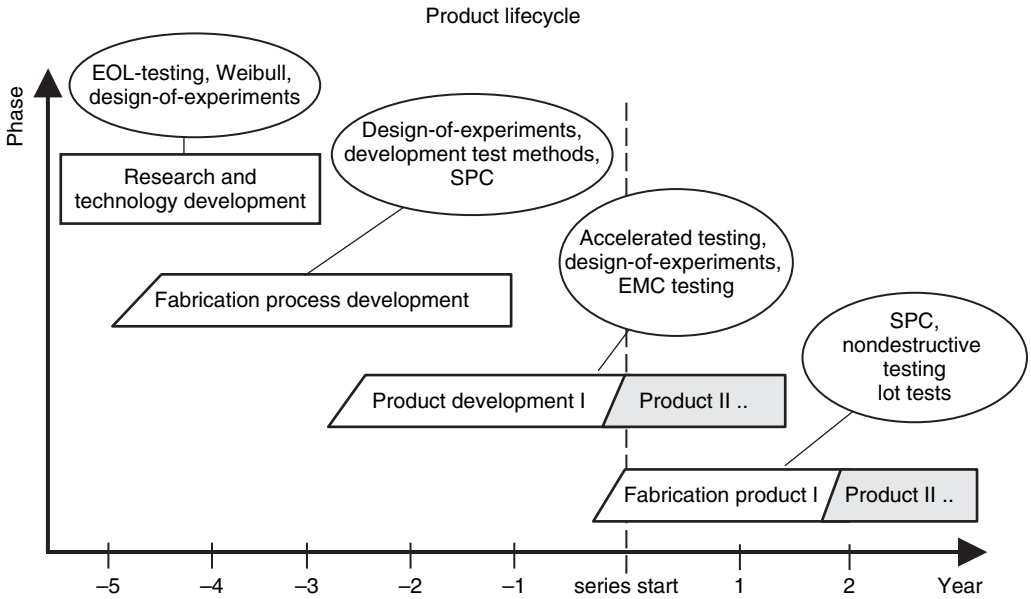


Figure 3 Reliability testing during different phases of the product generation cycle of microsystems. EOL, end-of-life; SPC, statistical process control.

Table 1 Characteristics of reliability tests in the technology-, process-, and product-development phases of MEMS hardware

Phase of life cycle	*Motivation of testing*	*Tests and test samples*	*Test scheme*
Technology development	Analyze and optimize design parameters, which affect reliability	Technology samples, dummies Accelerated tests Systematic failure analysis Simulation	Systematic design variations
Process development	Analyze process parameters and disturbance variables, which affect reliability	Robust fabrication processes Accelerated tests Optional failure analysis	Systematic process variations
Product development	Verify that new product meets all specifications	Test for conformity with requirements	Stable processes Some variations
Product series production	Monitor minimum quality and reliability levels	Statistical process control Accelerated tests Functional tests	No variations Constant fabrication
	Avoid faulty products	Nondestructive testing Failure analysis of defects	

(1) In the first phase, it is necessary to perform research projects in order to characterize a new technology.
(2) In the second phase, fabrication processes must be qualified for the new technology.
(3) In the third phase, products will be developed and qualified using the new technology.
(4) In the last phase, the products are fabricated and a requalification of the series products is verified on a lot-by-lot basis.

In **Table 1** the different requirements and purposes of testing in the subsequent phases of the product life cycle are summarized. The methodology as well specific reliability tests will be described in Section 1.17.7.

1.17.3 Process Monitoring

1.17.3.1 Statistical Process Control

The MEMS facility must produce devices that fulfill their specifications with respect to performance, quality, and reliability. Typically, at the end of a MEMS fabrication process the devices will be subjected to comprehensive final tests to verify their accordance with the specifications. Unfortunately, this concept usually does not allow for fast process monitoring and control. Therefore, the method of statistical process control (SPC) has been established in order to check whether a process is controlled and capable. Also it is an efficient technique to analyze the transient quality in comparison with the long-term behavior.

The total process chain must be partitioned into individual steps, which can be analyzed efficiently. In the SPC, charts are used to visualize the quality level of each step. For control charts, it is necessary to measure certain properties that are relevant for the quality and for subsequent fabrication steps.

(1) In the first step, all properties to be monitored must be determined, such as layer thickness, adhesive strength, and etch rate.
(2) These values are recorded over a certain period of time in lots of size n. From these mean values $\widehat{\mu}$ and variances $\widehat{\sigma}^2$ are computed for the total process. A test on the normal distribution should be performed.
(3) For the control charts upper and lower limits (ULs and LLs) are computed. These are the warning limits and the control limits.
(4) Within short periods of time lots of a limited size are taken repetitively. For each lot, mean values and variances are computed.

There are two principal types of control charts. The first one is the chart for mean values, which is applicable when always the same target value, e.g., a certain film thickness, is expected. The second is the control chart for the variability (scatter), which is sensible when the target property is variable.

1.17.3.1.1 Shewhart control charts for mean values

In order to determine the control limits, the mean value $\widehat{\mu}$ and the standard deviation $\widehat{\sigma}$ of the total process are computed as follows.

	In each lot	*Total process*
Mean value	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\hat{\mu} = \bar{\bar{x}} = \frac{1}{k}\sum_{i=1}^{k} \bar{x}_i$
Standard deviation	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\bar{x}-x_i)^2}$	$\hat{\sigma} = \sqrt{\frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\hat{\mu})^2}$
Upper limit UL	$\mathrm{LL} = \hat{\mu} - \left(\frac{u_{1-\alpha/2}}{\sqrt{n}}\hat{\sigma}\right)$	$\mathrm{UL} = \hat{\mu} + \left(\frac{u_{1-\alpha/2}}{\sqrt{n}}\hat{\sigma}\right)$
Lower limit LL		

x_i is the property of the product which is regarded.

$u_{1-\alpha/2}$ is taken for a certain confidence level $1-\alpha$ based on the normal distribution. α is the significance level. For $1-\alpha = 0.99$, $u_{1-\alpha/2} = 2.576$.

$1-\alpha$	0.90	0.95	0.99	0.999
$1-\alpha/2$	0.95	0.975	0.995	0.9995
$u_{1-\alpha/2}$	1.645	1.960	2.576	3.291

Upper and lower control limits (UCL and LCL) are taken for the confidence level of $1-\alpha = 0.99$ and warning limits UWL and LWL for the confidence level of $1-\alpha = 0.95$. This leads to four characteristic limits shown in the chart as follows:

Control limit	*Warning limit*
$1-\alpha = 0.99$	$1-\alpha = 0.95$
$\mathrm{LCL} = \hat{\mu} - 2.576\,\frac{\hat{\sigma}}{\sqrt{n}}$	$\mathrm{LWL} = \hat{\mu} - 1.960\,\frac{\hat{\sigma}}{\sqrt{n}}$
$\mathrm{UCL} = \hat{\mu} + 2.576\,\frac{\hat{\sigma}}{\sqrt{n}}$	$\mathrm{UWL} = \hat{\mu} + 1.960\,\frac{\hat{\sigma}}{\sqrt{n}}$

In **Figure 4**, a typical control chart is shown for the oxidation process of silicon wafers at a temperature of 950°C. The limits have been computed according to the equations above.

1.17.3.1.2 Control charts for variability

A typical test distribution is the χ^2-distribution, which is the basis for control charts for the variability of the product properties:

Control limit	*Warning limit*
$1-\alpha = 0.99$	$1-\alpha = 0.95$
$\mathrm{LCL} = \sqrt{\frac{\chi^2_{n-1;\ \alpha/2}}{n-1}}\,\hat{\sigma}^2$	$\mathrm{LWL} = \sqrt{\frac{\chi^2_{n-1;\ \alpha/2}}{n-1}}\,\hat{\sigma}^2$
$\mathrm{UCL} = \sqrt{\frac{\chi^2_{n-1;\ 1-\alpha/2}}{n-1}}\,\hat{\sigma}^2$	$\mathrm{UWL} = \sqrt{\frac{\chi^2_{n-1;\ 1-\alpha/2}}{n-1}}\,\hat{\sigma}^2$
$M = \hat{\sigma}\cdot a_n$ for s or $M = \bar{R}\cdot d_n$ for $\bar{R}$	

Typically, the scatter of each lot is recorded in the control chart. It is computed as the standard deviation of each sample. In some cases it is easier to monitor just the span R, i.e., the difference between the maximum and minimum values of each lot. Correction factors for the middle of the chart are taken from the following table:

Sample size n	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Factor a_n	0.798	0.886	0.921	0.940	0.952	0.959	0.965	0.969	0.973
Factor d_n	1.128	1.693	2.059	2.326	2.534	2.704	2.847	2.970	3.078

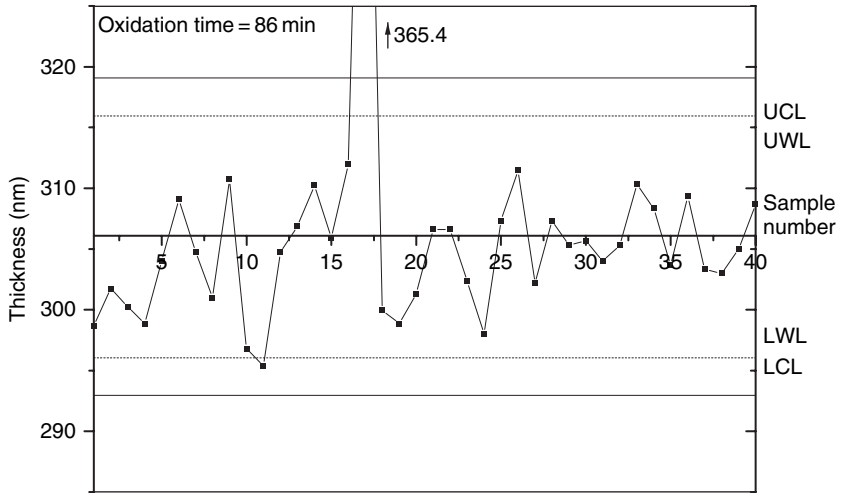


Figure 4 Statistical process control (SPC): Shewhart mean values control chart of wet oxidation of silicon at 950°C. Upper and lower control limits (UCLs and LCLs) and upper and lower warning limits (UWLs and LWLs).

1.17.3.2 Test Structures for Process Monitoring

In MEMS technology many individual process steps must be monitored by specific tests, which are listed in Table 2.

There are mainly four basic groups of test, which can be used on these structures:

Visual inspection	Optical microscope, scanning electron microscope (SEM), focused ion beam (FIB)
Electrical measurement	Resistance, dielectric properties
Optical measurements	Elipsometry, refractive index, infrared (IR) transmission
Mechanical tests	Adhesion, microhardness, internal stress

Many of these tests can be performed on wafers without specific test structure, but in many cases well-defined test features must be generated.

Only a part of the test methods is described by standards. A number of national and international standards define the test structures. Many of these have been established in very large-scale integration (VLSI) silicon processing. A well-known source is the US National Institute of Standards and Technology (NIST), which has defined test structures for contact resistance, feature sizes, or electromigration.

When a manufacturing process of a device is monitored, metrology of the device-specific structures on product wafers will yield the most realistic results. The main purpose of special test structures is standardization. This will make it possible to measure the properties of materials or process data under comparable conditions. In semiconductor technology, the use of test structures is established for decades and many test structures have been standardized by the NIST. A typical example is the NIST electromigration structure that requires a well-defined line width and contact geometry in order to generate comparable reliability data for an individual metallization. Another important property is the contact resistance. It depends on contact geometry, size, doping concentration, intermediate layers, and heat treatment. All parameters of the test structures must be variable in order to allow an adaptation to all products. There are several general requirements to test structures for MEMS. These are as follows:

- Versatile application over a wide range of technologies, materials, and sizes.
- Generation of data, which can be compared to existing and future structures.
- Ease of implementation on test wafers as well as on production wafers.
- Representative of target product and technology.

In the following the required test structures for diffused or implanted resistors in silicon will be discussed as illustrated in Figure 5. Such elements are frequently used as stress-sensitive transducers in pressure sensors or load cells.

The schematic structure exhibits a conductor realized by p-diffusion in an n-substrate. A layout as well as a cross section and a parallel section are shown. The resistance is affected by the following factors:

- Sheet resistance of the doped regions, which depends on the diffusion profile as well as on the specific resistance (doping concentration).

Table 2 Tests and test structures for the monitoring of the basic MEMS process steps

Process step	*Property*	*Test structure needed*
Oxidation, nitridation	Thickness and deposition rate, stress, dielectric breakdown strength, pinhole density per area, refractive index	No
Photoresist	Thickness, pattern definition and size, aspect ratio	Yes
Metallization and conductors	Thickness, pattern definition and size, adhesion, stress, resistance, contact resistance, electromigration resistance, aspect ratio, edge coating, bondability	Yes
Implantation or diffusion (resistors)	Doping profile, concentration, implantation depth, implantation width, resistance, contact resistance	Yes
Dielectric layers and insulators	Thickness, pattern definition and size, aspect ratio, adhesion, stress, resistance, planarization, insulation resistance, dielectric breakdown strength, edge coating, refractive index	Yes
Surface micromachined structures	Pattern definition and size, aspect ratio, stress, mechanical quality factor, resonance frequency	Yes
Deep etching	Deepness and etch rate, aspect ratio, wall angle, anisotropy, roughness	Yes
Wafer bonding	Adhesive strength, hermeticity, defects per bonded area	No

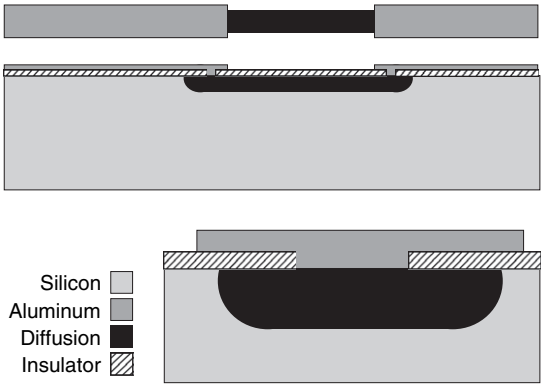


Figure 5 Cross section of an implanted resistor in silicon.

- Cutoff voltage and the reverse current of the insulating pn-junction.
- Contact resistance from silicon to the metal conductor.
- Sheet resistance of the metal line.
- Pattern quality, e.g., edge displacement and edge roughness.
- Contact resistance between test probes and bonding pads.

These individual properties cannot be determined on complete resistors. Also the cause of defectiveness can only be cleared up systematically on test structures, which are produced simultaneously with the resistor. In Figure 6 the following layers of the test structure are used:

01	diff1	Red	Doped region
10	pass1	Blue	Contacts between metal line and doped region
11	pass2	Turquoise	Aperture of the passivation on the bond pads
20	metall1	Green	Aluminum conductor

Using the upper left structure in Figure 6, the sheet resistivity of the aluminum line is measured. It is a four terminal structure, which is not affected by parasitic resistances of the wires or probes. Similarly, the resistivity of the doped region is measured with the second pattern in the top line.

The contact resistance between silicon and metal is measured using the third pattern in line 1. The contact is divided into small areas in order to avoid size effects as a result of current crowding. Also many contacts can be connected in series in order to measure the effective contact resistance and to evaluate the reliability at increased test currents with the first

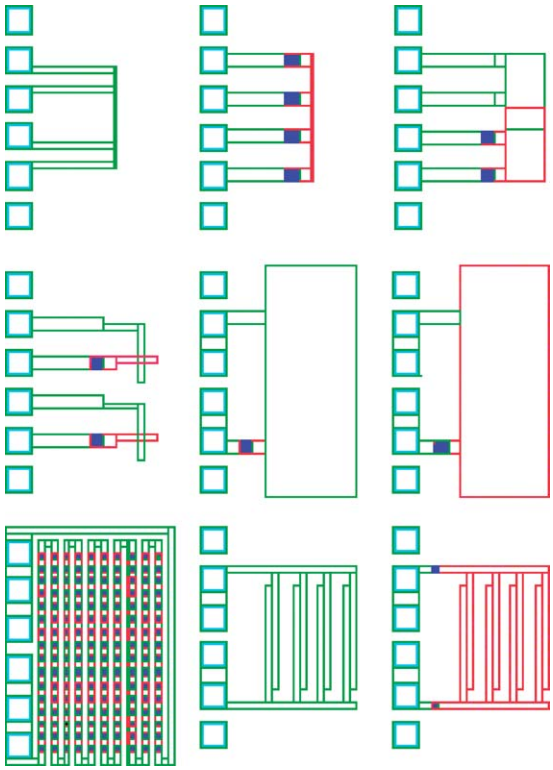


Figure 6 Test structures for monitoring structural and material parameters of implanted resistors in silicon.

structure in line 3. The structures in line 2 were developed to measure the breakdown strength and the residual current of insulating pn-transitions. In line 3 the dielectric breakdown strength of the passivation on top of the aluminum conductors is tested with the test pattern 2, while the last structure is used to check for defects between the two doped conductors.

Typically, test structures are scaled in size over a wide range because they must represent the actual product features. Also size-dependent effects such as edge displacement or underdiffusion must be taken into account. Unfortunately, in MEMS industry, many of the test structures are kept confidential as they can reveal the quality level of the fabrication facility. Also there is no comprehensive standard for BMM or SMM structures, which is equivalent to the microelectronic standards defined by the NIST.

1.17.3.3 Test Structures for Fault Detection

On microelectronic systems two basic types of faults can be distinguished. One is catastrophic faults, which lead to complete nonfunction. The second is

parametric faults, which can be regarded as parasitic elements, which can change voltages, currents, or the transient behavior. For mechanical elements of MEMS, complete failure is often a result of fractured structures or of nonreleased structures (Table 3). When functional parameters are affected, in most cases the mass, the stiffness, or the damping factor are out of range. In these cases the static as well as the dynamic behavior will be affected significantly. Therefore, it is possible to conclude from functional misbehavior back to physical defects of the micromechanical elements. Mir *et al.* (2000) proposed failure modeling in order to analyze fault effects on the functionality or on parameters, which can be measured. To that purpose, a physics-of-failure model and a subsequent simulation using the finite element method will be powerful tools to evaluate the effectiveness of possible test concepts. An example is the influence of vacuum quality on the damping properties of a resonant structure and therefore on the natural frequencies and the quality factor of the vibration (Table 4).

Table 4 lists the possible parametric faults in the mechanical and the electrical domain of micromachined thermal MEMS. Geometry and structural features as well as material properties are mainly affected by the fabrication process. As a result, the coefficients that determine the properties of the electrical or mechanical element will be affected. Examples are stiffness, mass, capacity, or resistance. On the behavioral level, properties such as resonance frequency, offset, and sensitivity can be monitored in tests.

Beroulle *et al.* (2001) worked out a scheme to conclude from relatively simple tests on system level to physical defects of the electrical or mechanical structures (Table 5). Unfortunately, such a concept has some limitations:

- The basic failures will be detected only with indirect methods.
- Measurement of resonance frequency or sensitivity requires physical stimulation.
- For complementary metal oxide semiconductor (CMOS) integrated MEMS the electrical properties on the physical level will only be accessible when adequate test points are integrated into the design. This requires a consequent design-for-testing concept.

Table 3 Catastrophic faults for CMOS-compatible suspended thermal MEMS due to manufacturing defects

Electronic structure faults (gauge)	*Mechanical microstructure faults*
Short faults	Break
Polysilicon, one level	Outside gauge
Polysilicon, two levels	Around gauge
Between poly-silicon and metal	Including gauge
Metal, one level	
Metal, two levels	
Open faults	Locking faults
Polysilicon level	Nonreleased microstructure
Metal level	Stiction
Polymetal contact	Cavity protuberances
	Miscellaneous
	Asymmetrical microstructure

Source: Mir *et al.* (2000)

1.17.4 Functional Testing

1.17.4.1 Wafer-Level Testing

A silicon-integrated microsensor exhibits electrical input connections for parameterization, programming, and power supply. The typical output is made up by the sensor signals and in most cases also by some status information. Microsensors need a physical stimulus to generate the measurand. In practice, sensors will need more than one stimulus as a result of cross sensitivities. Typically, the dominant disturbance variable is temperature but other effects such as pressure, moisture, or mechanical stress must also be taken into account. When testing at the package level, the device is already interconnected to the signals and to the physical world, as shown in Figure 7. Therefore, in the case of complete devices, correct physical stimulation and measurement are relatively easy to accomplish.

Such a package-level test concept alone has significant disadvantages: One is the delay between fault generation and fault detection due to the time of the manufacturing cycle. The second is the higher costs of complete devices. According to an analysis by MEMUNITY and SUSS, a typical costs distribution for a MEMS device is as follows:

- Cost of die: €3.00
- Cost of wafer-level (WL) test: €0.90
- Assembly and packaging: €7.00

Hence, 70% of all manufacturing costs are because of the back-end process and only 30% because of front-end fabrication. Therefore, if a silicon die is defective, the economic loss per piece is €3. When the bad MEMS chip is further packaged, the total loss will be €10 per defective unit.

Table 4 Classification of parametric faults on micromachined thermal MEMS

Mechanical domain	*Electrical domain*
Structural properties	Structural properties of conductors
Edge quality, width, length, thickness	Width
Cativity protuberances	Length
Friction and stiction	Thickness
Missing structures	Doping profile
Partial suspension or release of structures	
Vacuum quality or pressure	
Thermomechanical material properties:	Thermoelectrical material properties:
Mass density	Resistivity of polysilicon and metal
Young's modulus	Contact resistivity
Thermal coefficient of expansion (TCE) of metal	TCR of polysilicon and metal
Heat conductivity and capacitance	Piezoresistive coefficient of Si
Beam emissivity	Seebeck coefficient poly and metal
Convection coefficient	Permittivity and loss factor dielectrics
Friction coefficient	
Variation of mechanical coefficients	Parasitic electrical components:
Stiffness, compliance	Parasitic resistances
Mass	Parasitic capacities
Damping factor/viscosity	Parasitic inductivities
System behavior:	System behavior:
Natural frequency	Offset shift
Quality factor	Sensitivity change
Nonlinearity	Impedance-level change

Source: Mir *et al.* (2000)

Table 5 Electrical, mechanical, and environmental faults of a CMOS-integrated magnetic sensor

Electronic element fault	*Electrical test parameter*
Gauge reference resistor out of range	Offset
Gauge piezoresistive coefficients out of range	Sensitivity of offset
Force line resistance out of range	Sensitivity
	Offset
Gauge reference resistor short or open	Offset ($=V_{DD}/2$)
Bridge shorts	Sensitivity (null)
Force line short or open	Sensitivity (null)
Mechanical element fault	*Electrical test parameter*
Stiffness out of range	Resonance frequency
Mass out of range	Sensitivity or resonance frequency
Damping coefficient out of range	Sensitivity or quality factor
Cavity protuberances	Nonlinear behavior
Structure break or nonreleased	Sensitivity

Source: Beroulle *et al*. (2001).

Another aspect is the time, which it takes to feed back the failure information into the wafer production process. Failure analysis of a packaged device is tedious. The silicon chip must be unpacked without producing defect artifacts. Also, assembly and packaging is not performed in a clean room where the wafer was processed but in a factory at another site or even in another company. It will then take weeks to months to feed back the error information into the wafer fabrication. Therefore, the data will not be up-to-date and cannot influence the quality level of the wafers immediately.

Early-stage testing of microsystems on unpackaged devices is an important factor in a modern test strategy. It has been demonstrated for non-MEMS integrated circuits (ICs) that the WL testing of devices is a more

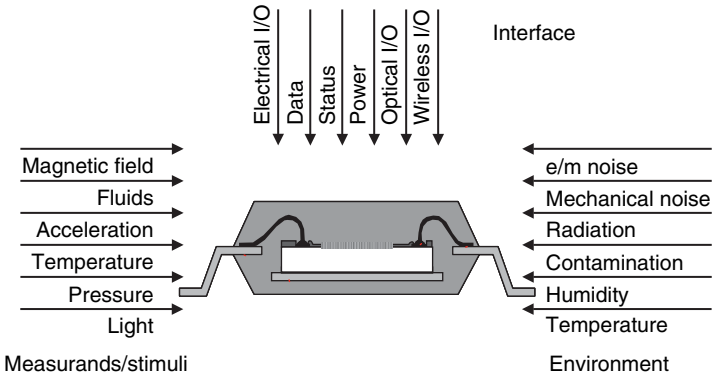


Figure 7 Signals and stimuli during the package-level testing of microsystems.

efficient strategy than the testing of single bare dies. Functional testing at the die level for known good dies requires expensive adapters and would be much more costly than the 10% costs portion of WL tests. A comparison of WL testing and package-level testing of MEMS is shown in **Table 6**. It should be noted that modern test strategies take advantage of both test concepts in order to optimize throughput, accuracy, dependability, and information gain altogether. This is the reason why testing and calibration costs are relatively high compared with the manufacturing costs.

In the following sections, the available technical solutions of testing MEMS on WL stage will be evaluated.

1.17.4.1.1 Wafer probe stations

Because the housing process is one of the most expensive steps during the fabrication of a device, most microsystems are partially tested on WL before the back-end process. A system for wafer testing will consist of four main components. They are as follows:

- The probe station.
- Single probes or a probe card for electrical interconnection to the device.
- A hot chuck that keeps the wafer accurately at a constant temperature level and
- The electronic measurement equipment.

The wafer is loaded manually or by an automatic transport system on a chuck of the wafer probe station and is fixed by a vacuum. There are fully automatic systems on the market, which are mainly used for mass production monitoring; semiautomatic ones for development purposes, and manual ones for research or rare usage (**Figure 8**).

Table 6 Comparison of wafer-level and package-level MEMS tests

Test level	*Wafer level*	*Package level*
Physical stimulation	Not possible for all measured variables	Easy, as all interfaces are present (electrical, mechanical, fluidic, optical)
Test equipment	With physical stimulation expensive Sequential testing	Test adapter costs can be high Expensive multichannel ATE
Measurable variables	I/O variables Electrical parameters on intermediate system levels (requires suitable test points)	Normally only I/O variables after internal signal processing Test modes optional if implemented
Tested elements	Specific tests of silicon element without interference of package	Test of complete microsystem, including packaging effects
Purpose	Measures basic functional properties MEMS process metrology and control Eliminates defective dies with reduced costs	System test Measure customer-relevant features Functionality and reliability Burn-in
Test time	Fast tests Allows fast control loops to the microfabrication processes	Slow tests Aimed at stability and long-term behavior

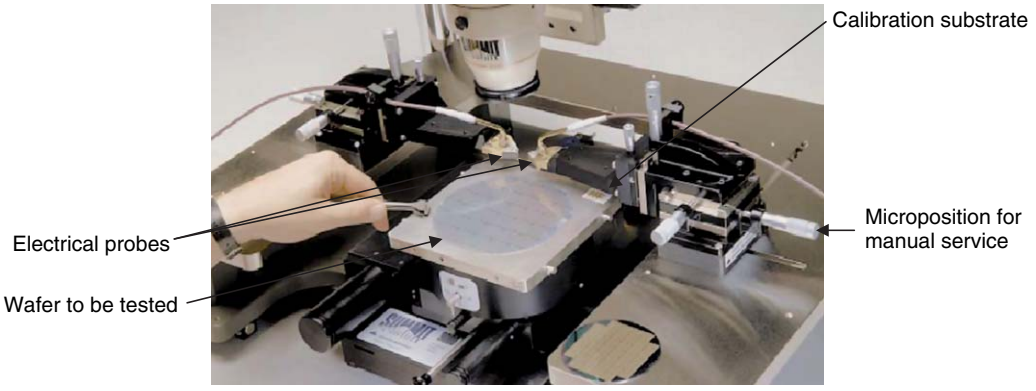


Figure 8 Close-up of a manual wafer probe station on the square chuck with independent vacuum control, which holds a calibration substrate (Impedance Standard Substrate (ISS), back to the right), and a test wafer (to the front) simultaneously. (Photo courtesy: Cascade Microtech.)

1.17.4.1.2 Temperature control

Usually electrical parameters such as impedances or transfer functions are measured using a wafer probe station. Some probes also allow applying a DC voltage between signal and ground. As many electrical properties of silicon structures are highly temperature sensitive, control of the temperature during WL testing is crucial for the accuracy of sensors. A special version of chucks, called temperature chucks, allow for a control of the temperature of the wafer and thus of the device under test (DUT) (**Figure 9**). This is accomplished by electrical heating of the backside of the chuck. For cooling, special air or liquid cooling channels or Peltier elements are integrated into the chuck. The typical test temperature range is between −65°C and 200°C, and special equipment ranges up to 300°C. To achieve a uniform temperature over the wafer, a low thermal resistance of both the chuck and the wafer backside is necessary, and also a high throughput of the cooling liquid is needed.

With a thermal chuck, reasonable care must be exercised to match the thermal expansion coefficients of the wafer and the chuck. Furthermore, the thermal expansion in the z-direction must be controlled very precisely. Otherwise, it would open the electrical contacts of the tips or, worse, could destroy the tips. Usually, special tips are applied to allow some thermal expansion.

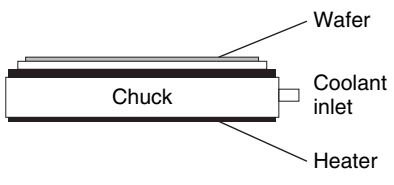


Figure 9 Schematic of the setup of a thermochuck.

1.17.4.1.3 Needle probes and microwave probes

To contact the DUT electrically to the measurement system, wafer probes with special needle-like contact elements are used (Wallitzer *et al.* 2001) (**Figure 10**).

A well-designed probing system also includes the die pad pattern and high-quality transmission line interconnections in order to minimize all errors as much as possible (**Figure 11**; see ground–signal–ground (GSG) versus ground–signal (GS) pads in Section 1.17.5.2.5.(v)).

The probes will indent into the metallic layer of the contact pads. The maximum allowed bending of the probes usually is very small. Thus, special care has to be taken to provide an extremely high planarity of the wafer adjustment to guarantee a good electrical contact with the probes, and to avoid breaking of the tiny needles. The contact of the probes to the pads usually is performed with the help of advanced micromanipulators and monitored by a microscope to check the correct alignment between probes and contact pads and to control the correct distance (**Figure 12**).

1.17.4.1.4 Device bias considerations

Power may be supplied through the analyzer's (e.g., VNA, vector network analyzer) bias ports, which combine the DC and radio frequency (RF) signals together at the measurement ports using bias tees. If the performance of the analyzer's internal bias tee is inadequate, or if the analyzer does not provide one,

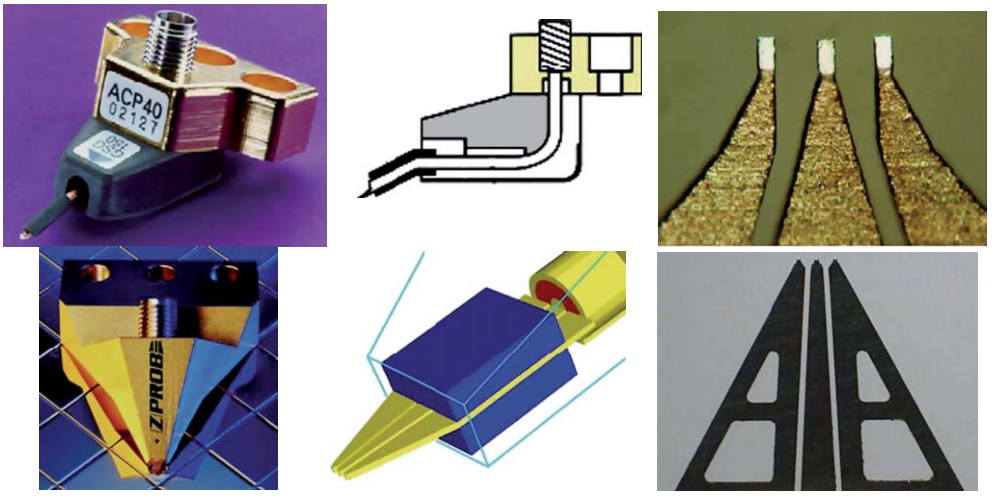


Figure 10 Close-up of a microwave probe used to electrically contact pads of a microelectromechanical systems (MEMS) die on wafer level. (Photo courtesy: Cascade Microtec.) *On the left*, microwave probe with SMA-type electrical connectors; *center*, schematic diagram of the construction of the probe; *right*, close-up view of the tips. (Photo courtesy: SUSS MicroTec.)

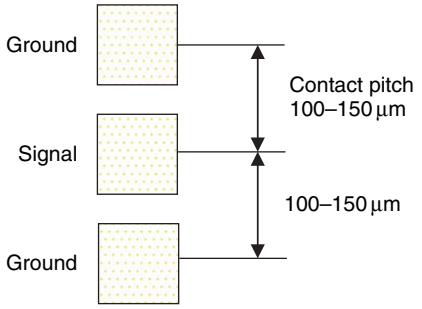


Figure 11 Typical coplanar microwave pad pattern (ground–signal–ground) for each die contact.

then an external bias tee may be preferred. A bias tee is a three-port network with the properties of a diplexer. That is a wide-bandwidth port, which is coupled to the bias port via a low-pass filter and to the RF port through a high-pass filter. Important selection parameters of a bias tee are crossover frequency, bandwidth, current and power rating, resistance and loss, and connector type.

The series and shunt resistances of the bias tee will be significant when measuring individual devices where a precise bias point is desired. For moderate to large currents, the series voltage drop will make a

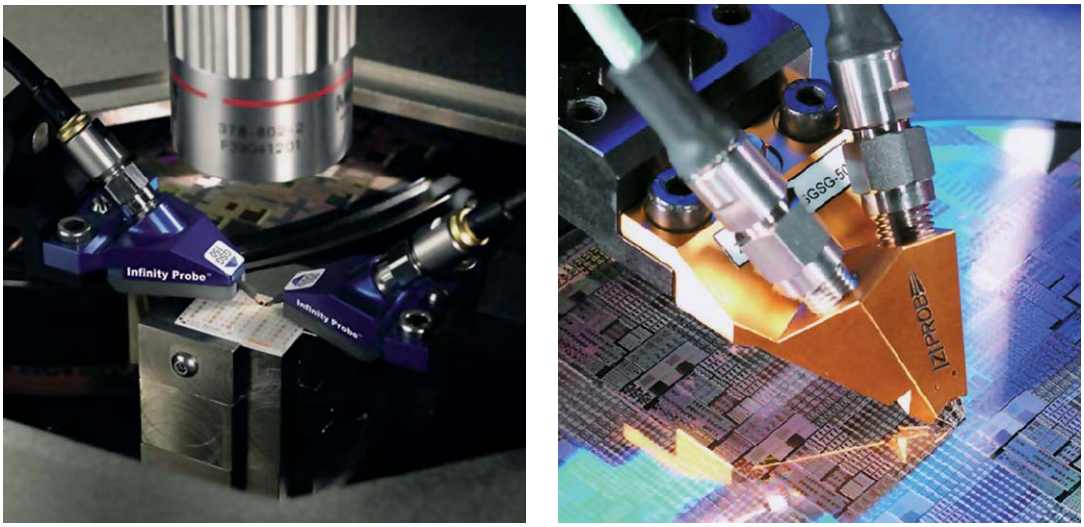


Figure 12 Close-up of microwave probes and cables on positioners, which provide planarity adjustment and cable restraint. *Left*, measurement of a calibration standard; *right*, devices on the wafer. (Photo courtesy: Cascade Microtech and SUSS MicroTec.)

change from the expected bias voltage. Actual device currents will vary from the expected ones due to the current path provided by the shunt resistance. This resistance is high, but not high by the standards of parametric measurements with currents possibly ranging from 1 fA to 100 mA. Probes, cables, and probe stations designed specifically for DC parametric measurements are generally preferred for these cases.

1.17.4.2 WL Physical Stimulation

To evaluate sensor functions or cross sensitivities of the devices, additional environmental loads are required, such as temperature, pressure or gas loads, or magnetic fields. Biotechnological and microfluidic devices frequently need additional duct systems for media. Examples will be given in the next sections.

For the WL test of integrated MEMS, three basic strategies are possible (**Table 7**). The conventional WL test of the analog and digital circuits of MEMS is identical to the WL test of ICs and will only be treated here with respect to RF testing and calibration. Physical stimulation on WL is a new technology, which is very prospective but also has limitations (**Figure 13**). The third concept aims at replacing physical stimulation by electrical excitation (**Table 7**).

This section evaluates the possibilities and requirements of physical testing of MEMS on the WL. Due to the rapid increase of the MEMS components market, also the suppliers of test equipment increased their product portfolio significantly. Consequently, wafer probing of MEMS with physical stimulation has already become the state of the art in several applications.

In **Table 8** a comparison has been made for the physical properties, which can be tested on the WL of MEMS. For many of the relevant sensor types commercial solutions exist, by which physical stimulation with the measured property is possible. Also WL test solutions for micromirror actuators and RF-MEMS are available. Unfortunately, for all the devices, strong cross sensitivities can appear, which must be taken into account. The dominant cross sensitivity of almost all devices is temperature, but also the pressure in the environment or in the device can affect the function (**Table 8**). Furthermore, several on-wafer tests are possible, which are specific for mechanical structures. These will provide information on structural features and will not characterize the full functional behavior. According to **Table 9** geometry, the topology and the binematic and dynamic behavior are tested.

Table 7 Comparison of wafer-level (WL) test strategies for MEMS

Method	*Advantage*	*Disadvantage*
Electrical testing of the analog–digital circuitry	Standard wafer probing	Only partial test of system
Physical stimulation	Realistic and comprehensive testing	Requires WL equipment
Electrical stimulation of the sensor structure	Test efficiency eliminates need for stimulation equipment	Requires design for test

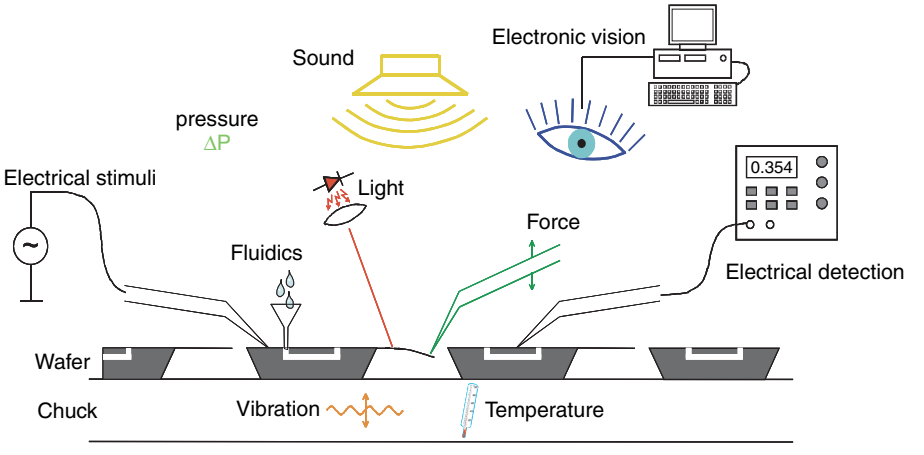


Figure 13 Examples of physical stimulation for the wafer-level testing of microelectromechanical systems (MEMS) (SUSS).

Table 8 Physical properties of MEMS, which can be tested on wafer level (WL) and dominant cross sensitivities, which must be taken into account

Device type	*Physical property*	*Main cross sensitivity*
Temperature sensor	Temperature	–
Pressure sensors	Pressure	Temperature Environmental pressure
Gas sensors	Gas composition Flow rate Pressure	Temperature
Microbolometer	Radiation power per area	Temperature
Acceleration sensors	Acceleration	Out-of-axis acceleration Mechanical stress Pressure
Gyroscopes	Yaw rate	Out-of-axis acceleration Mechanical stress Pressure
Silicon microphones	Sound pressure	Temperature
RF MEMS	Resonance frequency	Temperature, pressure, vibration, and acceleration
Micromirrors	Angular deflection	Vibration

Table 9 Additional test equipment for MEMS wafer-level test

Analysis type	*Test equipment*
3D mechanical motion analysis	Laser vibrometer or interferometer
Feature recognition	Camera systems with automated image processing
Topology analysis and profiling	Optical test systems, interferometers

1.17.4.2.1 Concepts of WL test equipment

The three basic test concepts will be explained on the example of a pressure sensor. The basic transducer element of a pressure sensor is a diaphragm, which is deflected by the pressure difference between both sides. Pressure sensors can either measure an absolute pressure using a reference cavity or a differential pressure between two pressure supplies. The working principle requires resistive strain gauge structures in order to measure the stress induced in the silicon membrane. The pressure range of silicon sensors extends from several millibars to approximately 100 bar. Measurement of some functional parameters of the sensors such as sensitivity requires a physical stimulus and that the mechanical structure be provided with a defined pressure. The available methods to apply pressure to wafers are summarized in **Table 10**:

- Use of a pressure probe on the wafer topside in addition to the electrical probes.
- A chuck that can supply a pressure or a vacuum on the wafer bottom side; or
- Operation of the complete prober station inside a pressure or vacuum chamber.

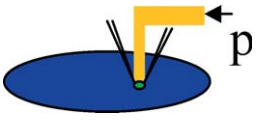
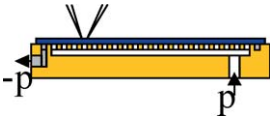
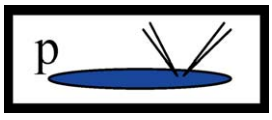
Each of these concepts has specific fields of application. Also the achievable accuracy as well as the equipment and the test costs varies significantly as compared in **Table 10**.

- Both the pressure-probe module (**Figure 14**) and the pressure chuck have the advantages that only limited changes must be made to standard equipment. Both are also limited with respect to pressure range and accuracy.
- When the complete equipment is mounted inside a pressure chamber (**Figure 16**), the system will be very versatile and many options are possible such as additional temperature control. The pressure can be controlled in a range between rough vacuum and high pressures. Such systems will be expensive, as only motorized or fully automatic wafer probers allow efficient working.

1.17.4.2.2 WL test with a stimulation via probes or via chuck

With a pressure-probe module as shown in **Figure 14**, a gas pressure is applied from the topside. A representative pressure range is from atmospheric pressure to 7 bar. An option for temperature control of the gas from ambient to 125°C is important. It must be used to prevent cooling of a sensor's diaphragm, which would lead to a strong signal disturbance. Pressure modules can be used with manual, semiautomatic, and fully automatic wafer probers.

Table 10 Methods and features of pressure sensor testing on the wafer level

Pressure application method	*Pressure probe*	*Pressure chuck*	*Pressure chamber*
Schematic			
Function	Probe on the wafer topside under impact pressure	Pressure from backside	Complete equipment inside pressure chamber
Control of cross sensitivity	Optional temperature control ambient to 125°C	Optional temperature control	Optional temperature control
Limitations	Very flexible, upgradeable	Limited pressure range, vacuum testing	Very versatile, many options
Sensor types	Absolute and differential sensors	Individual chuck pattern for each wafer pattern	Few limitations
Pressure ranges	Pressure range up to 7 bar	Rough vacuum to low overpressure	Vacuum to high pressure 50 bar
Prober type	Used with all probers	Used with all probers	Motorized or fully automatic wafer prober
Cost	Good price performance ratio	Low priced	Expensive

Source: Data assembled on the basis of information from SUSS.

Figure 14 Pressure-Probe Module Suess PPM, which applies pressure from the topside. (Photo courtesy: SUSS.)

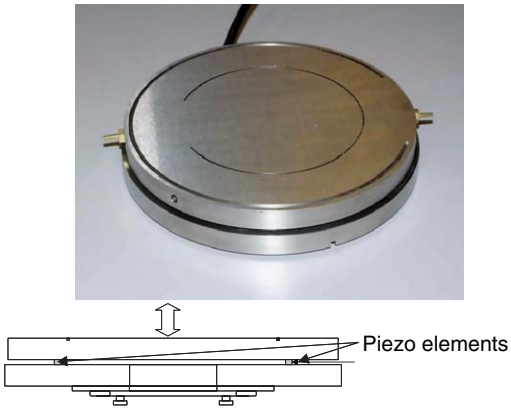


Figure 15 Acceleration chuck for wafer probing of acceleration sensors. *Upper*, photo; *lower*, drawings are a schematic cross section. (Photo courtesy: SUSS.)

In automobiles, acceleration sensors are widely used for comfort, safety, and security applications. The relevant range spans from less than 1 *g* to hundreds of *g*. For WL stimulation an acceleration chuck has been developed, which generates out-of-plane vibrations of the wafer (**Figure 15**). According to the manufacturer (SUSS) typical acceleration values are 3–5 *g* at 500 Hz, 14 *g* at 1 kHz, and more than 50 *g* at 2 kHz. In this frequency range, deflections between 100 and 200 µm are generated, which still can be tolerated by compliant probe tips.

1.17.4.2.3 WL testing in a pressure/vacuum chamber

For wafer probers that are mounted inside a pressure chamber, only motorized or automated systems are sensible. Typical data of commercial systems with a pressure chamber are as follows (SUSS):

- Regulation of pressure in a range from rough vacuum to 50 bar.
- Controlled temperature by a thermal chuck from −65°C to 250°C.

- Nonaggressive gas loads or controlled humidity are possible.
- Wafer sizes up to 200 mm.
- Eight probe manipulators with four RF probes up to 65 GHz, four to eight DC probes or alternatively probe cards can be equipped.

Many MEMS devices need to be tested in vacuum or with a special well-controlled atmosphere. The most important of these are RF-MEMS, optical switching micromirrors, resonator elements, and inertial sensors including accelerometers and gyroscopes. Also microbolometers for IR imaging, pressure sensors, chemical sensors, and many MEMS designed for aerospace applications must be tested at low pressures. If the testing has to be performed on WL, a special vacuum station is needed. Such a device encloses an entire probe station with micromanipulators in a vacuum chamber (**Figure 16**).

A vacuum probe station must be fully motorized either with manual or computer control. The control provides the x–y-stage movement, the platen z-lifting and lowering, x–y-microscope movement, and manipulator x–y–z-movement. Depending on the vacuum chamber, a pressure range from high vacuum (10^{-7} mbar) up to atmospheric pressure can precisely be regulated (**Figure 16**).

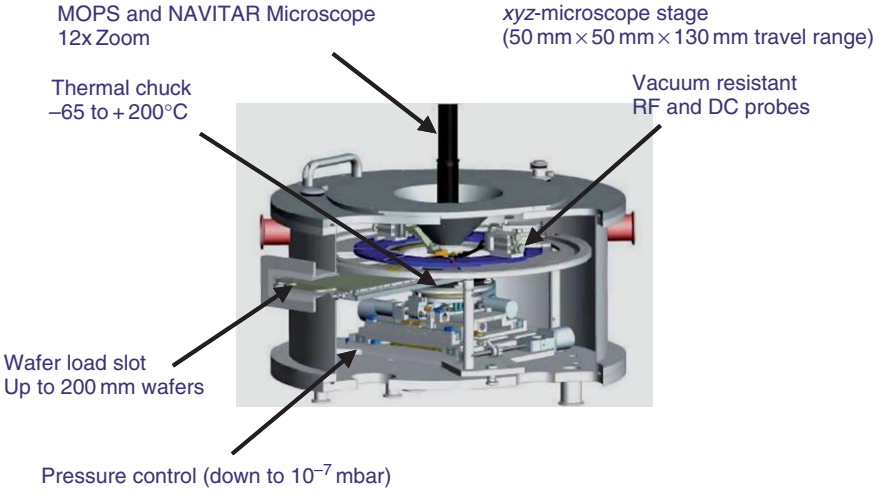


Figure 16 *Above*, Vacuum probe station; *below*, vacuum chamber of a low-pressure semiautomatic probe station. (Photo courtesy: SUSS.)

1.17.4.2.4 Applications of WL testing

1.17.4.2.4.(i) Gas sensors MEMS gas sensors are becoming more and more important due to legislative restrictions as well as due to safety and comfort requirements. There is an increasing demand to monitor exhaust and hazardous gases such as CO_X, NO_X, and others in automotive, industrial, or domestic applications. Using a vacuum/low-pressure vessel and mass flow controllers, contact-free and exact dosage of gases down to the parts per million range is possible.

An economic alternative is to supply the gases to wafers bearing gas sensors using a gas nozzle in combination with the electrical probes (**Table 10** and **Figure 14**). Unfortunately, this only allows quantitative measurements of limited accuracy and should be restricted to nontoxic media.

1.17.4.2.4.(ii) RF-MEMS For MEMS operating at high frequencies, testing in a vacuum environment is mandatory. For mechanical switching devices, the vacuum will avoid sticking effects. Also the vacuum will prevent damping of mechanical structures and electrical signals caused by the viscosity of the air. Elimination of a gas can also minimize acoustic interference within the cavities. For such tests a semiautomatic vacuum prober with a sufficient number of manipulators and RF probes is recommended. Probes for frequencies up to 100 GHz are available. Also many applications require an additional thermal chuck for temperatures from −65°C up to 250°C. Additional features of such test equipment, which are already in use, are as follows:

- Laser trimming capability
- Mechanical motion analyzer

1.17.4.2.4.(iii) MEMS infrared arrays Microbolometers are used to measure the temperature distribution of objects emitting IR radiation. For testing, such devices are exposed to IR radiation with controlled wavelength via an IR transmission window. The IR radiation is generated with IR black bodies, which are mounted on a shutter. Also the temperature of the wafer must be controlled. Here the test assembly is mounted into a vacuum environment to prevent heat transfer by convection. For the sensor wafers, special probe cards are used with front-end electronics on top and close to the tips in order to suppress wiring artifacts.

1.17.4.3 Electrical Failure Detection and Self-Test

1.17.4.3.1 Detection of faults by electrical measurements

Many manufacturing defects in MEMS devices will lead to topological or geometric deviations from the intended structures. In many cases, faulty structures can be detected by methods such as automated microscopic inspection using image processing and pattern recognition software. In order to avoid too many different tests, it is desirable to detect defects using electrical measurements on the WL. A very efficient strategy is the detection of faults by electrical stimulation. This concept has been tested by Dumas *et al.* (2006) on bending beams, which were produced by BMM and which are used as magnetic Lorentz force sensors. The possible classes of defects after free etching as well as their causes in the process are listed in **Table 11**.

The principal questions to be answered are as follows:

- Which physical phenomenon or property is sensitive to the defect?
- How can it be stimulated electrically?
- Which reaction can be detected electrically?

To check the feasibility of electrically stimulated tests, physical modeling of the defect and subsequent simulation using the finite element method may be applied. In this way the sensitivity to typical defects is evaluated. For the reported magnetic sensor there are two phenomena, which are affected by the etch defect:

(1) In case of an incomplete etch release, the heat conduction from a beam will be increased and also the effective thermal mass will be higher. As a result, the time constants of heating and cooling will be affected as well as the temperature rise per amount of heat.
(2) A second effect is that a free beam will have characteristic resonance frequencies, which will be changed by fixing the structure.

In the first case the beam must carry a coil, which can be heated by an electrical pulse while the coil's

Table 11 Classification of defects in the deep-release etching of beam structures

Structural defect	*Cause*
Fully released	Etch okay
Pointwise fixed	Etching almost complete
Strongly fixed	Partial etching
Not released	No etching

Source: Dumas *et al*. (2006).

transient resistance change is monitored. In the second case the beam is subjected to a harmonic stimulation by repeated heat pulses. At a characteristic frequency resonance occurs and can be detected using the strain gauge structures, which are normally used for deflection measurements. Also other micromachined sensor structures have been successfully tested such as accelerometer paddles, which are also sensitive with respect to transient thermal behavior and resonance frequency (**Table 12**).

In principle, an electrical test is possible to detect certain defects even after full processing and encapsulation. Nevertheless, this concept requires design for testing and is only feasible for a certain portion of all possible failures.

1.17.4.3.2 Built-in self-test

In many cases stimulation of a microsystem by the property to be measured, such as pressure or acceleration, is not convenient due to practical reasons. Physical testing often becomes very demanding when it must be performed on the WL at high or low temperatures and when the device is already in the use phase. Therefore, the concept of the built-in self-test (BIST) has been developed.

Several years ago band-gap references were introduced for selfcalibration of operational amplifiers or analog–digital converters. Besides calibration, BIST can also be used to detect whether a device is defective or if it has failed. In the case of MEMS, such a concept requires provisions to stimulate the sensing structure internally. Although BIST is still more a concept, it has already been used in industrial MEMS (Charlot *et al.* 2001), **Table 13**. One of the successful techniques is the impulse response method in combination with pseudo-random testing (Rufer *et al.* 2005). BIST usually requires that a monolithic MEMS device will be integrated with the test circuitry.

1.17.5 Calibration and Compensation

In this section all errors which limit the accuracy of a NWA measurement will be discussed and the most used calibration and verification methods will be presented.

There are three sources of errors, which limit the accuracy of any measurements: systematic errors, statistical or random errors, and drift errors (**Figure 17**). Statistical errors are caused by a low signal-to-noise ratio during the measurement and by a nonrepeatability of the measurement system. They can be minimized with averaging or adequate filtering of measurement data.

Systematic errors are assumed to be time invariant and thus predictable. They are mainly caused by imperfections in measurement equipment and setup. Systematic errors are fully eliminated by calibration. It is inevitable to use calibration in most systems that interface with the real world. Unfortunately, testing

Table 12 Deep-release etching of beam structures – results of electrothermal detection of defects by heat pulses and repetitive heating

Structural defect	*Time constant (ms)*	*Magnitude*	*Cut-off frequency (Hz)*	*Flat-band response (dB)*	*Thermally induced resonant peak*
Fully released	4.2	2.6 V	19	38	Yes
Pointwise fixed	4.2	2.6 V	19	38	No
Strongly fixed	1.5	1.1 V	57	31	No
Not released	0.2	17 mV	600	−5	No
Test	Pulse heating		Repetitive heating		

Source: Dumas *et al.* (2006).

Table 13 Examples of successful implementation of built-in self-test (BIST) of MEMS

Device	*MEMS sensor principle*	*BIST principle*	*Purpose*
Accelerometer	Capacitive comb structure (SMM)	Deflection by electrostatic forces	Fault detection, test sensitivity
Accelerometer	Seismic mass on bending beam (BMM)	Thermal bimetal deflection	Fault detection, test sensitivity
IR sensor gas-flow sensor	Array of thermopiles (BMM)	Built-in heater structure	Test function and sensitivity of all pixels

Source: Charlot *et al.* (2001).

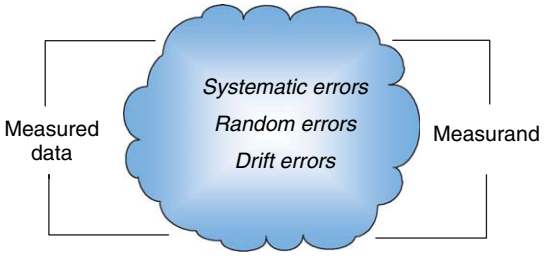


Figure 17 Three sources of errors, which limit the accuracy of a measurement. (Drawing courtesy: Agilent Technologies.)

and calibration is time-consuming and accounts for a large portion of the product generation costs.

Individual calibration of microsystems is used to reduce the effects of individual variations in the characteristics due to variations and drift in the fabrication process over time.

Most microsystems are calibrated using accurate operational test measurement at the end of the fabrication process. The calibration is done either on the hardware or on the software level. For hardware calibration, the characteristics of the device are changed by a trim process using a laser or similar techniques. For software calibration, the calibration data are stored in a nonvolatile memory in the device.

Both hardware and software calibration are limited by drift errors, which arise due to a change of the system performance after the calibration has been done. Drift errors are caused primarily by temperature variation and by aging. They are eliminated by a proper recalibration. A new approach uses an online calibration technique of microsystems similar to those implemented on analog-to-digital converters. Such an online calibration is performed continuously during operation of the device and should not interfere with the normal operation of the system, apart from making it more accurate.

1.17.5.1 Vector Network Analyzer

The VNA has become the workhorse of most on-wafer measurements above 1 GHz. This is because at high frequencies direct measurements of voltage and current become difficult and so the stimulus–response approach has been developed for the high-frequency regime. For getting the best on-wafer measurement results, a solid understanding of measurement system components and their interaction is required. Due to drift of the internal electrical parameters of any VNA and due to variations of the electrical contact of the DUT to the VNA, the measurement system including the VNA must be calibrated before any meaningful measurement.

A VNA measures vector ratios of reflected or transmitted energy to the energy incident upon the DUT. As a stimulus–response measurement, a VNA measurement determines the properties of devices rather than the properties of signals. Signals will be measured using instruments such as oscilloscopes or spectrum analyzers.

The block diagram (**Figure 18**) shows the key elements of the VNA. The RF source at the top provides the stimulus a_1 or a_2 for the DUT. A forward/reverse

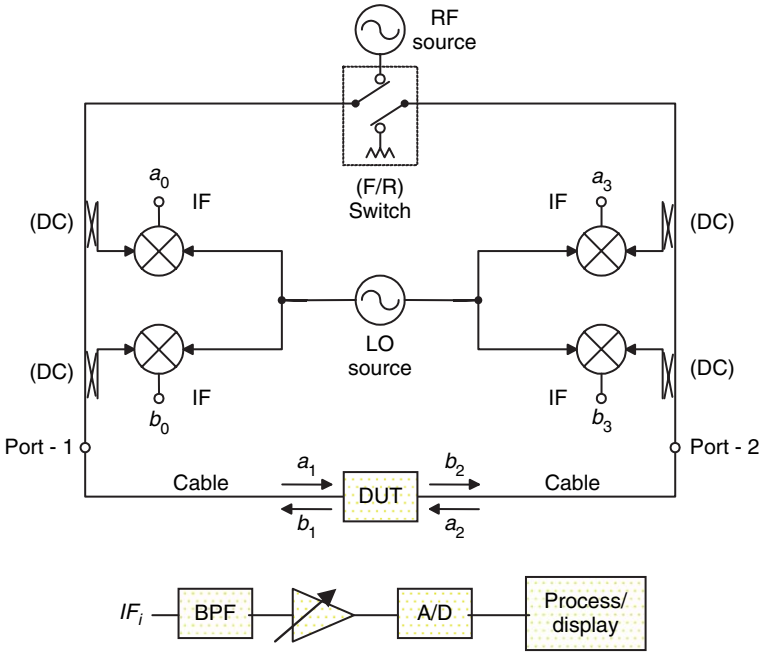


Figure 18 Main hardware blocks in a vector network analyzer (VNA).

(F/R) switch directs the RF energy to either DUT port. The test set uses directional couplers (DCs) or bridges to pick off the forward (a_1, a_2) and reverse (b_1, b_2) waves traveling to and from each port, and down-converts these signals to four intermediate frequency (IF) sections (a_0, a_3, b_0, b_3). Each IF section IF_i band pass filters, amplifies, and digitizes (A/D) the signals for further digital processing and display.

A significant challenge in stimulus–response analysis is to define exactly where the measurement system ends and the DUT begins. In on-wafer VNA measurements, this boundary is known as the reference plane of the measurement and will often be located at the probe tips. Ideally, the system will measure the characteristics of whatever is connected to the measurement reference plane. A typical wafer probing VNA setup is shown in **Figure 19**.

If we ignore any nonideal behavior of the cables, couplers, mixers, and so forth, then the ratios of wave amplitudes (a and b) inside the machine correspond directly to the so-called S-parameters of the DUT. For example:

$$\frac{b_0}{a_0} = S_{11} \text{ and } \frac{b_3}{a_0} = S_{21}$$

of the DUT.

1.17.5.1.1 Impedance measurements

S-parameters are just one type of network representation used for linear, small-signal AC analysis. A reflection ratio at a device port, with all other ports terminated, such as S_{11} or S_{22}, is also known as a reflection coefficient. The load impedance, Z_L, is related to the reflection coefficient, Γ, by the bilinear transform pair as follows:

$$Z_L = Z_0\left[\frac{1+\Gamma}{1-\Gamma}\right] = r + jx, \quad \Gamma = \frac{b_i}{a_i} = \frac{Z_L - Z_0}{Z_L + Z_0}$$

where Z_0 is the reference impedance of the measurement system. S-parameters can be converted to, e.g., Z-parameters (or to y-parameters, h-parameters, etc.) and back:

$$S_{11} = \frac{(Z_{11}-1)(Z_{22}+1) - Z_{12}Z_{21}}{(Z_{11}+1)(Z_{22}+1) - Z_{12}Z_{21}}$$

$$Z_{11} = \frac{(1+S_{11})(1-S_{22}) + S_{12}S_{21}}{(1-S_{11})(1-S_{22}) + S_{12}S_{21}}$$

$$S_{12} = \frac{2Z_{12}}{(Z_{11}+1)(Z_{22}+1) - Z_{12}Z_{21}}$$

$$Z_{12} = \frac{2S_{12}}{(1-S_{11})(1-S_{22}) + S_{12}S_{21}}$$

$$S_{21} = \frac{2Z_{21}}{(Z_{11}+1)(Z_{22}+1) - Z_{12}Z_{21}}$$

$$Z_{21} = \frac{2S_{21}}{(1-S_{11})(1-S_{22}) + S_{12}S_{21}}$$

$$S_{22} = \frac{(Z_{11}+1)(Z_{22}-1) - Z_{12}Z_{21}}{(Z_{11}+1)(Z_{22}+1) - Z_{12}Z_{21}}$$

$$Z_{22} = \frac{(1+S_{22})(1-S_{11}) + S_{12}S_{21}}{(1-S_{11})(1-S_{22}) - S_{12}S_{21}}$$

Figure 19 Two typical radio frequency (RF) device characterization setups are shown above. On the left are vector network analyzers; in this case an Agilent PNA system. Cables from the test set ports route over the probe station to a pair of microwave probes, which are aligned through the microscope or camera system onto calibration substrates and test devices. The right sides show the monitor of the PCs, which controls the prober and runs the calibration and measurement software. (Photos courtesy: Cascade Microtec. and SUSS MicroTec.)

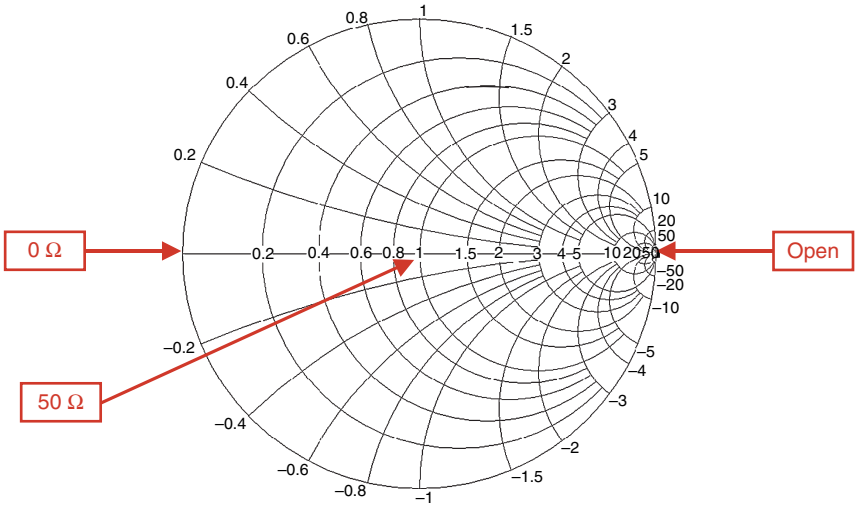


Figure 20 The Smith chart is a polar graph of the reflection coefficient Γ with grid lines showing normalized impedance values Z_L/Z_0.

Typically the VNA automates the display of a wide variety of formats: Smith chart, polar, and linear and log magnitude formats. The Smith chart (Figure 20) is a polar graph of the reflection coefficient Γ with grid lines showing the normalized impedance values Z_L/Z_0. All passive impedances are represented in a compact unit circle reflection format.

The Smith chart is very useful in microwave theory, as it visualizes the influence of a reactance or susceptance (Figure 21). The Smith chart can be used like a roadmap for matching purposes.

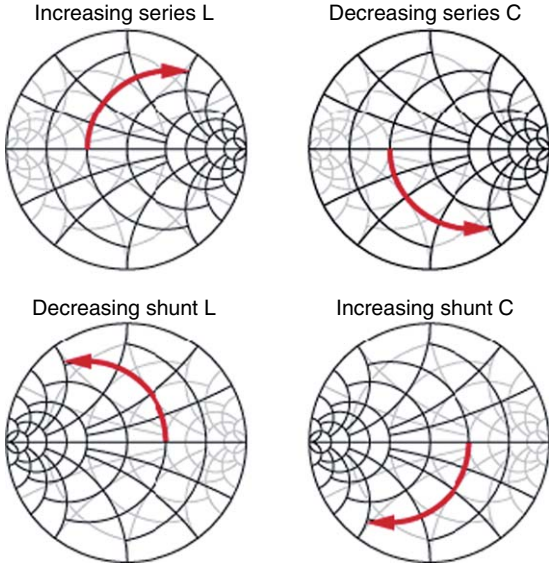


Figure 21 The paths of circuit elements in the Smith chart: The graphs demonstrate how shunt and series reactances change the circuit reflectance on the Smith chart.

1.17.5.2 Calibration Techniques

This section introduces the concept of calibration or systematic error correction of VNA.

1.17.5.2.1 VNA calibration fundamentals

VNA calibration is the process of measuring devices with known or partly known characteristics and using these measurements to establish the measurement reference planes. VNAs rely on calibration for accuracy even in measurements with the reference plane defined at the instrument front panel connectors or at the ends of cables. Calibration also corrects the imperfections of the measurement system. These imperfections include not only the nonideal nature of cables and probes but also the internal characteristics of the VNA itself.

The VNA is calibrated in much the same manner that the zero function on an ohmmeter subtracts out the resistance of the test leads. On an ohmmeter, when the zero function is activated, it stores a resistance measurement, which is then subtracted from all future measurements. The obvious error model is simply a resistance in series with the test port.

Where the ohmmeter's calibration simply used a short circuit (connecting the test leads together) to determine the extra resistance term in the error model, a VNA uses multiple calibration standards – typically open circuits, short circuits, loads, and through (thru) connections.

A VNA does the same thing as the ohmmeter but uses a more complex error model with several terms for each frequency point. The measurement system is

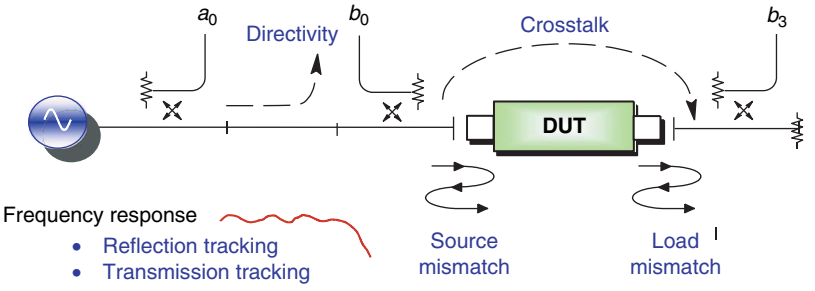


Figure 22 Six error terms are needed for the forward model. (Drawing courtesy: Agilent Technologies.)

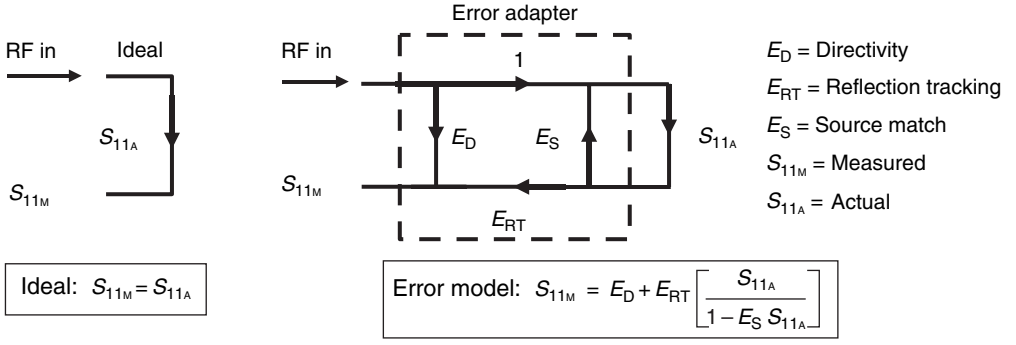


Figure 23 One-port model of reflection. (Schematic courtesy: Agilent Technologies)

described as an ideal VNA with an error adapter network that models all of the system's nonidealities: directivity of the couplers, imperfect match looking back into the reflectometer (test set ports), imperfect frequency response of the reflectometers and the transmission between ports, and the cross talk between ports (**Figure 22**).

b_0/a_0 gives the measured value S_{11M} of the actual reflectivity S_{11A} of the DUT. **Figure 23** shows the one-port error model in detail. The actual reflectivity S_{11A} is modified by the directivity E_D of the VNA, which leads to some signal b_0 without any reflected signal from the DUT, by a not fully source match E_S of the VNA cable, which together with S_{11A} might cause some standing waves, and by the reflection tracking E_{RT}, which causes a damping of the signal together with a phase shift. The damping and phase shift of the forward signal can be fully considered also in a proper choice of E_S and E_{RT}; thus, without any restriction we can set the forward signal path of the error adapter equal to 1.

To solve for these error terms, we measure three standards in the calibration to generate three equations for the three unknowns. This equation system is then solved at the end of the calibration routine, and we get the complex values of E_D, E_S, and E_{RT}. The corrected reflectivity S_{11A} can then be calculated from S_{11M} by the following equation:

$$S_{11M} = E_D + E_{RT}\left[\frac{S_{11A}}{1 - E_S S_{11A}}\right]$$

Figure 24 shows the effect of calibration of a measured one-port S_{11}. The source mismatch leads to a standing wave pattern in the uncalibrated data, which is effectively eliminated by the calibration. The full error model for a VNA is shown in **Figure 25** and the corresponding signal flow graph in **Figure 26**.

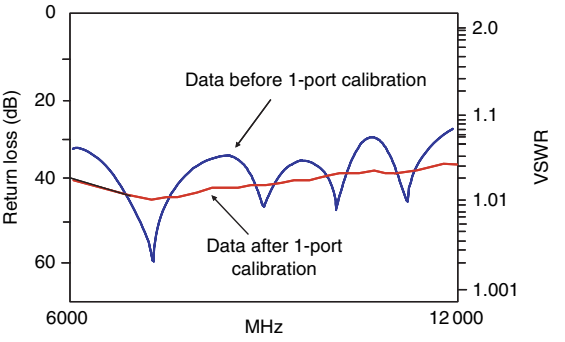


Figure 24 Measured return loss and voltage standing wave ratio (VSWR) of a device under test (DUT) before and after a one-port calibration. (Figure courtesy: Agilent Technologies.)

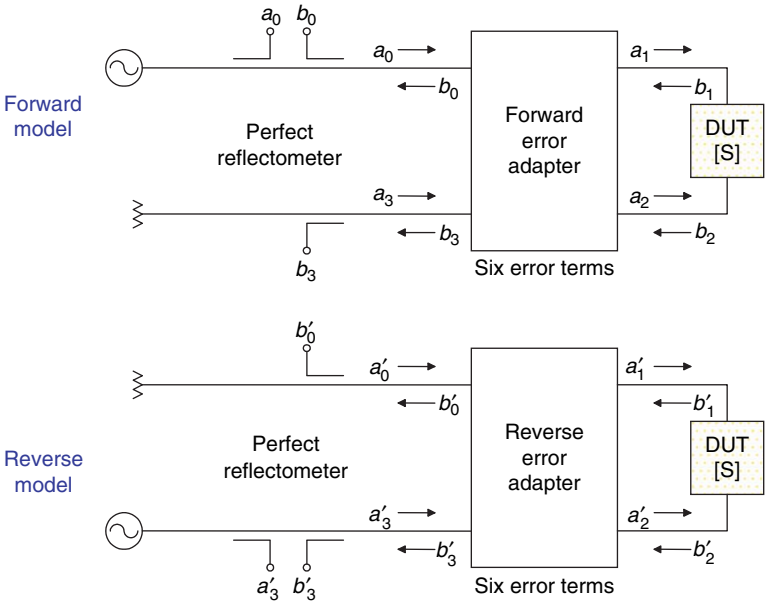


Figure 25 Error model for a typical network analyzer. All errors are modeled using an error adapter in front of an otherwise perfect system.

The full error model, shown in **Figures 25 and 26**, results in 12 unknown error terms. To solve these error terms, we measure 12 standards in the calibration to generate 12 equations for the 12 unknowns. The corrected four S-parameters of a two-port DUT (S_{11A}, S_{12A}, S_{13A}, S_{14A}) are then calculated from the measured values (S_{11M}, S_{12M}, S_{13M}, S_{14M}) by (Agilent)

$$S_{11A} = \frac{\left(\frac{S_{11M}-E_D}{E_{RT}}\right)\left(1+\frac{S_{22M}-E_{D'}}{E_{RT'}}E_{S'}\right)-E_L\left(\frac{S_{21M}-E_X}{E_{TT}}\right)\left(\frac{S_{12M}-E_{X'}}{E_{TT'}}\right)}{\left(1+\frac{S_{11M}-E_D}{E_{RT}}E_S\right)\left(1+\frac{S_{22M}-E_{D'}}{E_{RT'}}E_{S'}\right)-E_{L'}E_L\left(\frac{S_{21M}-E_X}{E_{TT}}\right)\left(\frac{S_{12M}-E_{X'}}{E_{TT'}}\right)}$$

$$S_{21A} = \frac{\left(\frac{S_{21M}-E_X}{E_{TT}}\right)\left(1+\frac{S_{22M}-E_{D'}}{E_{RT'}}(E_{S'}-E_L)\right)}{\left(1+\frac{S_{11M}-E_D}{E_{RT}}E_S\right)\left(1+\frac{S_{22M}-E_{D'}}{E_{RT'}}E_{S'}\right)-E_{L'}E_L\left(\frac{S_{21M}-E_X}{E_{TT}}\right)\left(\frac{S_{12M}-E_{X'}}{E_{TT'}}\right)}$$

$$S_{12A} = \frac{\left(\frac{S_{12M}-E_{X'}}{E_{TT'}}\right)\left(1+\frac{S_{11M}-E_D}{E_{RT}}(E_S-E_{L'})\right)}{\left(1+\frac{S_{11M}-E_D}{E_{RT}}E_S\right)\left(1+\frac{S_{22M}-E_{D'}}{E_{RT'}}E_{S'}\right)-E_{L'}E_L\left(\frac{S_{21M}-E_X}{E_{TT}}\right)\left(\frac{S_{12M}-E_{X'}}{E_{TT'}}\right)}$$

$$S_{22A} = \frac{\left(\frac{S_{22M}-E_{D'}}{E_{RT'}}\right)\left(1+\frac{S_{11M}-E_D}{E_{RT}}E_S\right)-E_{L'}\left(\frac{S_{21M}-E_X}{E_{TT}}\right)\left(\frac{S_{12M}-E_{X'}}{E_{TT'}}\right)}{\left(1+\frac{S_{11M}-E_D}{E_{RT}}E_S\right)\left(1+\frac{S_{22M}-E_{D'}}{E_{RT'}}E_{S'}\right)-E_{L'}E_L\left(\frac{S_{21M}-E_X}{E_{TT}}\right)\left(\frac{S_{12M}-E_{X'}}{E_{TT'}}\right)}$$

Each actual S-parameter is a function of the four measured S-parameters. Thus, the VNA must perform a forward and a reverse sweep to update each S-parameter. The calibration functions allow the user to store measurements of standards, compute the error models, and automatically apply corrections to DUT measurements. Corrected VNA measurements are sometimes referred to as de-embedded. **Figure 27** shows as a comparison the uncorrected, response calibrated, and full two-port calibrated insertion attenuation of a filter.

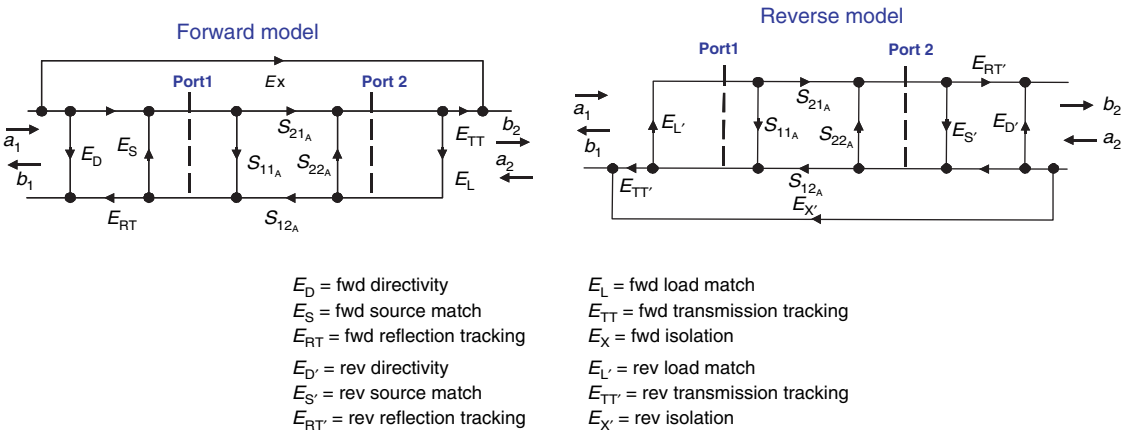


Figure 26 Signal-flow diagram of the error model for a typical network analyzer. (Schematic courtesy: Agilent Technology.)

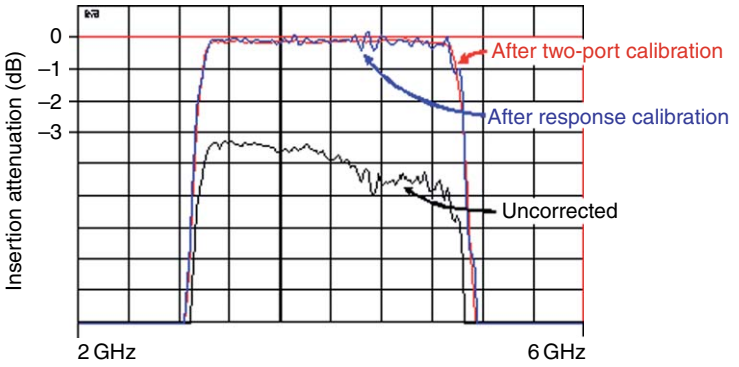


Figure 27 Response calibration versus two-port calibration of filter insertion attenuation. (Schematic courtesy: Agilent Technology.)

1.17.5.2.2 Calibrated VNA accuracy

As with any corrected measurement, the absolute accuracy of a calibrated VNA measurement is determined by the techniques and completeness of the error model used as well as by the accuracy of the description of the reference devices (calibration standards), and by the repeatability of the measurement system.

Improper calibration will introduce errors. The impedances used in the calibration must be accurately known and entered. Assumption of ideal behavior of standards is a recipe for errors, but worse things can happen. For example, the improper entry of a description of a short-circuit or open-circuit standard will lead to useless results that may not always be obviously wrong. The calibration routine of the VNA will accept everything that is entered. Accurate descriptions of the electrical behavior of the various calibration standards must be supplied to the VNA.

Corrected measurements rely on the repeatability of the measurement system. Just as an ohmmeter cannot correct a changing resistance of its test leads or drift in the ohmmeter circuitry, a VNA cannot correct for random errors such as noise or dynamic range, cable repeatability, or instrument drift. Any change in the measurement due to, for instance, drift of the VNA test set, thermally induced cable length changes, or even the effects of noise due to VNA dynamic range can invalidate the correction. Sensitivity of cable electrical performance (such as phase delay), of the microwave probe contact (such as ohmic resistance), or even of the calibration elements by environmental changes is a significant element of quality and suitability for VNA use.

VNA error models are based on the use of *S*-parameter representations of network properties. *S*-parameters are signal-flow- and transmission-line-based. Only a single mode of propagation at the device terminals is assumed. Situations that violate this assumption such as using waveguides that can propagate multiple modes, radiation, or parasitic coupling

between networks other than at the device terminals are not properly handled. If a second mode, radiation, or extra parasitic does not change for any and all devices being measured then it will drop out with the calibration. These extra modes, however, usually show a behavior that is DUT-dependent and the VNA calibration will not account for these effects. A well-designed probing system with good quality transmission-like interconnections will minimize these errors as much as possible (**Figure 11**). An extensive discussion of VNA calibration and accuracy is given (e.g., NN Calibration and Accuracy Factors – Summit High – Frequency Probe Station Reference Manual. Cascade Microtech, part number 103–375A).

1.17.5.2.3 VNA calibration options and standards

If not calibrated, the VNA will measure the uncorrected *S*-parameters (**Figure 28**). Uncorrected parameters are very convenient to measure, but not very accurate and, thus, rarely used. A response calibration is simply a vector magnitude and phase normalization of a transmission or reflection measurement. Thus, it removes frequency response errors. It is sometimes used when highest accuracy is not required (**Figure 29**).

Typically, a full calibration of all the error parameters is used as a reference or to assure the highest accuracy (**Figure 30**) (Rytting 1986). For a two-port full error calibration each port must be connected to some well-known standards, like an open (O), a short cut (S), and a load (L). Additionally, both ports must be connected using a through connection. A full calibration removes all errors concerning directivity, source match, load match, reflection tracking, transmission tracking, and optional, also cross talk.

Figure 28 Uncorrected vector network analyzer (VNA) measurement.

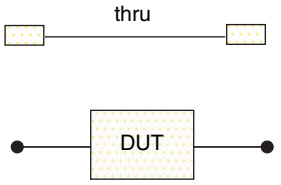


Figure 29 *Above*, Response calibration; *below*, measurement.

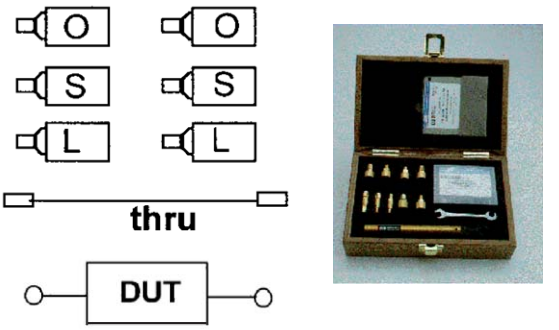


Figure 30 Full two-port calibration elements. *Left*, schematic; *right*, calibration standards (SOLT) for SMA adapters (Agilent Technologies).

1.17.5.2.4 Calibration standards

On-wafer calibration standards most often are precision thin-film resistors, short-circuit connections, and 50-Ω transmission lines fabricated either on the wafer containing the DUT or on a separate calibration substrate (typically alumina). High-performance load resistors are trimmed to within 0.1% (SUSS MicroTec) of their desired DC-value (usually 50 Ω).

A true thru standard does not exist for on-wafer measurements because probes cannot directly connect to each other and must instead use a very short transmission line as a thru standard. **Figure 31** shows typical short-load-thru (ISS (impedance standard substrate) family, Cascade Microtech) and short-open-load-thru (SOLT) (CSR family, SUSS MicroTec) substrates.

It is important to note that some specific electrical behavior of the standards depend on the probe pitch used. The calibration data are therefore supplied with the probe, rather than being a single value for a standard as would be the case for coaxial standards where there is no ambiguity about what connections can be made to it. Calibration data are normally specified for a specific probe spacing used with a particular impedance standard substrate, and is only valid for a particular probe–calibration standard combination. The contact pattern, e.g., GSG or GS (see **Figures 11, 38 and 39**), will also impact calibration data. Important for a wide application range of a calibration standard is a good temperature stability of the load resistance and a high accuracy of the characteristic impedance of the through standard over a wide frequency range (see **Figure 32**).

On-wafer standards, fabricated on the same wafer as the DUT, are sometimes desirable since the probe-to-standard transition can be designed to be very similar to the transition to the DUT. This is helpful

Figure 31 A calibration substrate has multiple repeated patterns of Thru-Short-Load Standards (*left*, impedance standard substrate (ISS) family, Cascade Microtech), Thru-Short-Open-Load (*right*, CSR family, SUSS MicroTec), and alignment marks for setting proper probe separation. A set of transmission lines is also provided for Thru-Reflect-Line (TRL) Calibration.

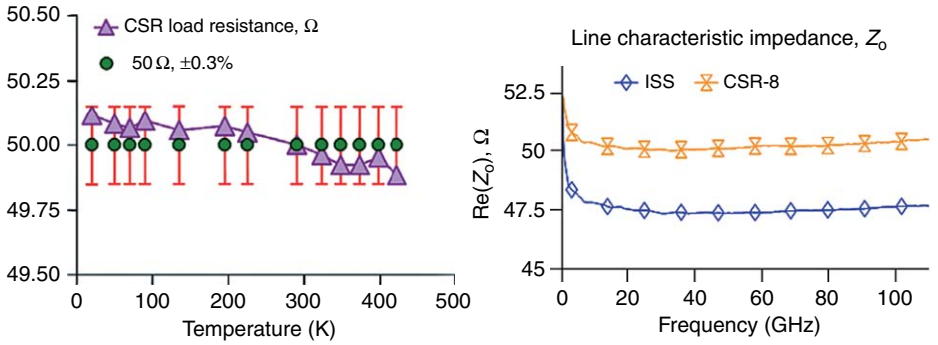


Figure 32 Typical temperature stability of load resistance and accuracy of the thru standard of coplanar calibration substrates (Geissler, Rumiantsev *et al.* 2006).

at frequencies above 20 GHz since a primitive transition may introduce extra parasitic elements or modes, which, since they are occurring at the probe-tip reference plane, may not be corrected by calibration. In many cases these transition errors may be de-embedded using simple lumped element equivalent circuit models (Walker and Williams 1997).

The use of off-wafer standards for calibration is very convenient and calibration results have been demonstrated, which are comparable to results obtained by using the special on-wafer standards methods recommended by the US NIST. In cases where the transition on-wafer is dramatically different to that of the calibration standard substrate, such as on silicon substrates, the use of a set of dummy pads on-wafer allows the removal of additional parasitic elements of the pads. Algorithms to remove pad parasitic are discussed (e.g., Van Wijnen).

1.17.5.2.5 Short-Open-Load-Thru calibration

By far the most commonly used method is the SOLT calibration (**Figures 33 and 34**). This calibration is the combination of 2 one-port short-open-load calibrations with additional measurements of a thru standard to complete the two-port calibration.

The SOLT standard elements are reasonably well modeled with simple lumped elements: open-circuit capacitance, short-circuit inductance, load inductance, and thru delay (and loss). All of the standards must be fully known and specified. This is very difficult to achieve for highly reflective standards (open, short) at higher frequencies,

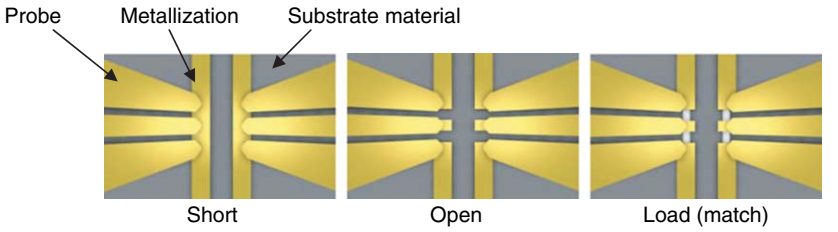


Figure 33 Close-up view of coplanar lumped calibration standards. (Figure courtesy: SUSS.)

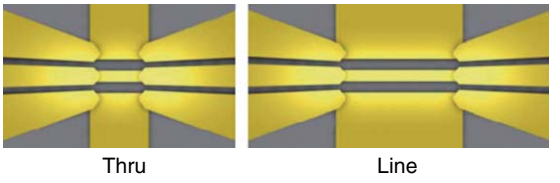


Figure 34 Close-up view of coplanar distributed calibration standards. (Figure courtesy: SUSS.)

thus rendering SOLT impractical for measurements above 20 GHz.

1.17.5.2.5.(i) Advanced calibration methods

An alternative to the SOLT calibration is the SOLR (short-open-load-reciprocal) calibration (Ferrero 1992). This algorithm is similar to the SOLT calibration for one-port corrections, but the redundant information available in the standards set allows the reciprocal standard to be any reciprocal two-port network. No other knowledge of the standard is required to complete the calibration. The SOLR algorithm is thus very powerful when only nonideal thru standards are available, like orthogonal probes, excess thru line length, glossy, or mismatched thru lines, or incompatible port geometries (pitch). However, all reflection standards should be fully defined, like SOLT, limiting calibration accuracy to 20 GHz.

The thru-reflect-line (TRL) calibration algorithm uses the characteristic impedance of a transmission line as the reference for calibration (see **Figure 34**) (Engen and Hoer 1979). Line pairs differing only in length are measured along with a reflect standard that is known only to the sign of the reflection coefficient. The line length difference must not be near 180° of phase since a half wavelength transmission line mimics a zero length line and provides no additional information. In practice a line pair will only provide useful results over a 20–160° phase difference range. Wider bandwidths require multiple line lengths to provide suitable phase differences over a wide frequency range.

In the TRL calibration (and its variations), the thru line must be fully known to allow the measurement reference plane to be moved to the probe tips from the midpoint of the thru line. In some applications using on-wafer standards, the DUT is effectively embedded at the thru midpoint, eliminating the need for this reference plane change. The TRL calibration process itself does not determine the characteristic impedance that is the reference for the resulting *S*-parameters. Additional assumptions and measurements may be used to determine the Z_0, and renormalize the *S*-parameters to 50 Ω (for example).

TRL calibration requires programmable probe positioners for fully automatic calibration. Dispersion, skin effects, losses, and fabrication tolerances of planar lines as well as requirements for multiple lines and probe repositioning significantly reduce the application and the frequency range of TRL at the WL.

The line-reflect-match (LRM) calibration was developed to overcome the limitations of the TRL calibration. Instead of using a set of different lines, the lumped match (load) element is used for calibration. However, good calibration accuracy can be guaranteed only if using purely resistive, highly symmetrical 50-Ω loads, which is impossible to fabricate at WL. Therefore, the accuracy of LRM significantly decreases as in-pair asymmetry of the load standard and measurement frequency increase.

A variation of the LRM, the line-reflect-reflect-match (LRRM) calibration provides calibration accuracy comparable to the NIST multiline TRL algorithm described below without the need for changing probe separation (Davidson *et al.* 1990, Pence 1993). The standards used for LRRM are identical to those used for SOLT but without the requirement for careful specification of the short and open standard behavior (see **Figure 33**). In LRRM, only the thru line delay (and loss for longer lines) and load standard must be specified. By using only one Match standard any ambiguity in load inductance between ports is eliminated. However, LRRM applies assumptions about the match. Namely, the match measured on one port is assumed

to behave identically on the other port. This can lead to an error in the return loss measurement (S_{11} and S_{22}) when the two ports are not actually the same, for instance, if the measurement setup includes bias tees, splitters, etc. (Wartenberg 2002).

Another modern variation of the LRM is the extended line-reflect-match (LRM+) (Scholz *et al.* 2004). LRM+ defines the measurement system reference impedance for each port individually overcoming limitations of the LRRM calibration procedure. Therefore, asymmetrical setups can be accurately calibrated with LRM+. This also applies to calibration with test chips (customized calibration elements, embedded on the wafer), as LRM+ is the only calibration method that does not assume that load pairs have identical values. Additionally, LRM+ calibrates the system to the DUT, including parasitic influences from contact pads, which means pad de-embedding is no longer required (see **Figure 37**). The accuracy of LRM+ has been verified using independent standards from NIST. Over a frequency from DC to 110 GHz, LRM+ consistently outperformed all other WL calibration methods available (Doerner and Rumiantsev 2005, Rumiantsev *et al.* 2006).

1.17.5.2.5.(ii) NIST multiline TRL Research by the Microwave Metrology Group of the US NIST has advanced the quality of rigorous on-wafer VNA measurements. The focus of this work has been the development and application of a multiline TRL calibration algorithm using the full redundancy of the standards. While normal TRL selects line pairs for applicable frequency ranges, the NIST multiline TRL uses a continuously varying optimally weighted average of all the data (Marks 1991, Williams *et al.* 1991). **Figure 35** compares the quality of the most-used on-wafer calibration techniques up to 110 GHz.

1.17.5.2.5.(iii) Measuring calibration standards The process of completing the calibration entails placement of probes on standards and directing the VNA measurement system to acquire data. This process is somewhat error prone as it is not always easy to determine if the standard is properly connected.

While uncalibrated or raw *S*-parameters from a VNA are often far from accurate, they can be used to help verify proper contact with the calibration standards. The raised probe open measurement does not have any contact issues and can be used as a reference. The measurement of a short standard will show ideally a 180° phase rotation of the reflection coefficient. This can be seen as a rotation of the raw data

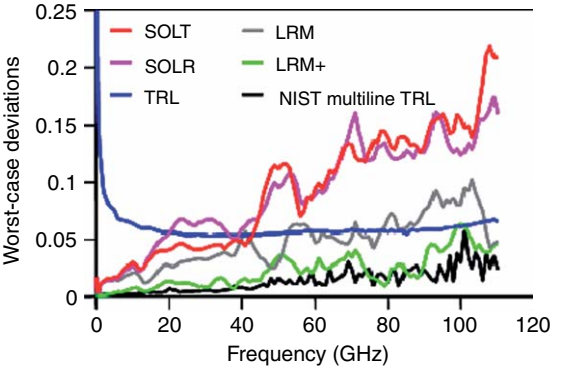


Figure 35 Comparison between the worst-case measurement deviations of a vector network analyzer (VNA) calibrated using different calibration techniques. (Figure courtesy: SUSS.)

Smith chart constellation or a shift in a phase display occurring when the probe makes contact. In practice, second-order terms in the error box will also somewhat impact the actual change of observed reflection, but the shift will be very noticeable and provide a good indication confirming visual identification of proper alignment and contact.

The load standard shows a significant reduction in the magnitude of the uncorrected reflection coefficient when contact is made. Contact with a 50-Ω load will tend to reduce the Smith chart display toward the center dot. Particularly at higher frequencies, the various reflections from adapters, cable connectors, and probes can degrade the overall return loss of the load standard. The measured reflections will normally be reduced by more than 10 dB when the load is in contact. Smaller reductions are signs of poor system return loss that may require correction.

The thru line measurement will exhibit the matching characteristics of the load and will also show an increase in the raw S_{21} transmission measurement.

1.17.5.2.5.(iv) Calibration verification It is a good practice to perform one or more verifications after completing every calibration. The verification of proper calibration kit entry and the correct measurements of the calibration standards can be demonstrated by measuring a long open circuit or a long transmission line. The reflection coefficient of the long open circuit should exhibit a smooth monotonic inward spiral on the Smith chart. The long transmission line verification standard should show a linear phase in the transmission response. Some common verification standards are shown in **Figure 36**.

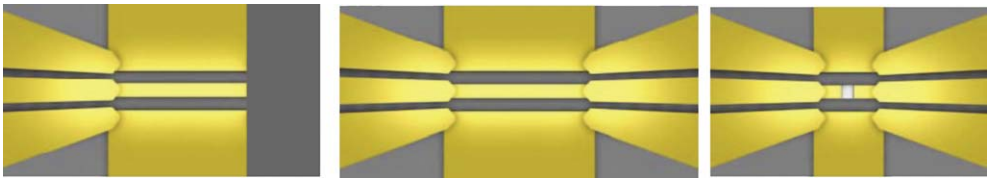


Figure 36 Typical on-wafer verification elements. In many cases measurement of such structures will catch the most common errors and problems. (Figure courtesy: SUSS.)

The remeasurement of a calibration standard is useless for verification, as only the repeatability of the measurement will be revealed. A VNA cannot recognize the errors caused by using incorrect impedance standards. A load resistor of the wrong value will not be recognized: it will be accepted as 50 Ω. The VNA will even accept interchanged open and short standards without complaint. Either mistake will result in significant measurement error.

Measurement of calibration standards for the advanced calibrations routines where standards do not have to be fully specified is useful for verification. Examining the open that is used for the reflect standard in a TRL or LRM calibration can identify problems. After calibration, the open may not have perfectly zero return loss. The probe in air will actually have slightly less loss than the probe in contact with the wafer, so in addition to some negative capacitance a small amount of return gain from the calibrated open may be observed.

The thru line standard is a useful verification element for an SOLR calibration. Again the verification should reflect expected behavior so this is only useful when the behavior is predictable. The SOLR can calibrate successfully even with a highly reactive and complicated frequency response of the reciprocal thru, which will not be predictable. However, a good line standard will exhibit attenuation with square root of frequency dependence consistent with thin-film transmission lines. Excess line length beyond the probe contacts will act like capacitive stubs resulting in increased attenuation at higher frequencies. Linear transmission phase would be expected unless the standard has dispersive characteristics.

A gold standard device – a device that has been characterized using a (hopefully) known good test system – is another way to validate a test system. However, it is difficult to validate a calibration and measurement method this way. Slight differences in calibration reference planes may appear as significant errors. Differences in circuit bias or power supplies can also appear as errors in the microwave measurement. A gold standard can, however, be an effective and simple test for qualifying a new system that uses identical equipment and methods as a reference measurement system.

1.17.5.2.5.(v) Ground–signal–ground vs. ground–signal pads After probe positioning and calibration errors, the most significant inaccuracy in microwave probing is associated with parasitic coupling at the probe tip (**Figure 37**).

A GSG interface terminates field lines effectively, but GS pads allow the open side fields to terminate on the wafer ground plane or chuck. The degree of this coupling is dependent on substrate thickness and pad spacing.

An undesirable mode of propagation where all topside conductors form a microstrip with the bottom of the wafer can occur in these structures. The best way to deal with this problem is usually to minimize the excitation of that mode. The GSG pad configuration excites the undesired microstrip mode less than the GS configuration. A reasonable rule is to avoid GS pads for good results above 10 GHz.

1.17.5.2.5.(vi) GS probing parasitics **Figure 38** depicts a GS probe tip, two connection pads, and a DUT, along with the main reactive parasitics of the connection. The ground is the backside of the wafer, the wafer chuck, or a high conductivity substrate such as silicon.

The transmission line uses the inductances in series with the DUT and the corresponding bridging capacitors. The resistor Z_{absorber} can be used to absorb energy in the undesired mode. This attenuates energy propagating up the ground, which could resonate with the probe holder.

Proper selection and use of a microwave probe will allow the parasitic C_{SProbe} and C_{GProbe} to be small and repeatable. However, differing pad sizes or conductor geometries on test structures can cause C_{GDut} or C_{SDut} to be different from one structure to another, or from the calibration standards. Only when these values are equal for calibration standards and measurement devices will these parasitics be removed by

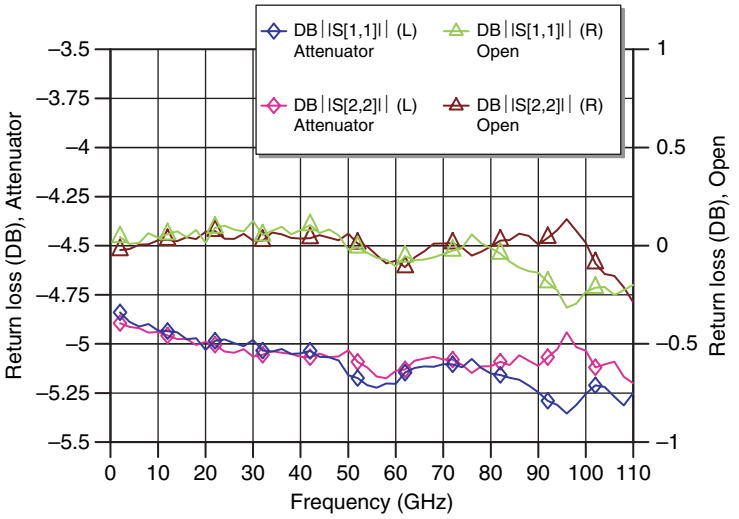


Figure 37 Verification example of the wafer-embedded extended Line-Reflect-Match (LRM+) calibration procedure using a symmetrical series attenuator and a symmetrical open elements on SiGe:C up to 110 GHz. (Source: Scholz R, Korndoerfer F, Senapati B, Rumiantsev A 2004 Advanced technique for broadband on-wafer RF device characterization. *63rd ARFTG Conference. Digest*, pp. 83–90, June 2004.)

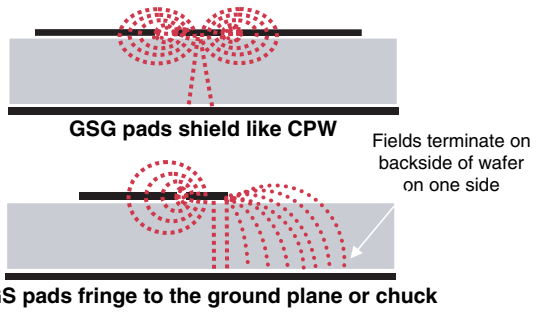


Figure 38 *Top*, ground–signal–ground pads shield electrical fields to the ground plate; *bottom*, ground–signal pads fringe to the ground plane or chuck.

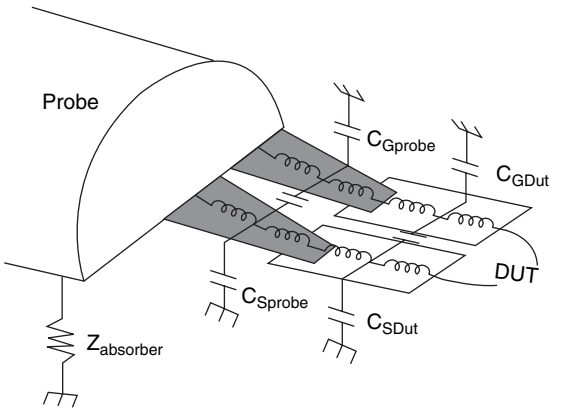


Figure 39 Ground–signal (GS) probing parasitic elements. C_{SDut} and C_{GDut} must be the same as the calibration standards. (Figure courtesy: Cascade Microtech.)

VNA calibration. But changes in metal configuration will often occur. For measurements above 10 GHz, the GS configuration is inadequate and the superior shielding of GSG is preferred.

1.17.5.2.5.(vii) Cable sensitivity Correcting the network analyzer measurement through a glossy cable is similar to the correction we subconsciously use to see an image through a dirty window. We correct for the attenuation due to dirt on the window as well as reflections from the window's surfaces. It is important that the cables be highly repeatable as well as reasonably low loss, so that there is enough dynamic range after looking through and correcting for the lossy cable. Calibrated measurements of a short-circuit standard before and after the cable has been significantly flexed, shown in **Figure 39**, demonstrates the importance of measurement repeatability. The resonance at about 20 GHz would significantly distort further DUT measurements. When troubleshooting the responses of a VNA, it is useful to keep in mind that impedances that vary rapidly with frequency cannot be due to errors near the DUT. When wafer probing there is insufficient electrical length on the wafer or in the probes to cause resonances or impedances that change rapidly with frequency.

1.17.5.2.5.(viii) Troubleshooting **Table 14** may be helpful in troubleshooting VNA measurement problems.

Table 14 Troubleshooting table for NWA measurements problems

Symptom	*Likely problem*	*Try*
Measurements do not repeat at any frequency	Cables or adapters are loose or faulty	Calibrate in coaxial with and without cables and adapters
	Analyzer fault	Calibrate in coaxial directly at test set port
	Probe contact fault	Verify probe contact visually or at DC
	Probe head fault	Calibrate in coaxial at probe connector; if good, replace probe head
	Oscillating DUT	Use different bias condition to verify; change low-frequency DUT terminations
Measurement noisy at all frequencies	Insufficient averaging or too large IF bandwidth	Increase averaging or decrease IF bandwidth
	Poor ramp sweep repeatability	Use step mode
Measurement noisy at certain frequencies	Sweeper band switching	Increase averaging, use step mode. Change sweep rate
	Loose connection	Tighten connectors
	Faulty probe resonating	Check open stub on ISS – should be a smooth inward spiral
Measurement noisy at high frequencies only	Poor return loss at high frequencies	Measure ISS load with corrections off – should be at least 10 dB below the open or short trace
DUT measurements noisy at low frequencies only, or Transistor S_{21} compresses at low frequencies (passive elements OK)	DUT is being biased by rectified RF, or RF is compressing DUT gm	Reduce RF power level, add device input port attenuation, use power slope
Passive element, reflection coefficient outside Smith chart	Incorrect calibration coefficients	Adjust calibration coefficients
	Poor probe placement on standards	Use alignment marks to improve placement accuracy

Source: Data assembled on the basis of information from Cascade Microtech.

1.17.6 Testing for Quality Assurance and Reliability

1.17.6.1 Reliability Testing

1.17.6.1.1 Definition of reliability parameters

Reliability $R(t)$ is typically defined as the probability that a device will still be functional after operation over the time interval $\{0, t\}$. In this context, the absolute time, the number of missions, the operating cycles, or even the mileage of a car can be taken as a measure of time. The complementary property of reliability is the failure probability $F(t) = 1 - R(t)$. Starting from a certain number of specimens N, the reliability is a function of the failed specimens n according to

$$R(t) = \frac{N-n}{N} \text{ or } F(t) = \frac{n}{N} \qquad [1]$$

Typical MEMS components cannot be repaired. Hence, the reliability of MEMS can only decrease over time. For electronic components frequently the mean time to failure (MTTF) is used to specify a lifetime parameter. MTTF is defined as the integral over the actual reliability function from zero to infinity.

$$\text{MTTF} = \mu = \int_0^\infty R(t)\mathrm{d}t \qquad [2]$$

In the case of a normal distribution, the MTTF is equivalent to 50% failed, i.e., MTTF $= t_{50}$. A certain portion of MEMS will have failed after some time. Therefore, specific lifetimes can be given for each failure probability, e.g., t_5, t_{10}, and t_{50} for failure portions of 5%, 10%, and 50%, respectively. For reliability engineering in most cases the exponential distribution as well as its more generalized form, the Weibull distribution, is used.

In the case of an exponential distribution, MTTF is taken at 63.2% failures, i.e., MTTF $\approx t_{63}$. The Weibull curve is characterized by a lifetime parameter, which is also taken as $T \approx t_{63}$ but is not identical to the MTTF.

For the failure rate, the definition in **Table 15** is valid. It can be regarded as the relative number of

Table 15 Distribution functions used in reliability engineering

Distribution function	*Failure probability* $F(t) = \int_0^t f(\tau)d\tau$	*Failure rate* $h(t) = \frac{f(t)}{1-F(t)}$
Exponential	$F(t) = 1 - e^{-\lambda t}$	$h(t) = \lambda = const.$
Weibull	$F(t) = 1 - e^{-\left(\frac{t}{T}\right)^b}$	$h(t) = \frac{b}{T} \cdot \left(\frac{t}{T}\right)^{b-1}$

failures per time in relation to the components that are still functional at that time. The failure rate can vary over time. In the case of an exponential distribution it is constant.

In the production phase of MEMS, engineers wish to achieve failure rates and failure portions that are as low as possible. Zero defects is a modern quality management strategy. On the contrary, sufficient failures are necessary in the development phase in order to monitor the reliability level and to give the development engineer optimization criteria. When the Weibull curve is verified experimentally, this is called end-of-life test.

For electronic components several definitions of failure are possible, which can be used specifically, which are as follows:

(1) Nonfunction of a component as well as open or short connections are hard failures.
(2) One of the most relevant failure criteria is deviation from specified functional parameters such as sensitivity, offset, and noise.
(3) Other criteria such as visual appearance, cracking, or a deteriorated surface finish can be taken as failures by definition.
(4) Criteria such as residual strength of a bond are useful in order to gain quality measures even when the MEMS device has not yet failed.

1.17.6.1.2 Real failure behavior of MEMS

Under operating conditions MEMS devices will not fail at a constant rate over their design lifetime. This behavior is represented graphically by the so-called bathtub curve (**Figure 40**). Over the life cycle three phases are distinguished.

In an early period, the failure rate is rather high, but it decreases over time. When early failures are analyzed, in many cases samples exhibit defects, which have led to the premature failures. The defectiveness is a result of inadequate fabrication processes. These processes generate a portion of devices that are functional but unreliable. In operation, these will fail first and the inferior portion of the produced lot will die out after some time.

According to the reliability theory, in the second phase the failure rate should be constant as failures here are only caused by chance when a local fluctuation of stress parameters exceeds the material's strength. The rate of these failures will increase

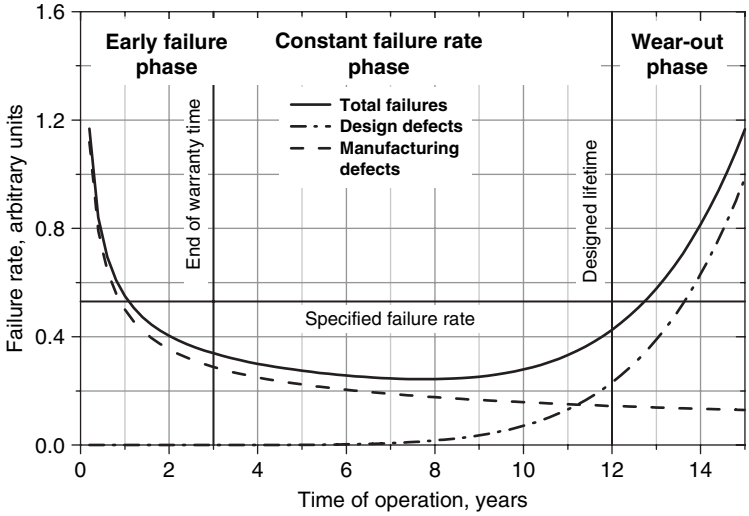


Figure 40 Schematic of the bathtub curve of failures of electronics hardware.

when the damaging conditions become more severe, e.g., when the temperature is increased.

In the last phase the failure rate will again increase due to wear-out processes. Wear-out reduces the ability of components to withstand the environmental loads. Therefore, the probability of failures is increased. Wear-out in MEMS is caused by phenomena like fatigue, corrosion, aging, or electrical degradation of dielectric layers.

The characteristic reliability behavior in the different phases of the bathtub curve is compared in **Table 16**. Clearly, different test strategies are necessary in the reliability management of MEMS. One has to distinguish between the designed reliability, which is tested by end-of-life-tests, and the manufactured reliability, which must be verified by tests monitoring the production.

Table 17 lists typical reliability data, which have been gathered from the data sheets of renowned MEMS manufacturers. A wide range of FIT values is characteristic of MEMS components.

1.17.6.2 Reliability Test Methods

Failures of an electronic component will occur when there are external, environmental loads or when the device is stressed by internal conditions due to operation. Several investigators (Guettler 2004) came to the conclusion that temperature is the dominant influence on the reliability of electronics. In mobile systems many other factors such as vibration, humidity, and lack of cleanliness will significantly contribute to failure causes (**Figure 41**).

Furthermore, operation of MEMS will lead to an increasing number of failures due to degradation (wear-out) by the use conditions. The individual conditions making up the load profile are very much application specific. Also the mechanisms can interact, making the problem rather complex. Loads that are most relevant to failures of MEMS are as follows:

- Temperature
- Temperature changes, including shock
- Mechanical shock and vibrations

Table 16 Characteristics of failures over the lifetime of MEMS

Phase	*Failure rate, tendency*	*Portion of specimens*	*Failure cause*	*Examples*	*Weibull exponent*
Early failures	Medium, decreasing	Defective ones, die out	Manufacturing defects	Defective solder Joints	$b < 1$; e.g., $b = 0.5$
Constant failure rate	Low, constant	Small, but possibly all	Random failures by fluctuations	Dielectric breakdown	$b = 1$
Wear-out failures	High, increasing	Aged ones	Design lifetime exceeded	Fatigue of interconnections	$b > 1$; e.g., $b = 2-10$

Table 17 Characteristics of failure rates of MEMS

Microsystem sensor	*Technology*	*Computed failure rate in FIT*[a]	*Confidence level, $1-\alpha$(%)*	*Temperature for failure rate (°C)*	*Use conditions*
Pressure	Sensor IC	18	60	25	Industrial
		44	90		
Pressure	Bulk micromachined	120	90	55	Industrial
		980		85	
Acceleration	Surface micromachined sensor IC	<1	60	55	Automotive, under hood
			90	125	
		1	60		
		33	90		
		49			
Magnetic sensor	CMOS Hall IC	2	60	65	Automotive, under hood

[a]Typical failure number was 0 to 10.
Typical test temperature is 125°C, used activation energy is 0.7 eV. Data were taken from several sources of manufacturers.

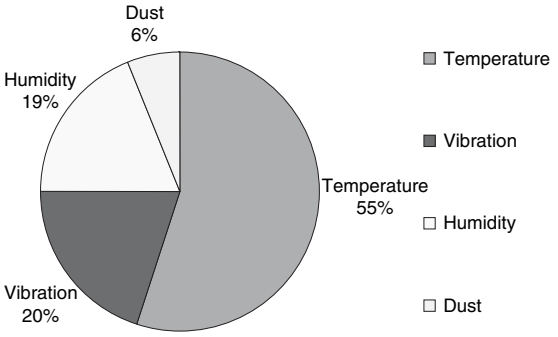


Figure 41 Environmental conditions as failure causes of electronic components. (Source: Guettler B 2004 (Original source: US Air Force Avionics Integrity Program).)

- Internal operating conditions (U, I, P)
- Environment, humidity, and climate
- Chemical agents

For verification, comparable loads must be imposed on the MEMS using standardized tests. Each of the tests will stimulate specific degradation processes and failure modes in the devices. Thus, many different laboratory tests are necessary to verify the expected operational reliability. In **Table 18**, a matrix summarizes the influence of these tests on specific failure modes of MEMS.

Testing must fulfill a number of prerequisites in order to provide the engineer with realistic results:

- The test conditions must be similar to operational conditions.
- All of the relevant failure modes must be stimulated.
- Test times must be sufficiently long.
- The data must exhibit sufficient statistical significance.

Some of the most relevant test standards for electronic components have been released by the Electronics Industry Association (EIA) as well as by the JEDEC as listed in **Table 19**.

A typical test procedure comprises the following steps:

(1) Visual inspection
(2) MEMS-specific functional test with parameter extraction
(3) Reliability testing
(4) Functional testing
(5) Decision whether the qualification goal is achieved

It is important that the degradation of functional parameters be monitored. Also other failure criteria can be applied in accordance with the predefined specifications. For reliability tests, the number of samples to be examined and the tolerable number of failures are specified according to EIA/JESD STD 47. Typically, three lots, each consisting of 77 components, must be used for temperature/environmental tests and three lots, each with 39 components, for mechanical tests. For production lots no failures are allowed. These tests numbers are applicable only for qualification of production lots but not in the development phase. There a sufficiently representative number of failures is desirable in order to determine lifetime parameters.

Successful lot qualification testing alone will not verify that the MEMS are reliable. The reliability results acquired during the technology qualification of the MEMS must be taken into account. The mathematical procedure is described in EIA/JESD STD 47. When a production lot of MEMS is qualified, N_{lot} is the lot size and M_{lot} is the number of lots to be tested, e.g., $N_{\text{lot}} \times N_{\text{lot}} = 77 \times 3 = 231$. No failures are allowed, so $c \overset{!}{=} 0$. The qualification is based on the data basis of the generic technology qualification. Due to statistical reasons, the complete database that has been gathered for a specific technology must be used even though it comprises many individual device types. N_{tq} is the generic lot size and $c = n$ is the number of failures in technology qualification:

$$\frac{\chi^2_{(c=n)}}{N_{\text{TQ}}} \overset{!}{\leq} \frac{\chi^2_{(c=0)}}{M_{\text{lot}} \times N_{\text{lot}}} \text{ or } \chi^2_{(c=n)} \overset{!}{\leq} \frac{N_{\text{TQ}} \times \chi^2_{(c=0)}}{M_{\text{lot}} \times N_{\text{lot}}} \quad [3]$$

The χ^2-value of the technology qualification must be smaller than the test value computed from Section 1.17.3.1 in order to clear the sample for production. The necessary degree of freedom ν is computed according to $\nu = 2 \cdot (c + 1)$. A second parameter is the confidence level $1 - \alpha$. α is the significance level. Typical confidence levels in reliability physics are $1 - \alpha = 90\%$ and $1 - \alpha = 60\%$. χ^2-values are taken from statistical tables or are computed using a spreadsheet program.

1.17.6.2.1 Field reliability

One of the most important questions in reliability engineering is the lifetime of the products in the application environment. After a certain time of operation in the field, some of the products will fail. Reliability is the probability of function after

Table 18 Test methods suitable for the stimulation of failures on MEMS components

Resulting failure	*Interconnections*	*MEMS attachment*	*Package defects*	*Leakage, hermeticity*	*External leads*	*Corrosion, metal*	*Chip metallization*	*High-aspect structures*	*Chip cracking*	*Electrical parameters*
Test method										
Temperature	X	X	X	–	–	–	–	–	X	X
Temperature cycles	X	X	X	X	–	–	–	–	X	X
Temperature shock	X	X	X	X	–	–	–	–	X	X
Mechanical shock	X	–	X	X	X	–	–	X	X	–
Vibration	X	–	–	X	X	–	–	X	X	–
Power cycling	X	X	X	X	–	–	X	–	X	X
Moisture and climate	X	–	–	–	X	X	X	–	–	X
Chemical media	X	–	–	–	X	X	X	–	–	X
Soldering cycle	X	X	X	X	–	–	–	–	–	X

Table 19 Environmental test methods for reliability testing of electronic components summarized from JEDEC/EIA Standards – these are also recommended for MEMS

Organization	*Number*	*Title*	*Purpose*
JESD	STD 26-A	General specification for plastic encapsulated microcircuits for use in rugged applications	General aspects
EIA/JESD	STD 47-D	Stress test-driven qualification of integrated circuits	General aspects
JESD	22-B101-A	External visual	Define optical criteria
JESD	22-A113-E	Preconditioning of surface mount devices before reliability testing	Define starting conditions
EIA/JESD	22-B116	Wire bond shear test method	Mechanical tests, strength
JESD	22-B104-C	Mechanical shock	Mechanical tests, strength
JESD	22-B103-B	Vibration, variable frequency	Mechanical tests, strength
JESD	22-A102-C	Accelerated moisture resistance – unbiased autoclave	Reliability in climate
JESD	22-A107-B	Salt atmosphere	Reliability in climate
JESD	22-A110-B	Highly accelerated temperature and humidity stress test	Electrical/climatic reliability
JESD	22-A100-B	Cycled temperature humidity bias life test	Electrical/climatic reliability
EIA/JESD	22-A101-B	Steady-state temperature humidity bias life test	Electrical/climatic reliability
JESD	22-A108-C	Bias life	Electrical/climatic reliability
JESD	22-C100-A	High-temperature continuity	Temperature
JESD	22-A103-C	High-temperature storage life	Temperature
JESD	22-A104-A	Temperature cycling	Temperature cycle
EIA/JESD	22-A105-B	Power and temperature cycling	Power cycle reliability
JESD	22-A106-B	Thermal shock	Temperature cycle reliability
JEDEC	22-A109-A	Hermeticity	Leakage at low rate
EIA/JESD	22-B106-C	Resistance to soldering temperature for through hole-mounted devices	Soldering cycle
EIA/JESD	22-B102-D	Solderability test method	Soldering tests

some time. The Weibull distribution in **Table 15** describes failure probability and reliability using the parameters b and T. These parameters must be made available by testing; n is the number of failures. Based on data from the qualification, a prediction for field application is possible in accordance with ISO 16750. The significance level α represents the probability of error, i.e., of failure despite a no-failure prediction. Typically, a minimum lifetime in the field t_{field} must be guaranteed and is therefore a specified value. Testing N devices over a time t_{test} without failure, a lifetime ratio $L_v = t_{\text{field}}/t_{\text{test}}$ can be defined. The reliability is then calculated according to ISO 16750:

$$R(t_{\text{field}}) \geq \alpha^{\frac{1}{NL_v^b}} \quad [4]$$

Vice versa, it is possible to compute the necessary defect-free test time according to:

$$t_{\text{test}} \geq \frac{t_{\text{field}}}{f}\left[\frac{\ln\alpha}{N\ln R}\right]^{1/b} \quad [5]$$

In this formula, an empirical acceleration factor between field and test has to be used, according to $f = t_{\text{test}}/t_{\text{field}}$. This factor must be determined experimentally or it is computed based on physics-of-failure models such as the Arrhenius equation or the Coffin–Manson law. Time can be defined in terms of hours as well as operation cycles or even miles.

For assemblies and modules integrating microsystems, specific standards will be applied. In the field of environmental tests, IPC standards as well as European standards IEC 60068-2, Environmental Testing, and ISO 16750 Road Vehicles-Environmental Conditions are examples. Generally, their physical test conditions are similar to the JEDEC standards.

It must be emphasized that a universal test set of general validity cannot be defined. Rather it is the first step toward a correct and complete product specification to analyze the real operating conditions in the application and to design the test program on this basis. Nevertheless, the use of best-practice test sets for specific MEMS applications like medicine or automotives can be a good starting point.

1.17.6.2.2 Failure analysis

Reliability tests must be accompanied by various failure analysis methods in order to understand the basic physical failure mechanisms. Helpful tests are as follows:

- Electrical testing
- Test of mechanical strength
- Chemical analysis of reactions products, corrosion residues, and impurities
- Measurement of thermal resistance

The investigations must be performed in the micro- to nanorange concerning dimensions and concentrations. The most relevant methods are compared in Table 20 like:

- Microscopy with light or electron beam
- Spectroscopy with electrons or X-rays for microanalysis
- Ultrasonic microscopy
- X-ray microscopy and tomography
- Test chips

1.17.6.3 Electrostatic Discharge

Electrostatic discharge (ESD) testing of MEMS is very important as these devices integrate micromechanical structures, analog and digital circuits, thin films, and CMOS memories. In a dry environment, typical human operations like walking, sitting, or working on an insulated ground will lead to electrostatic voltages up to 35 kV; in a humid atmosphere, the voltages are reduced to a maximum of 1500 V. Discharge onto the pins of a component will lead to high transient voltages and currents, which can damage internal structures. Two principal types of test have been standardized for the ESD of electronic components. The human body model (HBM) simulates the ESD of a person while the machine model (MM) tests the ESD of an equipment. For both tests, sensibility classes have been defined (Tables 21 and 22).

The test equipment must fulfill certain requirements with respect to rise time, peak current, and resonance frequency in order to perform correct tests (Table 23).

The test procedure is specified in the following list:

(1) Verify the test signal
(2) Perform parameter test and functional tests at ambient temperature
(3) Test at elevated temperatures
(4) Test with positive and negative pulse of all pin combinations; lot size is three components
(5) Perform parametric and functional tests
(6) Proceed with next voltage level

When the electrical parameters are out of the specified range, this will be the ESD failure criterion.

1.17.6.4 Burn-in, Run-in, and Nondestructive Testing

The idea of burn-in and run-in is based on the bathtub curve. The early failure rate of MEMS devices is significantly higher than the steady-state failure rate. This is caused by a certain portion of devices, which are not perfect and will fail during operation earlier than flawless MEMS components. During burn-in, devices will be operated at high temperatures preferably under conditions comparable to those of the application. Parametric and functional tests must be performed during burn-in. The mathematical basis of burn-in is the Weibull curve. The failed portion is

$$F(t) = 1 - e^{-\left(\frac{t}{T}\right)^{b}} \textit{ and the failure rate is}$$

$$h(t) = \frac{b}{T} \uparrow \left(\frac{t}{T}\right)^{b-1}$$

After a certain burn-in test time, a significant portion of the defective devices will fail. The remaining samples that are shipped to the customer will have a higher apparent reliability than the original batch. Due to high equipment costs, 100% burn-in test of large lots of MEMS devices is often not possible. A second aspect to use burn-in is the use as an indicator of the actual reliability level. When the early failure rate rises significantly, this indicates that also long-term reliability problems are imminent.

Table 20 Methods for the failure analysis of MEMS devices

Failed element	*Interconnections*	*MEMS attachment*	*Package defects*	*Leakage, hermeticity*	*External leads*	*Corrosion metal*	*Chip metallization*	*High-aspect structures*	*Silicon cracking*	*Electrical parameters*
Test/Inspection method										
Visual inspection (microscope)	X	–	X	–	X	X	–	–	–	–
Metallography	X	X	X	–	–	–	–	–	X	–
SEM/EDX	X	X	X	–	X	X	X	X	X	–
FIB/SEM	X	–	–	–	–	X	X	X	X	X
X-ray transmission	X	X	X	X	X	–	–	–	–	–
Ultrasonic microscope	–	X	X	X	–	X	X	–	X	–
Resistance measurement	X	–	–	–	–	–	–	–	–	X
Functional test	X	X	X	X	–	X	–	–	–	X
Thermal resistance	–	X	X	X	X	–	X	–	–	X
Shape measurement	–	–	–	–	X	–	–	–	X	–
Destructive testing	X	X	X	X	–	X	–	–	–	–

Table 21 Typical ESD sensitivities of different device technologies (Advanced Packaging)

Number	*Component type*	*Range of ESD sensitivity (V)*
1	MOSFET	10–100
5	EPROM	100–500
7	SAW device	150–500
8	Operational amplifier	190–2500
9	CMOS	150–3000
12	Thin-film resistor	300–3000

Table 22 Classes of ESD sensitivity of electronic components with the human body model (EIA/JESD 22-114A) and the machine model (EIA/JESD 22-115A)

Human body model		*Machine model*	
Class	*Failure at voltage* U	*Class*	*Failure at voltage* U
1	<2000 V	A	<200 V
2	2000 V to 4000 V	B	200–400 V
3	>4000 V	C	>400 V

Table 23 Parameters of short-circuit current waveform in the ESD testing with the human body model and the machine model

Model	*Voltage*	I_{peak}	*Resonance frequency*
Machine	400 V	5.8–8.0 A	11–16 MHz
Model	*Voltage*	I_{peak}	*Rise time, short circuit*
Human body	4000 V	2.40–2.96 A	2–10 ns

A value of $b = 0.5$ is a representative datum for electronic components and MEMS. Hence, burn-in can be a very effective method to monitor the quality level or to eliminate a portion of the bad samples by testing.

An important third aspect of burn-in is the stability of microsensors, i.e., regarding their electrical parameters. After manufacturing, many of the material properties affecting the electrical behavior are not yet stabilized, such as the resistance of thin films, the stress of metal layers, or the curing degree of encapsulation polymers. Packaging processes can have a strong influence on long-term functionality and drift of microsystems (Fischer *et al.* 2005). In order to save money, fabrication processes are optimized with respect to a high throughput. Therefore, transient processes will not have the required time to accomplish a final status. Consequently, the first material changes will be rapid enough to accomplish the product, which can be handled within seconds or minutes. The end status of the process is only reached after hours or days. Therefore, parameters defining the accuracy such as offset, zero point, or linearity cannot attain their final values in a microsensor directly after fabrication. The burn-in of the system at high temperatures can finalize the kinetic processes and therefore produce stable system properties over the use period. As a device is operational during burn-in, the functional test can also be used to monitor the parameter drift.

Nevertheless, the burn-in test strategy is avoided if possible due to several drawbacks, which are listed in **Table 24**. The main applications lie in fields with small lot sizes and very high-reliability requirements like in aerospace or medical technologies.

In some cases microsystems will exhibit defects, which can impair reliability or functionality. Many defects are caused by inadequate fabrication operations such as assembly or packaging. Nondestructive testing (NDT) is a method of risk elimination of the final product by defect visualization. The three most important NDT methods are visual inspection, X-ray inspection, and scanning acoustic microscope. The main advantages of NDT are that 100% inspection is possible and that no specimens will be lost during testing. As X-ray and ultrasonic testing will also reveal hidden flaws, these are of great importance in MEMS technology.

1.17.7 Test Standards

No special family of standards exists, which was specifically defined for microsystems technology. In typical applications, MEMS are regarded as electronic components and will be subjected to the respective standards and test procedures. A number of international organizations have regulated the field of components.

The American JESD standards and the JEP documents define the properties and quality requirements as well as methodologies for testing and statistical procedures. A part of the standards and technical papers published by JEDEC is free for download. More standards for electronic components have

Table 24 Advantages and drawbacks of burn-in of MEMS

Advantages	*Disadvantages*
Effective in elimination of defective samples	Costly equipment
Shorter quality control loops than for field failures (a week rather than years)	As an end-of-line test still forms wide control loop (hours are desirable)
Good early reliability indicator	Established mainly for thermal phenomena
Improve long-term stability of electrical parameters	High effort for 100% test of large series
Monitor off-set and zero-point drift	Temperature-driven, not for all damaging phenomena

been published in cooperation with the Electronics Industry Alliance (EIA) and are partly available from JEDEC.

In the field of manufacturing technology for electronic assemblies, a comprehensive set of standards was published by the IPC. It has been established worldwide and is used almost independent of specific markets. IPC standards are primarily aimed at printed circuit board assemblies. They specify methods, procedures, tests, and product requirements from the materials level up to the module level. A subset of the IPC publications is free for electronic download.

A major application field of MEMS is electronics for cars. The large equipment suppliers have formed the Automotive Electronics Council (AEC) who worked out a set of standards, AEC-Q100 to AEC-Q200 on the basis of modified JEDEC/EIA standards. Many AEC standards have been made free to the public.

Furthermore, many environmental tests have been defined by ISO standards and by European EN standards. Typical reliability test methods and tests for electromagnetic compatibility (EMC) have been defined and are widely used.

1.17.8 Conclusions

In the field of MEMS, testing is indispensable at all stages of the product generation cycle. It is very important to use efficient and versatile test strategies in order to accomplish fast, economic and dependable verification results with respect to functionality and reliability. These requirements are some of the strong drivers towards wafer-level-testing, which allows verification during intermediate fabrication stages but which also requires very costly equipment. Other future test concepts like built-in self-test or self-calibration are very prospective for MEMS but will require strong research efforts.

Acknowledgment

For his fruitful discussions and qualified remarks, the authors want to thank Mr. Rumianntsev from SUSS Microtec, who took over the proof-reading of the text.

References

Beroulle V, Bertrand Y, Latorre L, Nouet P 2001 Test and testability of a monolithic MEMS for magnetic field sensing. *J. Electron. Testing* **17**, 439–50

Charlot B, Mir S, Parrain F, Courtois B 2001 Generation of electrically induced stimuli for MEMS self-test. *J. Electron. Testing* **17**, 459–70

Davidson A, Jones K, Strid E 1990 LRM and LRRM calibrations with automatic determination of load inductance. *36th ARFTG Conference Digest*, pp. 57–63

Doerner R, Rumiantsev A 2005 Verification of the wafer-level LRM+ calibration technique for GaAs applications up to 110 GHz. *65th ARFTG Conference Digest*, Long Beach, CA, USA

Dumas N, Azais F, Latorre N, Nouet P 2006 Electro-thermal stimuli for MEMS testing in FSBM technology. *J Electron. Testing* **22**, 189–98

Engen G F, Hoer C A 1979 Thru-reflect-line: an improved technique for calibrating the dual six port automatic network analyzer. *IEEE Trans. Microwave Theory Tech* **MTT-27**, 987–93

Ferrero A 1992 Two-port network analyzer calibration using an unknown thru. *IEEE Microwave Guided Wave Lett.* **2(12)**, 505–7

Fischer S, Beyer H, Janke R, Wilde J 2005 The influence of package-induced stresses on moulded hall sensors. *Microsyst. Technol.* **12(12)**, 69–74

Guettler B 2004 Wenn's der Elektronik zu heiss wird. *Automotive Electron. Syst.* **1–2**, 59 (Original source: US Air Force Avionics Integrity Program)

Marks R 1991 A multi-line method of network analyzer calibration. *IEEE Trans. MTT* **39(7)**, 1205–15

Mir S, Charlot B, Courtois B 2000 Extending fault-based testing to microelectromechanical systems. *J. Electron. Testing* **16**, 279–88

NN Calibration and Accuracy Factors – Summit High-Frequency Probe Station Reference Manual. Cascade Microtech, part number 103-375-A

Pence J 1993 Verification of LRRM calibrations with load inductance compensation for CPW measurements on GaAs substrates. *42nd ARFTG Conference Digest*, San Jose, CA, USA, pp. 81–6

Rufer L, Mir S, Simeu E, Domingues C 2005 On-chip pseudorandom testing. *J. Electron. Testing* **21**, 233–42

Rumiantsev A, Doerner R, Thies S 2006 Calibration standards verification procedure using the calibration comparison technique. *36th European Microwave Conference Digest*, Manchester, UK, September 2006

Rytting D 1986 Appendix to an analysis of vector measurement accuracy enhancement techniques. *RF and Microwave Symposium and Exhibition*, Hewlett-Packard Inc.

Scholz R, Korndoerfer F, Senapati B, Rumiantsev A 2004 Advanced technique for broadband on-wafer RF device characterization. *63rd ARFTG Conference Digest*, June 2004, pp. 83–90

Van Wijnen PJ *On the Characterization and Optimization of High-Speed Silicon Bipolar Transistors*. Cascade Microtech, Beaverton, OR

Walker D K, Williams D F 1997 Compensation for geometrical variations in coplanar waveguide probe-tip calibration. *IEEE Microwave Guided Wave Lett.* **7(4)**, 97–9

Wartenberg S A 2002 *RF Measurements of Die and Packages*. Artech House, Boston

Williams D, Pence J, Davidson A 1991 Comparison of on-wafer calibrations. *38th ARFTG Conference Digest* pp. 68–81 December 1

Wollitzer M, Thies S, Schott S 2001 New probing technology now enables impedance controlled on-wafer probing. *GAAS*, London, UK

Organizations and Companies

AEC Automotive Electronics Council. Q 100 – Stress-Test-Driven Qualification of Integrated Circuits. http://www.aecouncil.com/AECDocuments.html

Agilent Technologies, Inc. Headquarters, 5301 Stevens Creek, Blvd, Santa Clara, CA 95051, USA: http://www.agilent.com

Cascade Microtech, Inc., 2430 NW 206th Avenue, Beaverton, OR 97006, USA: www.cascademicrotech.com

EIA Electronic Industries Alliance, Arlington, VA, USA: http://www.eia.org

IPC, Association Connecting Electronic Industries (former Institute for Interconnecting and Packaging Electronic Circuits) 3000 Lakeside Drive, 309 S Bannockburn, IL 60015, USA: http://www.ipc.org/

JEDEC Solid State Technology Association, Arlington, VA, USA: http://www.jedec.org/download/default.cfm

MEMUNITY The MEMS Test Organisation: http://www.memunity.org/on-wafer_testing.htm

SUSS MicroTec AG, Schleissheimer Str. 90, 85748 Garching, Germany: http://www.suss.com

The Micromanipulator Co., Inc., 1555 Forrest Way, Carson City, NV 89706, US: www.micromanipulator.com

Standards

EIA/JESD STD 47 Stress-Test-Driven Qualification of Integrated Circuits

IEC 60068-2 Environmental Testing

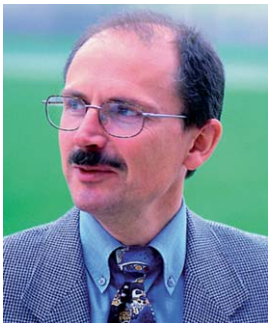
ISO 16750-1, Road vehicles – environmental conditions and testing for electrical and electronic equipment – Part 1: General. ISO, Geneva 2003

Biographies

Leonhard Reindl received his Dipl.-Phys. degree from the Technical University of Munich, Germany in 1985 and the Dr. sc. techn. degree from the University of Technology, Vienna, Austria, in 1997. From 1985 to 1999 he was a member of the microacoustics group of the Siemens Corporate Technology Department, Munich, Germany, where he was engaged in research and development on SAW convolvers, dispersive and tapped delay lines, ID-tags, and wireless passive SAW sensors. In the winter of 1998/99 and in the summer of 2000, he was guest professor for spread spectrum technologies and sensor techniques at the University of Linz, Austria. From 1999 to 2003 he was a university lecturer for communication and microwave techniques at the Institute of Electrical Information Technology, Clausthal University of Technology. In May 2003, he accepted a full professor position at the laboratory for electrical instrumentation at the Institute for Micro System Technology (IMTEK), Albert-Ludwigs-University of Freiburg.

His research interests include wireless sensor and identification systems, micro energy harvesting, surface and bulk acoustic wave devices and materials, as well as microwave communication sensor systems based on SAW devices. He holds 35 patents on SAW devices and wireless passive sensor systems and has authored or coauthored approximately 130 papers in this field.

Leonhard Reindl is an elected member of the Administrative Committee (AdCom) of the IEEE Ultrasonics, Ferroelectrics and Frequency Control Society. He is a member of the Technical Program Committee of the IEEE Frequency Control Symposium and the IEEE Sonics and Ultrasonics Symposium. He is also engaged in technical committees of the German VDE/ITG Association, and served as chairman for the German conference Sensors and Measurement Systems 2006.



Jürgen Wilde received a Diploma degree in Materials Science from the University of Erlangen-Nuremberg, Germany, in 1982. In 1989 he received a Ph.D. for his work on The Properties of Microstructured Bonding Wires and their Processing for Electronics from the Technical University of Clausthal, Germany. Subsequently, he worked as a research fellow and group leader in the DaimlerChrysler research institute in Frankfurt, Germany. His research was aimed at influencing design, materials, and fabrication processes on the performance and reliability of mechatronic systems. Since 1999 he has been holding the Chair for Assembly and Packaging at the University of Freiburg's Department of Microsystems Engineering in Germany. He teaches graduates and undergraduates giving lectures on Assembly and Packaging, Quality Management, Testing and Qualification, Systems Integration, and Reliability. He is a member of the VDI/VDE Society GMM, where he is chairman of the Board 5.5 Assembly and Interconnection Technology. His publication list comprises approximately 80 papers in the field of assembly, packaging, and materials science for electronic devices like sensors and microsystems. His current research is dedicated to new methods for accuracy prognosis, reliability verification, and testing of microsensors in automotive applications.

Craig Stewart has worked in the fields of microwave and fiber optic devices, communications, semiconductors and semiconductor equipment. He qualified with a first class honors degree in Electrical Engineering from Glasgow University and

then undertook research in guided wave optics and acousto-optic interactions. In 1975 he moved to the Plessey Allen Clarke Research Centre (Caswell) where he spent seven years working on optical fiber measurements, passive optical devices and acousto-optic signal processing. There followed a transfer to the Plessey Component Group and management assignments in various group companies as well as a short period studying business methods in Japan. On returning, he joined the Plessey Three-Five Group in 1984, a company formed to address the emerging market for III–V semiconductors. Initially tasked with formulating a marketing strategy, he then assumed responsibility for engineering development to establish viable processes and product portfolio, finally moving once again into sales and marketing management for the company. During this period, he first became familiar with on-wafer measurements, both as a customer for Cascade Microtech products and as the sponsor for an internal Plessey development program on ceramic based wafer probes. In 1988 he left Plessey and spent a short period with a cable equipment company before joining Cascade Microtech in 1990 with responsibility for European sales and applications support. Initially the market need was confined mostly to RF measurements on III–V semiconductor devices and Cascade was able to establish the premier position in the field. As the need for on-wafer measurement tools expanded into the mainstream semiconductor industry the need for complementary precision DC measurements emerged with consequently increased sales, applications and service responsibilities. In October 2006, Craig left Cascade Microtech to pursue other business interests although he is still active in semiconductor and communications technology and management via Candim Consulting.

1.18 Multiphysics and Multiscale Simulation

J. G. Korvink, A. Greiner, and Z. Liu, Department of Microsystems Engineering-IMTEK, University of Freiburg, Freiburg, Germany

1.18.1 Multiphysics Simulation and Design Optimization

The microelectromechanical systems (MEMS) design process is highly iterative in character. This is because design constraints tend to become more precise as each design evolves, and because complexity issues emerge only as a design takes shape. Computational tools aim at alleviating this aspect by providing two key capabilities:

- A high level of predictability of the physical processes in a given design realization
- A method to modify a design in the direction of optimality

For the second capability, the first is of course a prerequisite. In this chapter we describe some recent techniques that contribute to more powerful tools both for multiscale multiphysics simulations and toward the ultimate goal of automatic design optimization.

We have grouped the continuum approaches together and described them first, and have followed this up with the particle simulation approaches for soft matter simulations.

As dimensions reduce and soft matter becomes more relevant to MEMS, we see a gradual increase in the use of particle-based methods, and their emergence in commercial tools.

1.18.1.1 Multiphysical Essence of MEMS Simulation

Modeling denotes the process of developing an appropriate mathematical description of a particular system. Simulation is the process of using this description to actually compute the behavior of the system analytically or numerically.

Originally, silicon-based sensors and actuators constitute the basic building blocks of MEMS, and integrated MEMS employ materials and processing techniques originally designated for integrated circuits. Despite the inherent reference to electrical and mechanical phenomena, the field of MEMS has grown to encompass a much broader application range, including microsystems exploiting electrical, mechanical, thermal, magnetic, fluidic, selected chemical, and biological effects and possible coupling among these fields (Romanowicz 1998, Senturia 2000). In addition, a much wider range of material technologies has become established, including metals, polymers, and ceramics.

1.18.1.1.1 Classification of coupled physical fields in MEMS

In MEMS, certain measurements and outputs are detected in the bulk of MEMS devices, and others only on their surface. The same is true of coupled physical phenomena. For example, sensor effects are modeled by equations that are either coupled through domain effects, where a number of physical models are active within the same region and influence each other, or through common boundaries for the case where uncoupled models operate in nonoverlapping adjoining regions. This is relevant to a large number of physical models. We seek to clarify the basic operating principles of MEMS and to characterize the influence of coupling effects in MEMS devices. This classification is not the only possible one but provides an overview of the coupling effects in MEMS.

1.18.1.1.2 Domain-type multiphysical coupling

The electro-thermo-mechanical interaction is a commonly used coupling effect in MEMS. This coupling effect is also called domain coupling because the physical phenomena of coupling occurs completely in the material domain. For example, the energy of an electrical current passes through a material and is partially dissipated into heat. At the same time, the gradient of temperature causes thermal stress and results in structural deformation. This is a natural sequential coupling effect, which can be observed in daily life. In fact, there are fully coupled effects between electrical, thermal, and mechanical signals due to the nature of associated materials. If the independent variables, such as stress, electrical field, and temperature, and the dependent variables, such as strain, electrical displacement, and entropy, are chosen, then the coupling effects can occur between strain–stress as the elastic compliance, strain–electric field as the inverse piezoelectric effect, strain–temperature as thermal expansion, electrical displacement–stress as piezoelectric effect, electrical displacement–electrical field as permittivity effect, electrical displacement–temperature as pyroelectric effect, entropy–stress as piezocaloric effect, entropy–electrical field as electrocaloric effect, and entropy–temperature as heat capacity effect (Nathan and Baltes 1999).

On the one hand, it is possible to observe how the independent variables are coupled with each other, based on these coupling effects. On the other hand, a certain coupling may have very small amplitude when compared with other couplings. For example, the value of elastic coefficients may change only in a very small interval (<1%) when retrieved under constant entropy or isothermal conditions. In this case, the coupling effect can be decoupled and the accuracy of the numerical solution will still agree well with measurements.

1.18.1.1.3 Boundary-type multiphysical coupling

The electrostatic microactuation of a parallel plate structure is a widely used technique in MEMS. One electrode is typically a movable beam while the other electrode is fixed and separated by an air gap, which is within the micrometer range. This electromechanical interaction is a typical boundary coupling effect since the electrically induced mechanical loading is significant with regard to the electrical and mechanical interface. It is clear from the Poisson equation that the electrical force is perpendicular to the material boundary. Then the mechanical deformation of the structure alters the distribution of the electrical potential, thus changing the distribution of the surface charge density and hence the electrostatic forces acting on the structure. The self-consistent solution for both the mechanical deformation and the electrical potential distribution is coupled through the

interface of two energy subdomains. It is especially tricky to compute self-consistency when the pull-in in electrostatic actuation occurs.

1.18.1.2 Coupled FEM Simulation

The primary goal of using numerical simulation for MEMS development is to enable and to accelerate the design compromise. It is also an important aid in improving the understanding of some physical processes, given that the model is correct. Of course, simulation is not a replacement for experimental evidence, and it cannot eliminate a hardware development cycle. It also does not make sense to pursue modeling without measured material property data. This necessity cannot be overemphasized.

1.18.1.2.1 PDE expression of physical field

Physical effects can be described by partial differential equations (PDEs). A PDE relates an unknown scalar or a vector-valued function $u[x_i]$ to its partial derivatives for space and time:

$$F\left[x_1, x_2, \ldots, x_n, u, \frac{\partial u}{\partial x_1}, \ldots, \frac{\partial^k u}{\partial x_1^{k_1} \cdots \partial x_n^{k_n}}, \ldots\right] = 0 \quad [1]$$

The order of a PDE is given by the derivative of the highest order k present in eqn [1], and it is referred to as quasi-linear when all the highest order derivatives appear in a linear relationship. For most of the commonly used second-order PDEs, the characteristic at a point x_i is determined by the coefficients of the second derivative term. We speak of elliptic equations when all eigenvalues of the coefficient matrix have the same sign, of hyperbolic equations when the eigenvalues have differing signs, and of parabolic equations when the matrix is singular. Only by appropriately specified boundary conditions, such as Dirichlet, Neumann, and Robin boundary conditions, a unique solution of the PDE is defined (Morton and Mayers 1994).

1.18.1.2.2 Numerical simulation of single PDE using FEM

In most cases, analytical solutions to the partial differential questions are possible only for a linear PDE with certain simple geometries and boundary conditions. For arbitrarily shaped geometries there can be no closed-form solutions. Then a PDE has to be solved numerically using suitable approximations. In this chapter we use the standard weighted Galerkin finite element method (FEM) to discretize a PDE and obtain approximate solutions.

In the FEM, the choice of weighting and approximation functions in the discretized weak form of the corresponding PDE has an important influence on the solution. In general, the FEM procedure includes deriving a PDE weak form, choosing an approximation function, meshing the computational domain, assembling the stiffness matrix and load vector, adding boundary conditions, and solving a linear algebraic equation. For more detailed discussions about the accuracy and the efficiency of FEM solutions, see Zienkiewicz *et al.* (2005). Here we merely discuss some key points that are very important to obtain reasonable numerical solutions in various MEMS field simulations.

Mechanical analysis involves solving the Lamé equation with the given displacement and force boundary conditions. In deformation analysis, it is usual to choose a linear finite element to discretize the computation domain in order to balance the computational cost and accuracy. This is especially important for 3D simulations. If only the numerical solution of the displacement is needed, then the linear finite element is sufficient in most cases, whereas this is not true anymore if the strain or stress distribution needs to be checked. For a 2D beam that is rigidly fixed at one end and is loaded by a force at the other end, the numerical solution of stress with linear finite element shows a major difference (generally >15%) in comparison to the analytical solution. The discrepancy cannot be compensated using more refined meshes. This phenomenon is called self-locking in beam analysis. The reason for this excessively stiff behavior can be attributed to the missing approximate basis function describing the bending terms using a linear finite element. Similar problems can also be met when solving thin plate and shell problems that have large rotations and small strain. One of the simplest remedies is to use a high-order finite element. However, the computational cost of the high-order finite element is much more than that of a linear finite element. An efficient method to obtain an accurate stress solution is to use a nonconforming element. The nonconforming quadratic complement element approximates the solution by the standard linear interpolation function and a nonconforming quadratic function. The nonconforming degrees of freedom (dofs) are condensed at the element level so that dofs of the element stiffness matrix are the same as for the linear element after static condensing. The nonconforming

element improves the accuracy of stress distribution significantly, and there is no significant increase in the computational cost when compared with the linear element.

Electrostatic analysis involves solving the Poisson equation to compute the distribution of the electrostatic potential and electric field in a given domain. The accurate calculation of electrostatic fields is very important in order to simulate microdevice behavior, such as capacitive sensors and electrostatic microactuators. Generally, the Lagrange element, which has potential dofs only on the element nodes, is used. The electric field is calculated through the numerical solution of the potential. In this case, the accuracy of electric field is not as good as the potential. When the solution of the Poisson equation has a singular character caused by a nonsmooth boundary or an inhomogeneous material distribution, the accuracy of the electric field is not good. Hermite-type elements can be used in which both the potential and the electric field are dofs. However, the Hermite type of high-order element is sometimes awkward to use when solving the electrostatic problem, which includes material interfaces. On the nonsmooth points, the distribution of the potential is continuous but the electric field is not continuous. So the numerical solution with standard Hermite-type element is oversmoothed compared with the exact solution. In order to find a remedy for the oversmooth problem of Hermite-type high-order elements, releasing of derivative dof is required. It is well known that on the material interface the potential and electric displacement is continuous, but the electric field is not continuous. The jump of the electric field depends on the ratio of material dielectric coefficients. So only one unknown of electric on the interface should be used when using the standard discretization method. One simple remedy is to rescale the element stiffness matrix before assembling into the global stiffness matrix, by the relationship of the electric displacement. The transformation is completed on the element level and the sum of dofs does not increase when compared with the standard method.

Like the electrostatic problem, there is also a material interface problem when solving the electromagnetic problem based on the Maxwell equations. Based on a standard Lagrange element, the electric and magnetic fields and fluxes are determined indirectly by solving a scalar and vector potential. At the same time, the curl of electric field and the divergence of magnetic flux should be zero. It is well known that the tangential components of electric field and magnetic field are continuous across the material interface, and the normal components of magnetic flux and electric displacement are continuous across the material interface. The difficulty in the modeling of field strengths caused by discontinuities in the properties of the media can be overcome by using edge elements. An edge element has the property of ensuring the continuity of tangential field components across an interface between different media, while leaving the normal field components free to jump across such interfaces. Hence, edge elements can be used for modeling an electromagnetic field along the interfaces between two different media. It can provide sufficient continuity for the electromagnetic field with a conforming-type element.

In acoustics, sound pressure waves are often sinusoidal. Associated with a sound pressure wave is a flow of energy. The acoustic wave is extremely sensitive to changes in material properties, such as mass, elasticity, and viscosity of the propagation medium. Similar numerical discretization ideas as used for the electrostatic and electromagnetic problem can be used to deal with the material problem. Meanwhile, there is another problem regarding acoustic simulation with the standard Lagrange element, which uses polynomials as approximate basis. Since the basis solution of the acoustic problem can be expressed using superimposed sine functions, fine enough number of finite elements should be used to interpolate the high-frequency wave solution. Coarse meshes cannot reflect a sound wave. The basic reason for this problem is the limited approximation accuracy of the polynomial function for the high-frequency waves. Bubble elements have shape functions, which are zero at the boundaries of the mesh element, and have a sine or a cosine function in the middle of the mesh element. Combined with a linear basis, the bases used in the bubble elements are more suitable to simulate wave phenomena with coarse meshes when compared with the standard Lagrange element.

Fluid problems are also a challenge for the FEM. The challenges arise from two aspects. The first is the consistent interpolation of the velocity and the pressure, and the second is the stable discretization of a pure convection operator. For the first problem, a general rule to obtain a consistent discretization is that the number of pressure dofs should never exceed the number of velocity dofs. When solving incompressible Navier–Stokes equations (NSEs), this demand should be valid independently of the number of elements used to discretize the computational

domain. In order to satisfy this criterion, a generally accepted rule is that the order of approximation of the pressure must be one lower than the order of approximation of the velocity. So if the velocity is approximated by a linear polynomial, the pressure is approximated by a constant in each element and so on. For the second problem, nonstandard Galerkin FEM, such as streamline upwind Petrov Galerkin (SUPG) method, can be used to obtain a solution without nonphysical oscillations.

1.18.1.2.3 Numerical simulation of coupled PDEs using FEM

Until now we have merely discussed the numerical simulation of isolated physical effects that are described by a single PDE, but neglecting all coupling effects with other state variables. However, MEMS devices achieve their functionality through coupling of different physical effects. In the following, we introduce coupling simulation based on boundary and domain coupling effects. The treatment is far from being complete but intends to illustrate some possibilities to accurately simulate coupling mechanisms in MEMS.

1.18.1.2.4 Sequentially coupled method

The sequentially coupled method is typically used in boundary-coupled problems. It is also called partitioned analysis of coupled problems (Felippa 1999), and it was first used to solve the fluid–structure interaction problem. The numerical solution of the structural deformation and fluidic field is implemented separately, and the coupling effect is transmitted through the fluid–structure interface. This means that different programs can be called as black boxes separately, and the finite element meshes need not coincide at the interfaces (http://www.ansys.com/). The decomposition is driven by physical or computational considerations for sequentially coupled methods. Another case using the sequentially coupled method is the loosely coupled problem. For the electrothermal problem, a loosely coupled effect exists if the electric current is not influenced by the change in the temperature. The electric current is first calculated and then solved for the temperature distribution. The advantage of using several programs lies in exploiting the know-how contained in each program, enabling research effort to concentrate on the development of coupling mechanisms and solution strategies, especially since there are many commercially available general-purpose modeling tools, which are extensively used for MEMS simulation. However, the coupling between the different programs has to be performed via external files. This generates a dependency on the data formats of the different programs. The applicable solution strategies for coupled simulations are limited to decoupled iteration schemes, because the coupling Jacobian matrix is not available. At the same time, the sequentially coupled method is not straightforward when extended to more than two fields, because no general staggering iteration can guarantee the convergence for a variety of physical coupling characteristics in MEMS.

1.18.1.2.5 Fully coupled method

In practical MEMS simulation, many coupled problems are fully coupled. It is natural that one discretizes the coupled PDEs directly. In fully coupled discretization, a unified mesh is used in the whole computational domain and each nodal point has the same dofs, which includes unknowns from all the coupled PDEs. With the standard FEM, the off-diagonal submatrices of the global discretization matrices describe the coupling effect and associated relation (http://www.femlab.com).

1.18.1.2.6 Linearization of nonlinear PDEs

For most of the coupled problems in MEMS, the nonlinearity comes from coefficients or material properties that contain a dependent variable or from certain terms in PDEs that have a dependent variable of degree 2 or higher order. Nonlinear PDE sets should contain at least one nonlinear PDE.

In order to solve a system of nonlinear equations, an iterative procedure is necessary. In general a suitable initial estimation is given, and then linearized nonlinear equations are solved based on a previous solution until a converged solution is obtained. Generally there are two ways to derive the iterative method. First, one can discretize the nonlinear PDEs and then solve the nonlinear equation system. An alternative is to linearize the nonlinear PDEs and then to discretize the resulting linear equations (Belytschko *et al.* 2000, Gockenbach 2002). Sometimes both approaches are identical. The second approach is conceptually easier than the first one, and therefore it is widely used in MEMS simulation.

When linearizing the nonlinear PDEs, the nonlinear terms are treated using the Taylor expansion, leaving only the linear terms. This is called the Newton method. Alternative linearization is constructed by the so-called Picard iteration method.

The nonlinear terms are linearized in which all the nonlinear unknown terms are replaced with an iterative solution from the last time step. An important question with respect to these two iterative methods is how to give a good initial estimate. It is well known that the Newton method converges fast (quadratically) when the initial value is in the neighborhood of the converged solution. However, if the distance between the initial value and the converged solution is too large, the Newton method may converge slowly or may even diverge. The Picard iteration seems to have a larger convergence region, which means that it does not need as accurate an initial estimate. However, Picard iterations converge much slower than the Newton method (linearly). A possible strategy to balance the stable and fast convergence is to use Picard iterations first with an initial value and then use Newton iterations when the iteration residual is smaller than certain value.

In strong nonlinear problems, the convergence of nonlinear iterations may be very sensitive to the quality of the initial value even though one is using combined Picard and Newton linearization. In this case, the robustness of the iterative property can be improved using the continuation method (Allgower and Georg 2003). The continuation method defines an easy problem and the original problem that one actually wishes to solve. The solution of the easy problem is gradually transformed to the solution of the hard problem by tracing this path. The path can be defined by introducing an additional scalar parameter that can weaken the nonlinearity of the original problem.

1.18.1.2.7 Time-dependent solvers

The discretization of the time-dependent PDE results in a system of ordinary differential equations (ODEs). In order to solve this system of equations any classical method for solving ODEs can be used. In general one can distinguish between explicit and implicit methods, and between one-step and multistep methods. In the FEM, the time derivative can be replaced by a forward difference discretization (or method of lines):

$$M\left(\frac{U_{k+1}-U_k}{\Delta t}\right)+KU=F \qquad [2]$$

where k denotes the present time-level and $k+1$ the next time-level. A method is called explicit if the term KU is taken only at the time level k. When KU is also taken at the new time level $k+1$, the method is called implicit.

In the explicit method, one still has to solve a system of equations if the matrix M is not lumped. In the case of a lumped M matrix, the solution implies only the inversion of a diagonal matrix and the explicit method is relatively cheap. A disadvantage of the explicit method is that the time-step length must be restricted based on the Courant–Friedrichs–Levy (CFL) condition in order to obtain a stable solution. For a stiff ODE, the size of the time step in the explicit method can be so small that the overall computation is not efficient anymore. For the implicit method, it may be unconditionally stable for well-defined problems. The accuracy requirement is the main reason to restrict the size of the time step.

In coupled problems, which include both time-dependent PDEs and steady PDEs, the matrix M becomes singular. Differential algebraic equations (DAE) have to be solved after the space discretization of the coupled PDEs. This is also the case where there is a big difference in magnitude of the solutions of the coupled time-dependent problems so that the matrix M is almost singular. One has to note that DAEs are different from ODEs, although a general DAE can be transformed into an ODE theoretically by variable substitution. The variable-order, variable-step size backward differentiation formula (BDF) is an efficient method to solve a DAE problem. Solving DAEs is also an important step in the coupled optimization method (COM) for fully coupled physical fields discussed later on in this text (Ascher and Petzold 1998).

1.18.1.3 Mesh Adaptation

In MEMS, an efficient coupled field simulation procedure depends on many detailed numerical strategies. In the FEM, a good mesh is a very important issue, which influences the efficiency and the accuracy of the numerical solution (Thompson *et al.* 1985). When users have limited experience in the properties of the coupled physical fields, it is nearly impossible to decide upon a suitable mesh to discretize the computational domain in advance. In this case, mesh adaptation can bridge the gap. Basically, dynamic mesh adaption can be implemented in two general ways, i.e., adaptive mesh refinement (AMR) (Costa and Alves 2003) and adaptive moving mesh (AMM) (also called the r-adaptive method) (Cao *et al.* 2003). The mesh is locally remeshed in the AMR method, thereby capturing the important part of the

solution. This method is usually accompanied by a relatively complicated data structure updating procedure. For the AMM method, numerical solution of the moving mesh is based on a moving mesh PDE. The nodal points are moved throughout the region to best approximate the solution globally.

The AMR method is suitable for domain-type coupled problems in MEMS. When using the FEM to solve a PDE, there are two major AMR methods available, the h-method and the p-method. For the h-method, the mesh is refined or coarsened by adding or deleting mesh grids according to a local error analysis. For the p-method, the mesh is fixed and the order of a finite element is varied to implement a high-order approximation locally. There have been extensive studies on the h-method and p-method, both of which have shown good efficiency for both steady-state and time-dependent problems. The key point of the AMR method is to define a remeshing strategy. This guarantees the quality of the finite element solution and smoothens the expressions of the material domain, while keeping a reasonable number of design variables. Generally, an error estimator is used to determine the adaptation strategy, since the exact solution of the direct problem is unknown. For a single physical field problem, such as a linear elastic mechanical problem, the error estimator by Zienkiewicz and Zhu can be expressed as a least square approximation between the approximated nodal stress and the smoothened stress in the Gaussian points for all finite elements. Since the solution range for different physical fields can have difference in magnitude of several orders, the error estimators for coupled problem cannot be based on the numerical solution directly. One strategy is scaling the numerical solutions based on the first calculation. User can also provide an initial guess based on the physical meaning of the unknowns of the coupled PDEs.

In the AMM method, the mesh topology is kept unchanged but the mesh grid points are concentrated inside the physical domain where the solution of the physical problem has large gradient or large solution variations (Beckett *et al.* 2001). Compared with the AMR method, the AMM method does not require complicated data structures for mesh refinement. Furthermore, the AMM method does not need to interpolate the numerical solution between different levels of mesh refinement. For the boundary coupled problem, the AMM method coincides with the arbitrary Lagrangian–Eulerian (ALE) method, which is popularly used in the fluid–structure interface problem. In a Lagrangian formulation, the update of a mesh follows the change of the computational domain. The advantage of a Lagrangian expression for a moving mesh is that it can retain a sharp material interface when the mesh is deformed properly. However, one of the challenges of using a Lagrangian-type formulation is that the mesh can become severely distorted or can overlap itself. So after the computational domain is updated, the mesh points inside the domain should be smoothly adjusted using the AMM method. Of course, it can also be possible to use both the AMR and the AMM methods during simulation, e.g., when the domain and boundary couplings exist on different subdomains in one simulation.

1.18.1.4 Multiphysics Design Optimization

Microsystem components are hard to design, because manufacturing technology significantly constrains our creative possibilities, and the sensitive regions where sensor and actuator effects are active are usually not neatly lumped but extend over large portions of a device, making system design approaches hard to apply. Currently, very few automated design methods are available for MEMS (Sigmund 2001a, b, Yin and Ananthasuresh 2002). To establish such methods, a number of numerical components must be in place so that together they can form an integrated design and optimization tool. Here we introduce an integrated structural topology optimization method for multiphysics in MEMS.

1.18.1.4.1 Structural layout optimization

The optimization of structural layout potentially has a great effect on the performance of a structure through parameters such as weight, volume, stress distribution and dynamic response, etc. Generally, structural layout optimization can be classified into three categories: size, shape, and topology optimization.

For size optimization, only simple geometrical parameters, such as the thickness of a plate or the length of a bar, are chosen as the design variable. The shape of the structure is fixed throughout the optimization procedure. For shape optimization, the design strategy to improve the performance of the structure is to adjust the shapes of structural boundaries. The default hypothesis is that the topology of the structure or the connections of the structure remain unchanged. Therefore, shape optimization

cannot answer topological questions. The topology optimization method is a suitable and an efficient methodology in this case.

When compared with size and shape optimization, the topology optimization method is an advanced method, which can obtain an optimal topology, shape, and size at the same time (Rozvany 2001). The continuous topology optimization method has been used to design the layout of structures since the development of the homogenization method (Bendsoe and Kikuchi 1988, Bendsoe and Sigmund 2003). Based on the ground structure method, an initial design domain is discretized by finite element meshing. According to this fixed meshing, an optimal topology can be obtained without guessing an initial topology of the structure. It should be noted that the topology optimization method is equivalent to an inverse problem (first kind is Fredholm integral equation), which considers the material parameter distribution for coupled problems. There is no general theory to obtain a stable solution for this kind of inverse problem. Numerical regularization is generally used in most cases.

1.18.1.4.2 Simultaneous simulation and optimization method

To date, topology optimization has covered several applications, such as functional materials, micro sensors, and compliant actuators, in MEMS design. As an example, let us consider electrothermal microactuator, which is actuated through resistive heating and thereby thermal expansion effects, modeling of electric, thermal, and elastic fields, and then calculates the sensitivity from these three fields to implement the structural optimization iteratively. Numerical instabilities during the optimization procedure, checkerboard patterns and mesh dependency, have been overcome based on the numerical filters or the homogenization method. Geometrically nonlinear deformation, which generally occurs as a snap through or buckling phenomenon in an electrostatic actuated switch, is also a research topic in nonlinear topology optimization. However, further application of structural topology optimization in MEMS is impeded by the complexity of the optimization algorithm itself and the difficulty in finding a unified expression for the optimization procedure for diverse coupled problems in MEMS. For example, there are many surface-type coupling effects in MEMS, such as the structure actuated using electrostatic force. It is known that the deformation of the structure also influences the magnitude of the electrostatic force and vice versa. During structural optimization the electrostatic force is design-dependent. It means that the force is simultaneously changed following the change of the structural boundary in the material frame during the procedure of optimization. Here an efficient and accurate expression of these forces is a key point and has also been a challenging problem to date.

Traditionally, topology optimization is based on the sequentially COM. Specified physical problems can be discretized and solved first, and then sensitivity based on the discretized numerical solutions can be calculated. This is a discretize-then-optimize approach. Another optimization strategy is the COM. One can establish coupled physical equations and an optimization equation at the original PDE level, and then discretize these coupled PDEs together by a numerical method. This is an optimize-then-discretize approach. Here we describe the first as the sequential optimization method (SOM) and the second as the COM (Arora and Wang 2005).

The COM has several advantages when compared with the SOM. First, in the SOM, it is obvious that the effect of an optimization iteration is strongly dependent on the accuracy of their discretized physical problems, because the discretized solution of physical problems is often a noisy or a discontinuous approximation of the initial problem. Numerical errors pollute the quality of sensitivity and influence the convergence. For the COM, the sensitivity, which connects the physical equations and the optimization equation, is calculated by the true physical relationship without any numerical discretization error. So it is often less noisy and more accurate than the discretized optimization method. Second, in the SOM, the state variables are treated as implicit functions of the design variable. Then the design sensitivity analysis should include the derivative terms of the state variables and the design variables. In the COM, the state variables are set independent of the design variables, i.e., the state and design variables are defined in a composite vector. One of the distinguishing advantages of the method is that the derivation of the sensitivity is relatively straightforward. This is especially true when one wants to implement structural optimization in which the state variables are controlled by multiphysical problems. For example, equilibrium equations could come from mechanics, electrostatics, and thermoelectric effects. The state variables include displacement, electrostatic field, thermal stress, etc. The coupling effects include surface coupling of

mechanism and electrostatics, and domain coupling of mechanics and thermoelectric effects. In this case, the sensitivity analysis of the SOM is extremely complicated, and the expression of the COM is much simpler for coupled problems. Then one question is how to establish a relationship between the decoupled state and design variables. One of the possibilities is to derive a corresponding optimization equation based on the specified objective and to solve the fully coupled equilibrium equations and optimization equations simultaneously.

Typically, the element density-type method, such as the solid isotropic microstructures with penalization (SIMP) method, is used to implement structural topology optimization (Allaire *et al.* 2004). However, the expression of the structural layout is nonsmooth because piecewise constant density values are used to express the material domain. Even though this does not cause serious numerical problems for the structural optimization problems, nonsmooth expression of material leads to unexpected results for other types of structural optimization, such as thermal problems (Yoon and Kim 2006). One of the simplest remedies to overcome the problem of nonsmooth expression of a material is the nodal density-type method, where material density is interpolated smoothly with nodal density values. In MEMS simulation, the actuation force is normally either a body force or a surface force. During structural topology optimization, these two kinds of forces are referred to as design-dependent loads, which mean that the load vectors change with the evolution of the structural layout. With either the element- or the nodal density-type method, there is no explicitly defined material boundary. So the implementation of a design-dependent load is not straightforward (Du and Olhoff 2004, Hammer and Olhoff 2000). Basically, the level set method is shape optimization method. Because of its ability to merge holes during the optimization, it has recently become popular in structural topology optimization. When compared with the density-type method, one of the distinguishing advantages of the level set method is the material boundary with zero-level set contours; therefore the design-dependent load is easily implemented during the optimization procedure.

1.18.1.4.3 Derivation of the optimization equation

In this section we introduce the use of the level set method to transform a typical optimization procedure to a coupled PDE expression. We first introduce how to establish a coupled simulation and optimization for a single mechanical and electrostatic field problem, so that the structural optimization procedure can be solved as a pseudo-time-dependent coupled PDE problem. Then we extend this procedure to the electromechanical-coupled optimization problem in MEMS.

A typical structural topology optimization problem for 2D compliance minimization can be written as:

$$\begin{aligned} \text{Min:} \quad & c[\phi,\varepsilon] = \int_{\Omega} \frac{1}{2} E[\phi]\varepsilon^{\mathrm{T}} D\varepsilon \, \mathrm{d}\Omega \\ \text{s.t.} \quad & \nabla\cdot(E[\phi]\sigma) = f \\ & \int_{\Omega} H[\phi]\mathrm{d}\Omega = \mathrm{Vol}^{*} \end{aligned} \qquad [3]$$

where Ω is the design domain, u is the displacement field calculated using the linear elastic equilibrium equation, ε is the strain tensor, σ is the stress tensor, E is the design variable and is defined by the level set surface ϕ as $E[\phi] = H[\phi] + (1 - H[\phi])E_{\min}$, where $E_{\min}$ is a small number, such as 0.001, to avoid singularity of the stiffness matrix, D is the elasticity matrix in which the material Young's modulus is E_0, and Vol^{*} is the material volume constraint in the deign domain.

The level set function $\phi[x]$ is an implicit function for a given domain Ω with smooth boundary, which satisfies the following (Osher and Fedkiw 2002):

$$\begin{aligned} &\phi[x] > 0, \quad x \in \Omega^{+} \text{ (material)} \\ &\phi[x] = 0, \quad x \in \partial\Omega \text{ (boundary)} \\ &\phi[x] < 0, \quad x \in \Omega^{-} \text{ (hole)} \end{aligned} \qquad [4]$$

The area constraint condition, which is typically used in compliance minimization design, is presented. To derive the level set equation, which moves the material boundaries, we combine the objective function and constraints together by using a Lagrangian formulation, and then derive a corresponding Euler–Lagrangian equation (Nocedal and Wright 1999). Using the Lagrangian multiplier λ, we can rewrite problem [3] as

$$\mathcal{J}[\varepsilon,\phi,\lambda] = \int_{\Omega} \left(\frac{1}{2} E[\phi]\varepsilon^{\mathrm{T}} D\varepsilon + \lambda \left(H[\phi] - \frac{\mathrm{Vol}^{*}}{\mathrm{Vol}^{\Omega}} \right) \right) \mathrm{d}\Omega \qquad [5]$$

where Vol^{Ω} is the area of the entire design domain. From classical calculus, we know that the extrema of functional [5] is attained at the position where $\mathcal{J}' = 0$. To obtain the level set equation, which uses a level set surface to express the structural topology implicitly, we need to calculate the variation of the level set surface. Only the normal velocity of the material boundary influences the change of the shape. The

tangential velocity will not influence the deformation of the geometry, merely its parameterization (Sapiro 2001). We assume that the zero level set contour $\phi = 0$ moves only in the normal direction. For an infinitesimal variation δl, which is along the normal direction $\vec{n}$, a new zero-level set is ϕ' and we have

$$\begin{aligned}\delta\phi|_{\phi=0} &= \phi'[x] - \phi[x] = \phi[x + \delta l \cdot \vec{n}] - \phi[x] \\ &= \nabla\phi \cdot (\delta l \cdot \vec{n}) = |\nabla\phi|\delta\end{aligned} \quad [6]$$

where the normal vector $\vec{n} = \nabla\phi/|\nabla\phi|$. For problem [5], the variation of the level set surface on the material boundary can be expressed as

$$\delta_\phi \mathcal{J} = \int_\Omega \left(\frac{1}{2}(1 - E_{\min})\varepsilon^{\mathrm{T}} D\varepsilon + \lambda\right)\delta[\phi]|\nabla\phi|\delta l \, \mathrm{d}\Omega \quad [7]$$

The corresponding Euler–Lagrangian equation at the extreme value point is

$$\left(\frac{1}{2}(1 - E_{\min})\varepsilon^{\mathrm{T}} D\varepsilon + \lambda\right)\delta[\phi]|\nabla\phi| = 0 \quad [8]$$

In most cases, it is impossible to solve eqn [8] directly. One general technique is to solve set equation numerically

$$\frac{\partial\phi}{\partial t} - \left(\frac{1}{2}(1 - E_{\min})\varepsilon^{\mathrm{T}} D\varepsilon + \lambda\right)\delta[\phi]|\nabla\phi| = 0 \quad [9]$$

The auxiliary variable t is used as pseudo-time. The method is also referred to as gradient descent flow. In the case of a well-posed optimization problem, the solution converges to a local minimum based on different initial values. In this way the 2D compliance minimization topology optimization problem can be transformed into the following coupled plane stress equation and time-dependent reaction diffusion equation

$$\begin{aligned}\nabla\cdot(E[\phi]\sigma) &= f \\ \frac{\partial\phi}{\partial t} - \left(\frac{1}{2}(1 - E_{\min})\varepsilon^{\mathrm{T}} D\varepsilon + \lambda\right)\delta[\phi]|\nabla\phi| &= \alpha\,\Delta\phi\end{aligned} \quad [10]$$

with a suitable initial condition ϕ_0. Here an artificial diffusion term $\alpha\,\Delta\phi$ is added to overcome numerical oscillations during the evolution of the level set surface. The coefficient α is proportional to the mesh size and the reaction term. The Lagrangian multiplier is expressed as

$$\lambda = \frac{\displaystyle\int_\Omega \left(\frac{1}{2}(1 - E_{\min})\varepsilon^{\mathrm{T}} D\varepsilon\right)\delta^2[\phi]|\nabla\phi|\mathrm{d}\Omega}{\displaystyle\int_\Omega \delta^2[\phi]|\nabla\phi|\mathrm{d}\Omega} \quad [11]$$

Here we call the first equation in eqn [10] as the state equation for the direct problem and the second equation as the optimization equation for the structural optimization problem.

For the electrostatic problem, one of the most commonly used objective functions is the stored potential energy inside a design domain. The electrical energy can be expressed as $\mu[\phi]\nabla V^{\mathrm{T}}\nabla V/2$ where V is the electrical potential, ∇V is the electric field vector, μ is the dielectric permittivity, which is defined by a level set surface ϕ as

$$\mu[\phi] = \mu_{\mathrm{mat}} H[\phi] + (1 - H[\phi])\mu_{\mathrm{air}}$$

where μ_{air} is the dielectric permittivity of the air and μ_{mat} is the dielectric permittivity of the design material. The equilibrium equation, which controls the electrostatic problem, is the Poisson equation. Following a similar procedure as for the mechanical problem, a maximization of the electrostatic energy problem can be expressed as

$$\begin{aligned}\text{Max:} \quad C[\phi, \nabla V] &= \int_\Omega \frac{1}{2}\mu[\phi]\nabla V^{\mathrm{T}}\nabla V \, \mathrm{d}\Omega \\ \text{s.t. } \nabla\cdot(\mu[\phi]\nabla V) &= 0 \\ \int_\Omega H[\phi]\mathrm{d}\Omega &= \mathrm{Vol}^*\end{aligned} \quad [12]$$

A fully coupled PDE expression of topology optimization for the electric potential energy problem with area constraint can be similarly expressed as

$$\begin{aligned}\nabla\cdot(\mu[\phi]\nabla V) &= 0 \\ \frac{\partial\phi}{\partial t} - \left(\frac{1}{2}(\mu_{\mathrm{mat}} - \mu_{\mathrm{air}})\nabla V^{\mathrm{T}}\nabla V + \lambda\right)\delta[\phi]|\nabla\phi| &= \alpha\,\Delta\phi\end{aligned} \quad [13]$$

where

$$\lambda = \frac{\displaystyle\int_\Omega \left(\frac{1}{2}(\mu_{\mathrm{mat}} - \mu_{\mathrm{air}})\nabla V^{\mathrm{T}}\nabla V\right)\delta^2[\phi]|\nabla\phi|\mathrm{d}\Omega}{\displaystyle\int_\Omega \delta^2[\phi]|\nabla\phi|\mathrm{d}\Omega} \quad [14]$$

Hence as for the mechanical problem, an artificial diffusion term is added to the right-hand side of the second equation in eqn [13].

During the procedure, the amplitude and the direction of the electrostatic force will change with the update of the topology of the structure and its deformation. Hence we have a coupled optimization problem. There are two different methods to solve this surface coupling problem: the sequentially coupled solver and the directly coupled solver. To

implement structural optimization with the coupled method, the directly coupled solver is used to calculate the structural deformation. The design domain is shown in **Figure 1**. There are two subdomains. Subdomain 1, the upper strip of the beam marked in black, has specified constant material properties. Subdomain 2 is the optimization domain and is white in **Figure 1(a), 1(c), and 1(e)**. Here we discuss three different load cases. In the first case, a fixed load vector is added on the middle point of the upper surface of the whole domain. In the second case, the electrostatic force acts on the structure. In the last case, both the fixed load and the electrostatic force act together on the whole structure. The objective of structural optimization is still compliance minimization with area constraint. The material inside both subdomains is aluminum with Young's modulus of 70 GPa and a Poisson ratio of 0.33.

In the first case, we use the compliance minimization model in eqn [3]. In addition, we consider the geometrically large deformation for the elastic equilibrium equation. So the strain tensor ε has a nonlinear relationship with the displacement vector. Since we use the COM (Liu *et al.* 2005) in which the design variable ϕ and the displacement vector are independent of each other, the optimization equation is still the same as the second equation in eqn [10]. The first equation in eqn [10] is solved using a nonlinear solver for large deformation analysis. The optimized structural topology is shown in **Figure 1(b)**.

In the second load case, a consistent solution for the electromechanical problem is needed. To coincide with the fully coupled optimization method, we choose a direct coupled solver where the deformed mesh caused by the structural deformation is simulated by the ALE method (Jakobsen *et al.* 2001). The couple ALE method can be implemented using the following coupled equations:

$$\begin{aligned} \nabla \cdot (\mu[\phi] \nabla V_X) &= 0 \\ \nabla \cdot (E[\phi]\sigma) &= F_{\text{ele}} \\ \nabla \cdot (G \nabla X) &= 0 \end{aligned} \qquad [15]$$

where X represents the deformed nodal position because of structural deformation, G is the mesh monitor function, which controls a local property of the deformed mesh, V_X means that the value of the electrical potential is calculated on the deformed mesh, and

$$\mu[\phi] = \mu_{\max} H[\phi] + (1 - H[\phi])\mu_{\text{air}}$$

where $\mu_{\max}$ is a large number to approximate the distribution of the electrical potential inside the conductor. In this example $\mu_{\max} = 100$. Using the moving mesh equation (the third equation in eqn [15]), all of the displacement, force, and electrical potential boundary conditions can be implemented in the original coordinates. This method can deal with the electromechanical coupling directly before pull-in occurs. For the optimization examples we presented here, the fixed–fixed end beam in subdomain 1 is used

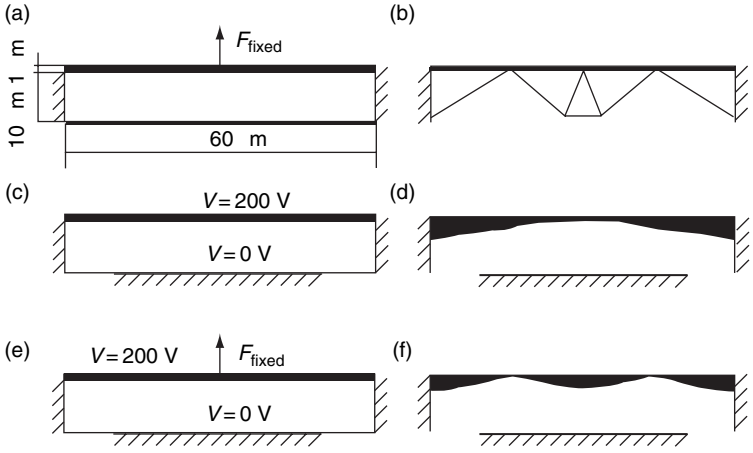


Figure 1 Compliance-minimized topology optimization of a fixed–fixed end beam and three load cases. (a) and (b) Initial design domain and optimized topology with fixed load vector; (c) and (d) initial design domain and optimized topology with electrostatic force; (e) and (f) initial design domain and optimized topology with both the fixed load vector and the electrostatic force. The beam in (a) has a stiffness of 0.019 N m^{-1} at its center point, and for the situation in (c), it has a theoretical pull-in voltage of 1700 V. In all simulations F was set to 10 μN. (Source: Lui Z, Korvink J G, Reed M L, Multiphysics for structural topology optimization. *Sensor Letters*. **4**, 191–199.)

to avoid pull-in occurring during the optimization procedure with reasonable electrical potential conditions. The coupled optimization equations are:

$$\begin{aligned}\nabla\cdot(\mu[\phi]\nabla V_X) &= 0\\ \nabla\cdot(E[\phi]\sigma) &= F_{\text{ele}}\\ \nabla\cdot(\nabla X) &= 0\\ \frac{\partial\phi}{\partial t} - \left(\frac{1}{2}(1-E_{\min})\varepsilon^{\mathrm{T}} D\varepsilon + \lambda\right)\delta[\phi]|\nabla\phi| &= \alpha\,\Delta\phi\end{aligned} \quad [16]$$

Here the mesh monitor function is the unit function so that the mesh is smoothly modified on the whole design domain. On each mesh node, there are six dofs, two displacement dofs, one electrical potential dof, two moving mesh position dofs, and one level set surface dof. However, the optimization equations remain unchanged despite the complicated coupled forward problems. The optimized result is shown in **Figure 1(d)**.

In the last case, both the fixed load vector and the electrostatic force act on the structure. So this is a multiple load optimization problem. Based on the analysis of the first two cases, the optimization equations for these two loads are the following:

$$\begin{aligned}\nabla\cdot(E[\phi]\sigma_1) &= F_{\text{fixed}}\\ \nabla\cdot(\mu[\phi]\nabla V_X) &= 0\\ \nabla\cdot(E[\phi]\sigma_2) &= F_{\text{ele}}\\ \nabla\cdot(\nabla X) &= 0\\ \frac{\partial\phi}{\partial t} - \left(\frac{1}{2}(1-E_{\min})\left(\varepsilon_1^{\mathrm{T}} D\varepsilon_1 + \varepsilon_2^{\mathrm{T}} D\varepsilon_2\right) + \lambda\right)&\\ \times\,\delta[\phi]|\nabla\phi| &= \alpha\,\Delta\phi\end{aligned} \quad [17]$$

where ε_1 and ε_2 are the strain vectors corresponding to two load cases. The optimized result is shown in **Figure 1(f)**.

1.18.2 Multiscale Simulation

Modeling and simulation of the behavior of materials in various applications of modern microsystems is a challenging task. On the other hand, high costs or restricted availability of samples for analysis requires modern microfluidic systems to operate in the picoliter regime and below. The ultimate goal is to make available simulation tools that emerge from a theoretical modeling that ranges from quantum physics of atomic-scale phenomena, covering up to continuum descriptions of macroscopic behavior, and going even further to system behavior. Such engineering tools, supporting design processes in microsystem technology, are still missing. In microsystems, the presence of length scales of several orders of magnitude requires the coupling of different tools appropriate for simulation at the respective spatial scale. The use of advanced computer-aided design (CAD) tools promises reduction in the extent of physical testing necessary to prototype a device. Activities to incorporate various physical models at different length scales (e.g., Nielaba *et al.* 2002) exist.

The simulation of complex processes in liquids and the modeling of the material properties of liquids are tasks that require resolutions on the molecular scale and below. The computational effort at the atomistic level is demanding. Consequently, atomistic approaches should undergo a coarse graining procedure. Simulating the behavior of a reasonable amount of liquid with various components in a common framework with a reasonable computational effort – as it is mandatory for the application of microfluidic systems in a design process – can be realized only by superimposing microscopic molecular dynamics (MD) approaches on coarse-grained particle dynamics methods. Typically, the microscopic MD simulation is computationally more expensive than coarse-grained particle dynamics solvers and therefore MD must be restricted to very specific parts of the whole system where an atomistic resolution is desirable. It is even conceivable that a successive coupling to continuum Navier–Stokes models is necessary. To overcome the problem of the costly particle dynamics calculation being repeated in every iterative loop of the continuum solver, a microscale simulation featuring noniterative and explicit time marching must be coupled to an equivalent model acting at a much coarser length scale. This eventually leads to a feasible simulation process for engineering CAD applications in multiscale microfluidic systems. A promising technique to fulfill the above-mentioned requirements is dissipative particle dynamics (DPD), a particle dynamic simulation technique for simulations above the molecular length scale (Hoogerbrugge and Koelman 1992), which is described in Section 1.18.2.1. A second approach to fluid dynamics that couple the DPD model to more macroscopical environments is given by the lattice Boltzmann method (Wolf-Gladrow 2000).

1.18.2.1 Particle Dynamics Methods

Molecular dynamics is the most popular particle dynamics method, and it has a long-standing tradition. Its main ingredients are the interaction potentials of the atomistically resolved particles together with an algorithm to integrate the set of differential equations for this many-body problem. In essence it is an

integration of Newton's equation of motion. Much effort has been put into the development of efficient algorithms. On the other hand the interaction potentials can be derived from first principles and their form is crucial for the properties of the simulated material.

Coarse-grained particle dynamics methods loose this direct connection to the first principles derivation of the interaction potentials. Instead, they must rely on either MD simulations or macroscopic measurements to adjust the respective parameters. Nevertheless, they benefit from the development of the efficient integration techniques for MD systems.

1.18.2.1.1 Coarse-graining microscopic methods vs refinement of continuum models

At a microscopic length scale on the order of interatomic distances in fluids or solids, MD represents an attractive simulation technique in computational fluid dynamics. However, at the length scale from 10 to 1000 nm a coarse graining of the molecular model due to rising computational costs is desirable.

There are two ways to arrive at a particle description of microfluidic phenomena: coarse graining of MD and a direct Lagrangian discretization of the equations governing hydrodynamics. The former leads to the DPD variants discussed in Section 1.18.2.1.2, while the latter is exemplarily discussed here by the method of smoothed particle hydrodynamics (SPH) (Koumoutsakos 2005).

SPH is a method for the simulation of hydrodynamic phenomena. Originally, SPH was developed to overcome hydrodynamic problems in astrophysics. Eventually it was adopted for the investigation of laboratory scale problems as well. It is a direct discretization of the macroscopic hydrodynamic equations on Lagrangian grid. The spatial derivatives are found by means of an interpolation formula. One of the biggest drawbacks of SPH is the fact that thermal fluctuations are not included in the equations of motion. For modeling flows at astronomical length scales, fluctuations most certainly do not play a role. In applying SPH to microfluidics where the dimensions are orders of magnitude smaller, fluctuations definitely do come into play. In a recent effort a generalization of SPH including thermal fluctuations has been proposed (Espanol and Revenga 2003) and termed smoothed dissipative particle dynamics (SDPD).

1.18.2.1.2 Dissipative particle dynamics

DPD, first introduced by Hoogerbrugge and Koelman (1992), is expected to be suitable for the intermediate range of length scale. Its relation to MD as well as to continuum methods has been investigated (Flekkoy and Coveney 1999, Flekkoy *et al.* 2000). DPD is a mesoscopic simulation method capable of bridging the gap between atomistic and mesoscopic simulations (Groot and Warren 1997).

In essence, DPD is a treatment of the dynamics of quasi-particles representing small sets of the liquid's molecules by stochastic differential equations similar to a Langevin approach. It combines features from MD and lattice gas methods. **Figure 2** provides a schematic representation of the context in which DPD has to be seen with respect to other simulation methods. Since its first introduction this method has been applied for the simulation of a wide range of phenomena, especially in material science. Thorough investigations have been carried out to understand the capabilities of DPD when applied to computational fluid dynamics. Recently, DPD has experienced an increasing interest of different research groups for its improvement in several aspects. Its algorithmic optimization was one focus, making it an appropriate method for application in engineering fields (Kauzlaric *et al.* 2005). The inclusion of energy and momentum conservation in the particle–particle interaction for the set of stochastic differential equations describing a DPD models is an important extension for heat flow applications. Recently, phase change models have been included, which is built on the energy-conserving DPD (DPDE) models. A solid–liquid phase transition and liquid–vapor coexistence can be modeled with DPD without turning down all the advantages of the method that come from using larger time steps than in MD.

We will briefly describe the basic ingredients of DPD. Every dissipative particle i with its position $\mathbf{r}_i$ and its momentum $\mathbf{p}_i$ or its velocity $\mathbf{v}_i$ is propagated according to Newton's law with the force $\mathbf{F}_i$ acting on a particle. Thus the equations of motion is given as follows:

$$\dot{\mathbf{r}}_i = \mathbf{v}_i \qquad [18]$$

$$\dot{\mathbf{v}}_i = \mathbf{F}_i = \sum_{j \neq i} \left(\mathbf{F}_{ij}^{\mathrm{C}} + \mathbf{F}_{ij}^{\mathrm{D}} + \mathbf{F}_{ij}^{\mathrm{R}} \right) \qquad [19]$$

where the force has three main contributions due to other particles in the flow, namely a conservative part $\mathbf{F}^{\mathrm{C}}$, a dissipative contribution $\mathbf{F}^{\mathrm{D}}$, and a fluctuating force $\mathbf{F}^{\mathrm{R}}$. The latter two are defined as

$$F_{ij}^{\mathrm{D}} = -\gamma_{ij} w[r_{ij}](\mathbf{e}_{ij} \cdot \mathbf{v}_{ij})\mathbf{e}_{ij} \qquad [20]$$

$$F_{ij}^{\mathrm{R}} = \sigma_{ij} w^{1/2}[r_{ij}]\mathbf{e}_{ij}\xi_{ij}[t] \qquad [21]$$

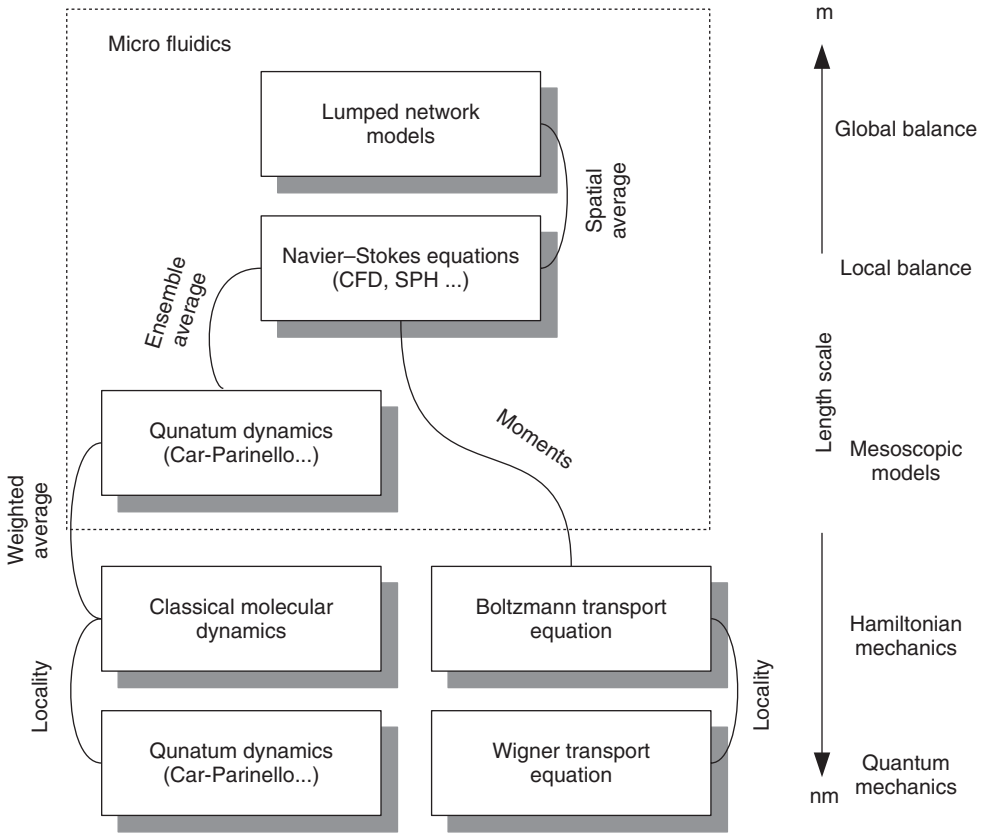


Figure 2 DPD model in the context of other simulation methods. BD, Brownian dynamics; CFD, computational fluid dynamics; DPD, dissipative particle dynamics; SPH, smoothed particle hydrodynamics. Deriving a mesoscopic technique from a microscopic one or a macroscopic technique from a mesoscopic one usually involves some sort of projection (averaging).

where γ and σ are the dissipation strength and the fluctuation amplitude, respectively. $w[r_{ij}]$ is a weighting function of the distance r_{ij} of the two dissipative particles i and j, while $\mathbf{e}_{ij}$ is the unit vector pointing from particle i to particle j, and $\mathbf{v}_{ij}$ is the relative velocity of the two particles. $\xi_{ij}[t]$ describes the Gaussian white noise according to the following equation

$$\langle \xi_{ij}[t] \rangle = 0 \qquad [22]$$

$$\langle \xi_{ij} i[t] \xi_{i'j'}(t') \rangle = (\delta_{ii'}\delta_{jj'} + \delta_{ij'}\delta_{ji'})\delta[t-t'] \qquad [23]$$

The modeling of material properties is done in DPD by adjusting the parameters entering the forces given in eqn [19]. These are γ and σ and form the conservative potential. An advantage in computational effort results only if $\mathbf{F}^{\mathrm{C}}$ is a soft repulsive potential. This allows for much larger time steps than in MD. On the other hand, the lack of an attractive part in $\mathbf{F}^{\mathrm{C}}$ does not allow for free surface flow. Hence a density-dependent part must be added to the conservative potential accounting for the excess surface energy in order to describe liquid–gas interfaces. Since this is an extension of the simple DPD model to many-body behavior, it is termed as MDPD in the literature (e.g., Paganobarraga and Frenkel 2001).

So far we treated only isothermal DPD even with its extension to MDPD. To include thermal properties of the material the internal energy of the individual dissipative particle and its interactions with the translational dofs should be taken into account. The internal energy as an example for the extension of the DPD to DPDE (e.g., Pastewka *et al.* 2006) devises a way for the introduction of various other dofs. Clearly, their equation of motion should be coupled to those of the existing dofs, and is a task for material modeling. With the extension to

MDPDE, the phase description of phase transitions has become possible, which makes this method also very attractive in engineering applications when compared with direct numerical treatment of the hydrodynamic equations.

1.18.2.1.3 Applications of DPD

The area of applications for DPD is growing very rapidly. Its range for different simulation tasks is listed as follows. A discussion on mesoscopic dynamics of colloids can be found in Dwinzel *et al.* (2002), the treatment of binary fluids and immiscible fluids is reported in Coveney and Novik (1996) and Dwinzel and Yuen (2000), the rheology of dense colloidal suspensions can be found in Boek *et al.* (1997), and many more. The important step was the derivation of a thermodynamically conservative interacting force from a given free energy density, which allowed for the implementation of phase coexistence and the description of phase boundaries (Paganobarraga and Frenkel 2001). The implementation of DPDE was a very important development to make the model applicable for phase transitions (Boneyt Avalos and Mackie 1997). Newer developments overcome the limitations of DPDE with respect to the time steps in simulations (Pastewka *et al.* 2006), and thus allow for its application to the study of complex fluids. There are recent investigations on the applicability of DPD in the framework of micelle formation (Schulz *et al.* 2004), which makes it a promising candidate for the task of bridging the gap of length and time scales for the simulation of ionic liquids as well, even if there are still strong implications for the applicability of the method. In Pool and Bolhuis (2006), the failure in the prediction of a critical micelle concentration by an error of thirteen orders of magnitude was exposed impressively. We noted that there is a need for further developments to make DPD applicable in specific materials systems, especially when it comes to difficult rheological properties (Kauzlaric *et al.* 2003a, b, 2004).

In **Figure 3** the filling of a bending specimen mold was simulated. Experimental findings reported on the jetting behavior around the obstacle in the flow field. This was exactly reproduced by the DPD simulation,

Figure 3 Filling simulation of a bending specimen mold performed by dissipative particle dynamics (DPD).

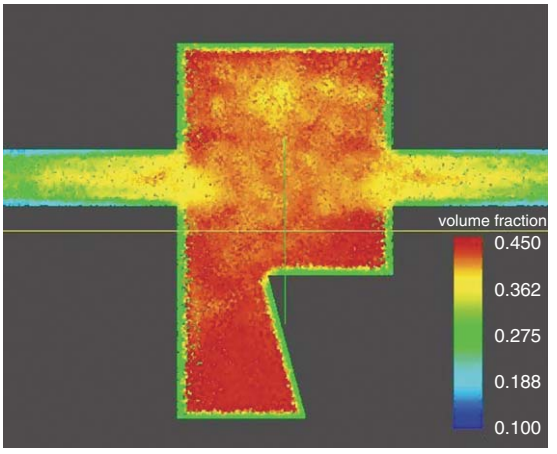


Figure 4 Characteristic powder distribution in micropowder injection molding.

while a conventional simulation tool failed in the prediction of jetting. A very important information in powder injection molding is the ceramic or metal powder distribution after the molding process. A homogeneous mold mass is subject to demixing due to the shear rates occurring in the molding process. **Figure 4** shows that by the extension of the DPD model a direct prediction of the powder concentration is possible.

1.18.2.2 Kinetic Models

Kinetic theory, as opposed to particle dynamics, starts from a different point of view for a system of many dofs. A state of N particles is given by its $3N$ coordinates r_i and its $3N$ momenta p_i to give $6N$ dofs. Instead of solving the coupled $6N$ equations of motion, as done in a particle dynamics approach, we rather regard the system as represented by a probability density $f_N[r_1, \ldots, r_{3N}, p_1, \ldots, p_{3N}]$. The quantity f_N has the meaning of a correlation function for the $6N$ dofs. By knowing the interaction potentials of the N particles we can form the Hamilton function $H[r_1, ..., r_{3N}, p_1, ..., p_{3N}]$ and write down the $6N$ Hamiltonian equations of motion as follows

$$\dot{r}_i = \frac{H}{\partial p_i} \qquad [24]$$

$$\dot{p}_i = -\frac{H}{\partial r_i} \qquad [25]$$

This is the starting point of particle dynamics methods, for example, MD, which proceed by solving eqns [24] and [25]. Instead of doing so we can write

down the equation of motion for f_N, which in this case in treated as an observable quantity, to give

$$\frac{\partial f_N[r_i, p_i, t]}{\partial t} + \frac{\partial r_j}{\partial t}\frac{\partial f_N[r_i, p_i, t]}{\partial r_j} + \frac{\partial p_j}{\partial t}\frac{\partial f_N[r_i, p_i, t]}{\partial p_j} = 0 \quad [26]$$

For solving eqn [26] the information from eqns [24] and [25] is needed and thus there is no advantage. In kinetic models the interest is not in the full correlation function f_N, but instead in the probability of finding one particle in phase space at a certain point f_1 [**r**, **p**]. This function can be obtained by integrating f_N over all the $6(N-1)$ variables that are not of interest. Together with Boltzmann's assumption of molecular chaos, the equation of motion for a single particle distribution function then is given as follows:

$$\frac{\partial f[\boldsymbol{r}, \boldsymbol{p}, t]}{\partial t} + \dot{\boldsymbol{r}}\frac{\partial f[\boldsymbol{r}, \boldsymbol{p}, t]}{\partial \boldsymbol{r}} + \dot{\boldsymbol{p}}\frac{\partial f[\boldsymbol{r}, \boldsymbol{p}, t]}{\partial \boldsymbol{p}} = C[\boldsymbol{r}, \boldsymbol{p}, t]_{\text{coll}} \quad [27]$$

where $C[\boldsymbol{r}, \boldsymbol{p}, t]_{\text{coll}}$ is a collision term that derives from the fact that we averaged over the $6(N-1)$ dofs. Eqn [27] is called the Boltzmann transport equation (BTE). Solution techniques for eqn [27] that make use of cellular automata is briefly discussed in Section 1.18.2.2.1.

1.18.2.2.1 Approximations of the BTE

Although it is a discretization method of the BTE, the lattice Boltzmann automaton (Chen and Doolen 1998) was introduced for the purpose of solving the NSE, which in turn can be derived from the BTE via a Chapman–Enskog Expansion (CEE).

The NSE is the equation of motion of the first moment of the single-particle distribution function. An approximation of the BTE up to a finite order satisfies the NSE by design, even though it could be more general. At first, the most simplified lattice Boltzmann model still consistent with the NSE has been sought. Examples for simplifications that do not necessarily improve efficiency are the single relaxation time approximation (Bhatnagar *et al.* 1954) and models with a minimal set of dofs like the D3Q13 model (13 speeds in three dimensions) (d'Humieres *et al.* 2001). In these models, efficiency is often decreased by the introduction of instabilities. However, there is a very simple way to make even the worst model stable, e.g., through increased viscosity. By applying the scaling laws of fluid dynamics it is still possible to simulate any flow condition by choosing appropriate time steps and grid spacings for the model-dependent lowest attainable viscosity. The performance penalty is severe and can be calculated from the same scaling laws. Two flow conditions are said to be similar if their Reynolds number is identical. The Reynolds number is defined as $Re = Lv/\nu$. Here L is a characteristic flow length, v is the flow's average velocity, and ν is the kinematic viscosity. Scaling laws are useful to derive numerically tractable dimensionless equations. It is easily seen that, for a constant Reynolds number, increasing viscosity can only be compensated for by increasing either v or L. Higher velocities will typically not result in more stable solutions since the lattice Boltzmann method is an explicit finite difference scheme and has to satisfy the CFL condition (Courant *et al.* 1967). The only remaining option is to take the expensive road and decrease L. This translates to finer grid spacings and shorter time steps. Since we are interested in time-varying 3D simulations we see that lowering the viscosity by one order of magnitude leads to four orders of magnitude decline in cell update operations and three orders of magnitude decline in the number of cells. In contrast, the above-mentioned simplifications result only in small gains per update step, typically smaller than a factor of 2 in memory and computational time. This cannot be compared at all with what we gain by finding a way of lowering the viscosity by many orders of magnitude without sacrificing stability. This is done, not by introducing new concepts or models, by merely removing other types of simplification, which crept into the scheme due to the effort to relate the lattice Boltzmann automata to the NSE rather than to see it as a discretization of the BTE. A recently developed new scheme called cascaded digital lattice Boltzmann automata (CDLBA) is able to overcome all the instability issues of conventional LBA (Geier *et al.* 2006a).

1.18.2.2.2 Applications of CDLBA

The stability of CDLBA is impressively demonstrated by the turbulent flows in **Figures 5** and **6**.

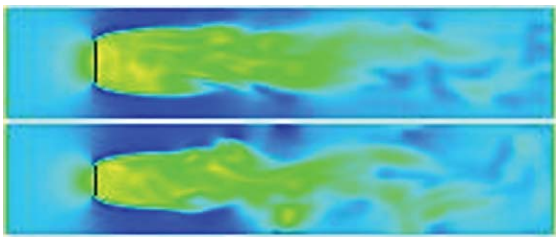


Figure 5 Top and side view of the wake field of a rectangular object in a gas flow, $Re \approx 500\,000$.

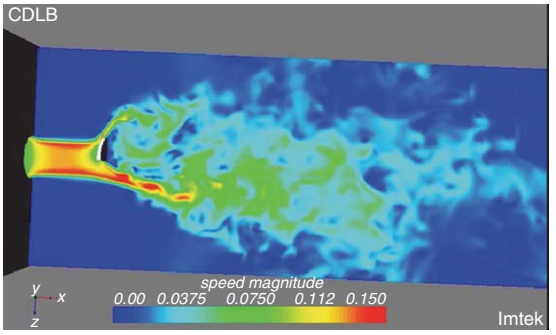


Figure 6 Arbitrarily shaped object in a gas flow from an orifice, $Re \approx 1\,000\,000$.

The flow over a rectangular plate has a Reynolds number of $Re \approx 500.000$ while the orifice flow has $Re \approx 1.000.000$. CDLBA is undoubtedly a major improvement for the calculation of flow fields in subsonic flow. In micrometer dimensions one may argue that turbulence does not play an important role. This is not true for all applications, especially when fast mixing devices are considered. In this application the onset of turbulence on short length scales is desired and the design of the respective structure can be supported by CDLBA, as it has already been done for the topology optimization of microflows (Geier *et al.* 2006b).

1.18.3 Conclusion

Whenever there is talk of automatic simulation tools, meaning that we wish to eliminate the need for a human operator, we deal either with highly linear and therefore deterministic decisions and computations or with a highly restricted application area and need to invest enormous effort to make the implementations robust.

Nevertheless, engineering design practice yearns for ever more automation as a means to become more efficient, especially in avoiding the tedious work of scanning for optimal design layouts for those cases where at least the target performance quantities are evaluable.

In our search for better tools to achieve these goals, we and others are working at cointegrating continuum and mesoscopic methods, and on extending algorithms to become more efficient and robust, and general enough to cope with the insatiable demand of MEMS engineers for design support tools in an ever-broadening application field.

Our choice for the immediate future has fallen on the level set method, treated as a coupled scalar field and tied to the rest of the physics, as a way to capture the engineering design topology. It provides a natural setting for a range of complex phenomena when performing topology optimization together with the finite element and related methods.

For soft matter physics, two particle approaches are followed: the DPD method and the digital lattice Boltzmann method. We are exploring ways to tie them together with conventional (continuum) methods, so as to achieve efficient multiscaling simulators, and we are embedding them in a topology optimization cycle, so as to bring these methods also into the design tool domain.

Acknowledgments

The authors wish to thank M. Geier, D. Kauzlaric, and L. Pastewka for their support in preparing the figures.

References

Allaire G, Jouv F, Toader A 2004 Structure optimization using sensitivity analysis and a level-set method. *J. Comput. Phys.* **194**, 363–93

Allgower E, Georg K 2003 *Introduction to Numerical Continuation Methods*. SIAM, Philadelphia, PA

Arora J S, Wang Q 2005 Review of formulations for structural and mechanical system optimization. *Struct. Multidisc. Optim.* **30**, 251–72

Ascher U M, Petzold L R 1998 *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equation*. SIAM, Philadelphia, PA

Beckett G, Mackenzie J A, Robertson M L 2001 A moving mesh finite element method for the solution of two-dimensional Stefan problems. *J. Comput. Phys.* **182**, 478–95

Belytschko T, Liu W K, Moran B 2000 *Nonlinear Finite Elements for Continua and Structures*. Wiley, New York

Bendsoe M P, Kikuchi N 1988 Generating optimal topologies in optimal design using a homogenization method. *Comp. Meth. Appl. Mech. Eng.* **71**, 197–224

Bendsoe M P, Sigmund O 2003 *Topology Optimization Theory, Methods and Applications*. Springer, Berlin

Bhatnagar P, Gross E, Krook M 1954 A model for collisional processes in gases I: Small amplitude processes in charged and neutral one-component system. *Phys. Rev.* **94**, 511–25

Boek E S, Coveney P V, Lekkerkerker H N W, van der Schoot P 1997 Simulating the rheology of dense colloidal suspensions using dissipative particle dynamics. *Phys. Rev. E* **55**, 3124–33

Boneyt Avalos J, Mackie A D 1997 Dissipative particle dynamics with energy conservation. *Europhys. Lett.* **40**, 141–6

Cao W, Huang W, Russell R D 2003 Approaches for generating moving adaptive meshes: Location versus velocity. *Appl. Numer. Math.* **47**, 121–38

Chen S, Doolen G 1998 Lattice Boltzmann method for fluid flows. *Annu. Rev. Fluid Mech.* **30**, 329–64

Costa J, Alves M K 2003 Layout optimization with H-adaptivity of structures. *Int. J. Numer. Meth. Eng.* **58**, 83–102

Courant R, Friedrichs K, Lewy H 1967 On the partial difference equations of mathematical physics. *IBM J. Res. Dev.* **11(2)**, 215–324

Coveney P V, Novik K E 1996 Computer simulations of domain growth and phase separation in two-dimensional immiscible fluids using dissipative particle dynamics. *Phys. Rev. E* **54**, 5134–41

d'Humieres D, Bouzidi M, Lallemand P 2001 Thirteen-velocity three-dimensional lattice Boltzmann model. *Phys. Rev. E* **63**, 066702

Du J, Olhoff N 2004 Topological optimization of continuum structures with design-dependent surface loading – Part 1: New computational approach for 2D problems. *Struct. Multidisc. Optim.* **27**, 151–65

Dwinzel W, Yuen D A 2000 Matching macroscopic properties of binary fluids to the interactions of dissipative particle dynamics. *Int. J. Mod. Phys. C* **11**, 1–25

Dwinzel W, Yuen D A, Boryczko K 2002 Mesoscopic modeling of colloids simulated with dissipative particle dynamics. *J. Model.* **8**, 33–43

Espanol P, Revenga M 2003 Smoothed dissipative particle dynamics. *Phys. Rev. E* **67**, 026705

Felippa C A 1999 Partitioned analysis of coupled mechanical systems. Report Cu-CAS-99-06

Flekkoy E G, Coveney P V 1999 From molecular dynamics to dissipative particle dynamics. *Phys. Rev. Lett.* **83**, 1775–8

Flekkoy E G, Wagner G, Feder J 2000 Hybrid the large model for combined particle and continuum dynamics. *Europhys. Lett.* **52**, 271–6

Geier M, Greiner A, Korvink J G 2006a Cascaded digital lattice Boltzmann automata for high Reynolds number flow. *Phys. Rev. E* **73(6)**, 066705

Geier M, Liu Z, Greiner A, Korvink J G 2006b Lattice Boltzmann based structural topology optimization of micro channels. *Second Int. Conf. Transport Phenomena in Micro and Nano devices*, Barga, Italy

Gockenbach M S 2002 *Partial Differential Equations: Analytical and Numerical Methods*. SIAM, Philadelphia, PA

Groot R D, Warren P B 1997 Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *J. Chem. Phys.* **107**, 4423–35

Hammer V B, Olhoff N 2000 Topology optimization of continuum structures subjected to pressure loading. *Struct. Multidisc. Optim.* **19**, 85–92

Hoogerbrugge P J, Koelman J M V A 1992 Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. *Europhys. Lett.* **19**, 155–60

http://www.ansys.com/

http://www.femlab.com

Jakobsen L A, Lund E, Møller H 2001 Shape sensitivity analysis of time dependent fluid-structure interaction problems using the ALE method. *Proc. 4th World Congr. Structural and Multidisciplinary Optimization*, Dalian, China

Kauzlaric D, Greiner A, Korvink J G 2003a An alternative simulation method for micro injection moulding. *Conference, Proceedings Polytronic 2003*, Montreux, Switzerland, pp. 79–85

Kauzlaric D, Greiner A, Korvink J G 2003b Modelling of microrheological behaviour of fluids with dissipative particle dynamics. *VDI Berichte 1803,* pp. 179–82

Kauzlaric D, Greiner A, Korivink J G 2004 Modelling micro-rheological effects in micro injection moulding with dissipative particle dynamics. *Tech. Proc. 2004 Nanotechnol. Conf. Trade Show*, Boston, MA, USA, vol. 2, pp. 454–7

Kauzlaric D, Greiner A, Korvink J G, Schulz M, Heldele R 2005 Modelling micro-PIM. In: Balites H, Brand O, Fedder G K, Hierold C, Korvink J G, and Tabata O (eds.) *Microengineering of Metals and Ceramics.* Wiley-VCH-Verlag, Weinheim, Germany, ISBN 3–527–31208–0

Koumoutsakos P 2005 Multiscale flow simulations using particles. *Annu. Rev. Fluid Mech.* **37**, 457–87

Liu Z, Korvink J G, Huang R 2005 Structure topology optimization: Fully coupled level set method via FEMLAB. *Struct. Multidisc. Optim.* **29**, 407–17

Morton K W, Mayers D F 1994 *Numerical Solution of Partial Differential Equations: An Introduction*. Cambridge University Press, Cambridge

Nathan A, Baltes H 1999 *Microtransducer CAD, Physical and Computational Aspects*. Springer, Berlin

Nielaba P, Mareschal M, and Ciccotti G (eds.) (2002) *Bridging Time Scales: Molecular Simulations for the Next Decade. Lecture Notes in Physics.* Springer, Berlin, Vol. 605

Nocedal J, Wright S J 1999 *Numerical Optimization*. Springer, Berlin

Osher S, Fedkiw R 2002 *Level Set Methods and Dynamic Implicit Surfaces*. Springer, Berlin

Paganobarraga I, Frenkel D 2001 Dissipative particle dynamics for interacting systems. *J. Chem. Phys.* **115**, 5015–26

Pastewka L, Kauzlaric D, Greiner A, Korvink J G 2006 Thermostat with local heat bath coupling for exact energy conservation in dissipative particle dynamics. *Phys. Rev. E* **73(3)**, 037701

Pool R, Bolhuis P G 2006 Can purely repulsive soft potentials predict micelle formation correctly. *Phys. Chem. Chem. Phys.* **8**, 941–8

Romanowicz B F 1998 *Methodology for the Modeling and Simulation of Microsystems*. Kluwer Academic Publishers, Dordrecht, The Netherlands

Rozvany G I N 2001 Aims, scope, methods, history and unified terminology of computer-aided topology optimization in structural mechanics. *Struct. Multidisc. Optim.* **21**, 90–108

Sapiro G 2001 *Geometric Partial Differential Equations and Image Analysis*. Cambridge University Press, Cambridge

Schulz S G, Kuhn H, Schmid G, Mund C, Venzmer J 2004 Phase behaviour of amphiphilic polymers: A dissipative particle dynamics study. *Colloid. Polym. Sci.* **283**, 284–90

Senturia S D 2000 *Microsystem Design*. Kluwer Academic Publishers, Dordrecht, The Netherlands

Sigmund O 2001a Design of multiphysics actuators using topology optimization – Part I: One-material structures. *Comput. Meth. Appl. Mech. Eng.* **190**, 6577–604

Sigmund O 2001b Design of multiphysics actuators using topology optimization – Part II: Two-material structures. *Comput. Meth. Appl. Mech. Eng.* **190**, 6605–27

Thompson J F, Warsi Z U A, Mastin C W 1985 *Numerical Grid Generation: Foundations and Applications*. North-Holland, New York

Wolf-Gladrow D A 2000 *Lattice Gas Cellular Automata and Lattice Boltzmann Models: An Introduction*. Springer, Berlin

Yin L, Ananthasuresh G K 2002 A novel topology design scheme for the multi-physics problems of electro-thermally actuated compliant micromechanisms. *Sens. Actuators A* **97**, 599–609

Yoon G H, Kim Y Y 2006 The element connectivity parameterization formulation for the topology design optimization of multiphysics systems. *Int. J. Numer. Meth. Eng.* **64**, 1649–77

Zienkiewicz O C, Taylor R, Zhu J 2005 *The Finite Element Method*. Elsevier, Amsterdam

Biographies

Dr. Jan G. Korvink obtained his M.Sc. in computational mechanics from the University of Cape Town in 1987, and his Ph.D. in applied computer science from the ETH Zurich in 1993. After his graduate studies, he joined the Physical Electronics Laboratory of the ETH Zurich, where he established and led the MEMS Modelling Group. This was followed by a move in 1997 to the Albert Ludwig University in Freiburg, Germany, where he holds a Chair position in microsystem technology and runs the Laboratory for Microsystem Simulation. He has written more that 180 journal and conference papers in the area of microsystem technology, and co-edits the review journal *Applied Micro and Nanosystems,* see http://www.wiley-vch.de/books/info/amn. His research interests cover the modelling and simulation and low cost fabrication of microsystems.

Andreas Greiner is a senior scientist and head of the group of physical modeling at the laboratory for Simulation of the Department of Microsystems Engineering (IMTEK), University of Freiburg, Germany. He graduated in physics from the University of Stuttgart, Germany, where he received his PhD in Theoretical Physics in 1992 and where he was concerned with the simulation of carrier transport in semiconductor lasers until 1995. From June 1995 on he held several grants as a research fellow at the University of Bologna, the University of Lecce and the CNRS research center for Microoptoelectronics in Montpellier. His research interest is the multiscale modeling and simulation of micro- and nanosystems as well as process simulation.

Zhenyu Liu obtained his PhD in mechanical engineering from the Dalian University of Technology, China in 2000. From 2001 to 2002, he worked as postdoc in the Department of Microsystems Engineering (IMTEK), University of Freiburg, Germany. From 2003 he joined IMTEK as a senior scientist and group leader at the laboratory for Simulation. His research interests cover the simulation and optimization of devices in MEMS.

1.19 Systematic Synthesis Methods

G. K. Ananthasuresh, Mechanical Engineering, Indian Institute of Science, Bangalore, India

Glossary

Analysis Evaluation of a given entity or system using physics-based modeling.

Calculus of variations A branch of mathematics dealing with minimization of functionals, which are mappings from a function space to real or complex number field.

Compliant mechanisms Mechanisms that use elastic deformation to transform or transmit given motion, forces, or energy.

Design variable A variable whose value is determined by the optimization algorithm to minimize the objective function in an optimal design problem.

Functional Functional is a mapping from a function space to real or complex number field.

Governing equation A mathematical equation, usually a partial differential equation, that described a physical law that governs the behavior of a system.

Inverse problem A problem in which the results are known and we need to find the data or model parameters that would give those results.

Karush–Kuhn–Tucker conditions The first-order necessary conditions for the constrained minimum of an optimization problem involving an objective function and equality and inequality constraints.

Parametric optimization An optimization problem in which a few parameters of a model, a system, or an entity are used as the design variables.

Shape optimization A structural optimization problem where the design variables determine the shape of the structure.
Simulation Modeling and solving for the behavior of a system or an entity using physics-based mathematical modeling.
State variable A variable that describes the state of a system. For example, the displacement of each point in an elastically deforming system is a state variable.
Stiffness measure A metric to quantify the stiffness of an elastic structure.
Strain energy Energy stored in a deformable body when it is subjected to forces.
Synthesis Creating an entity or a system for desired behavior.
Topology optimization A structural optimization problem where the design variables constitute a space of possible topologies of the structure.

1.19.1 Introduction

We begin this chapter by explaining how the term synthesis is interpreted here. This is followed by a brief discussion of the issues involved in the systematic synthesis of microsystems and their components.

1.19.1.1 What Is Synthesis?

The integration of different types of miniaturized components into a single system – sometimes as a single chip and sometimes in a package – is one of the key attributes of microsystems. The coming together of numerous entities of a wide variety at a close proximity in the micron scale ushers in new opportunities as well as challenging problems in simulation and design. The simulation of microsystems, owing to focused research and development efforts in the last decade of the 20th century, is well established today. There are quite a few new software packages that are exclusively devoted to this task and several other older ones that incorporated this as a new module. The number of research papers that deal with simulation of microsystems is rapidly growing. More importantly, most of today's papers in this field include some discussion on modeling and simulation unlike those in the 1980s and 1990s. The design of microsystems is less developed in comparison. Since the terms simulation and design are often misinterpreted – not only in the field of microsystems but also in others – some clarification is pertinent at the outset. One of the unambiguous ways to interpret these is to use their equivalent terms analysis and synthesis, respectively.

Analysis is an evaluation of something that is already there or is conceived anew. We use well-known laws, principles, and phenomenological models that govern the behavior of different phenomena, pose them as mathematical equations, and finally solve them either analytically or numerically. Thus, analysis or simulation is an alternative to experimenting with an actual entity. It of course helps in design but simulation is not the design in its entirety. Synthesis is the crux of design. Synthesis is the creation of something that is not already there. When it is conceived as a creative flash by an individual, it should be assumed that the process of synthesis took place in one's mind. The same is true when a new design is put forth based on the previous experience of an individual. In both the cases, simulation is very useful to check how well a newly conceived design performs before making its prototype. Most of the features of the existing microsystem software help immensely in this respect. They also provide the feedback or the insight to change the design to improve it. Perhaps this is the reason for the confusion between the terms simulation and design or their analogs, analysis and synthesis. Just as analysis can be systematically done using mathematical models of various phenomena, synthesis too can be done systematically using the same models. But the manner in which the problem is posed and the analytical or numerical methods that help solve it are significantly different for analysis and synthesis.

Synthesis can be considered as the reverse of analysis. While analysis (i.e., simulation) helps determine the resolution of the angular rate of a particular embodiment of a micromachined gyroscope, synthesis aims at something different: what systematic procedure or a direct algorithm, as opposed to a trial-and-error approach, helps determine the overall configuration and its parameters, dimensions, and materials that give the resolution that a designer wants? Systematic synthesis procedures alleviate or completely remove the tedium of repeated and incrementally improved designs that follow the trial-and-error approach. An additional advantage is that they

may also lead to designs that would otherwise not come about from intuition, prior experience, or creativity of human designers.

1.19.1.2 Significance, Opportunities, and Challenges

It is useful to understand the differences between synthesis-based and purely analysis-based approaches to design from a mathematical viewpoint. Since most physical, chemical, or biological phenomena involved in microsystems are governed by differential equations, let us consider a differential equation – an ordinary differential equation – to keep things simple and clear:

$$D(u(x), v(x)) + cu(x)v(x) = 0 \qquad [1]$$

where D is a differential operator, $u(x)$ and $v(x)$ are some functions of the independent variable x, and c is a constant. Let $u(x)$ denote the function that determines the design, and let $v(x)$ be the state variable on which a certain response of the system directly depends. The response is the quantity that a designer wants to optimize or have at a desired value. In analysis, eqn [1] is solved for $v(x)$ when $u(x)$ (the design) is known. This is the forward problem. But it may not be straightforward to solve it. Hence, in each physical or chemical domain, there are well-developed analytical and computational methods to solve the corresponding governing differential equation. In microsystems, there is an additional interesting feature that several differential equations are coupled together because the phenomena involved in them are physically or chemically coupled. Much of the computer-aided design (CAD) software that exists for microsystems today addresses this need: how can one efficiently determine the response of a microsystem component when the design is known? Since the design is not known, whenever a new task is taken up, the software programs expect the designer to guess a design or initiate one based on prior experience or intuition. With the help of the software, the designer computes the response and tries to adjust the design to match the desired performance by trial and error and using the insight gained from numerical experiments. Arguably, this is a tedious and time-consuming process. Furthermore, there is no guarantee that an optimum design is found.

In the synthesis-based approach, the guesswork would be circumvented or minimized. Here, when the designer knows what is needed (i.e., $v(x)$ is known), eqn [1] is solved for the unknown $u(x)$. Notice that it is simply the inverse problem wherein the roles of design variable ($u(x)$) and the state variable ($v(x)$) are interchanged. It must be noted, however, that the analytical or computational methods to solve this inverse problem are significantly different from those of the forward (analysis) problem. But, if we develop such methods, which we call synthesis methods, there is no need for a trial-and-error approach to design. This makes design computations efficient.

The synthesis-based direct approach to design not only provides computationally efficient means to obtain new designs that work but also reveals valuable insights into the problem. For example, a designer could be advised about what performance is reasonable to expect for the given specifications. That is, the limits of desired performance can be understood *a priori*. An example would help to appreciate this better. Let us consider the simplest elastic system of an axially deforming bar of a given area of cross section $A(x)$ and material's Young's modulus $E(x)$ under an axial load of $p(x)$ (**Figure 1**). Let its displacement under elastic deformation be $w(x)$. If we want to design this bar to be the stiffest (i.e., it deforms the least), $w(x)$ becomes the state variable deciding the desired performance and thus taking up the role of $v(x)$ in eqn [1]. Then, $A(x)$ is the design variable. For this system, the governing differential equation is given by

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(E(x)A(x)\frac{\mathrm{d}w(x)}{\mathrm{d}x}\right) + p(x) = 0 \qquad [2]$$

In which the state and design variables appear differently. This is almost always the case. Hence, the methods to solve the forward and inverse problems are significantly different. Furthermore, in the forward (analysis) problem $A(x)$ is completely known as a function because one would have the

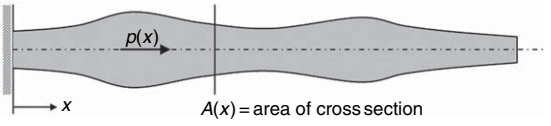


Figure 1 A simple example of an elastically deformable structure to illustrate the difference between analysis and synthesis. Is it sensible to desire displacement to the left if the force is applied to the right everywhere? The answer is no. However, this intuition proves wrong when a 2D elastic body is considered in a similar situation.

cross-section value at all points x. It is not so in the synthesis problem: we only know of or desire a certain level of response or performance or we are interested in its value for only particular values of x.

In this problem, if the stiffness is the desired performance, we know its value as given below:

$$\text{A measure of stiffness} = \int_{\Omega} p(x)w(x)\mathrm{d}\Omega \qquad [3]$$

where the symbol Ω denotes the spatial design domain of the bar that spans between 0 and L. That is, the function $w(x)$ is not known but the quantity that is controlled by it is known or it needs to be optimized for. In other situations, one may want to limit the deflection at the tip to a certain value. For example, one may stipulate in a certain application that it should move to the left (i.e., have a negative value). But if the load $p(x)$ is positive (i.e., pointing to the right) everywhere, it is not possible to achieve that with any physically meaningful $A(x)$. This is rather obvious in this simple problem. However, if this problem is considered in two dimensions (as in a planar sheet of some geometry with forces applied in one direction, say positive x, in the same plane), it is indeed possible to get the displacement in the negative x direction. So, if a trial-and-error approach is used, there is no easy way to know this in complicated design problems unless the designer has the required intuition about the problem. Similarly, one would also not know what value of the measure of stiffness (eqn [3]) is the best that can be achieved. Given that microsystems is a multidisciplinary field, it is often not practical to expect designers to have that intuition in all aspects of the problem. As opposed to this analysis-based iterative approach, when we solve the synthesis problem, often both the problems of the trial-based design iterations can be remedied to a large extent. The interested reader may refer to Ananthasuresh (2003, Section 1.2) for additional examples that explain the difference between analysis and synthesis. Much work is of course needed to gain insight into a given problem. But this will be the work done by synthesis researchers and it will not be a burden on the designers who use synthesis methods.

The above synthesis perspective toward designing the systems governed by differential equations offers unique opportunities in systematically creating designs that are not otherwise conceived by human designers. The designs obtained by the synthesis methods also aid the intuition of the designers. These points are discussed in the chapter after some of the synthesis methods are presented. First, we consider a general overview of the synthesis work done for microsystems.

1.19.1.3 Overview of the Synthesis Work on Microsystems

Many aspects of the synthesis of microsystems components and devices have been addressed in the literature. Among these, the maximum attention has been paid to mechanical synthesis wherein an elastically deforming body is designed. It is well known that most of the mechanical components of microsystems consist of elastically deforming bodies rather than rigid bodies assembled with joints. So, it is often necessary to design the geometrical form of an elastic body – be it a simple beam or a plate or a complicated 3D elastic body – so that it deforms in a desired manner under given mechanical forces. An example of such a problem is shown in **Figure 2**. The method to solve this type of problem first appeared in Ananthasuresh (1994), Ananthasuresh *et al.* (1994), Frecker *et al.* (1997), and Sigmund (1997). Comprehensive literature on this can be found in Ananthasuresh (2003) and Bendsoe and Sigmund (2003).

When forces are not just mechanical, which is often the case in microsystems, this basic method needs to be extended. For electrostatically actuated micromechanical structures, methods have been developed to design the geometry (Ye *et al.* 1998). This problem involves two coupled differential equations in the electrostatic and elastic energy domains. Similar coupled problems exist in electrothermally actuated structures (Mankame and Ananthasuresh

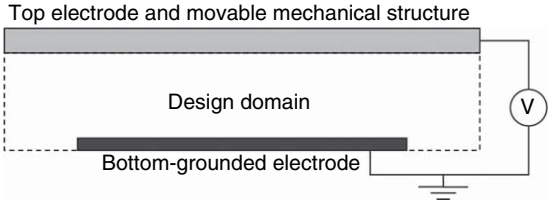


Figure 2 An example of a mechanical synthesis problem. What mechanical connection within the design domain indicated by dashed lines between the bottom grounded electrode and the top electrode would make the top movable electrode displace downward without any or minimal curving? This problem has practical relevance in micromachined polychromators. (Source: Hung E S, Senturia S D 1999 Extending the travel range of analog-tuned electrostatic actuators. *J. Microelectromech. Syst.* **8**, 497.)

2004, Moulton and Ananthasuresh 2001, Sigmund 2001a, b, Yin and Ananthasuresh 2002), magnetically actuated structures (Yoo and Soh 2005), acoustics (Bangtsson *et al.* 2003), and several others. In the case of electrothermally actuated structures, three different domains, viz., electric, thermal, and elastic domains, are coupled together.

When the material under consideration is *active*, it also leads to interesting synthesis problems. For example, with piezoelectric materials, one often wants to amplify the displacement. This can be done by mechanical displacement amplifiers that need to be synthesized for specific needs. This has received much attention in the literature (Ananthasuresh 2003, Silva and Kikuchi 1999, Silva *et al.* 2000). Since microsystems fabrication techniques make it possible to control the geometry at the microscopic level, it is even possible to design material microstructures such that extreme properties can be effectively achieved. These include negative Poisson ratios and negative thermal expansion coefficients as discussed in Sigmund (1994) and Sigmund and Torquato (1996). In fluids too, interesting problems are posed and solved to design flow paths (Klarbring *et al.* 2003). Attempts have also been made to synthesize microfabrication process flows and lithography mask layouts (Ananthasuresh 2003). Several other efforts are under development (Bendsoe and Sigmund 2003). While these focus on the generality of the synthesis procedures, several problem-specific synthesis methods have also been developed for microsystems. Among these, system-level parametric optimization methods (e.g., Mukherjee *et al.* 1998) are worth mentioning. In this chapter, a bird's-eye view of all these methods is presented by highlighting the underlying mathematical issues.

1.19.1.4 Organization of the Chapter

In Section 1.19.2, we first present the synthesis methods in general in addition to classifying them in a certain way. This is followed by a simple example that considers the parametric optimization. And then, the synthesis methods in different energy domains of microsystems are presented. In the subsequent section, geometric aspects of mask layouts and process synthesis are discussed. Closing remarks summarize the chapter and note opportunities and challenges for further work in this area.

1.19.2 Methods for Synthesis

General methods for synthesis are discussed in brief. They apply not only to microsystems synthesis problems but also to many other problems. We will also discuss how one can look at the synthesis problems in microsystems in a hierarchical way.

1.19.2.1 Direct vs. Indirect Synthesis

Referring back to eqn [1], in a direct synthesis method, when the desired response function is given (i.e., $v(x)$ is known), the differential equation is solved for the unknown function $u(x)$, the design. This direct method is often not possible for most differential equations that govern different physical and chemical phenomena. There are two aspects to it. First, we need to be concerned about the well-posedness of the problem and then the existence and uniqueness of the solution. The question of well-posedness arises because the desired response function might not satisfy the particular features of the solution of the forward (analysis) problem. The existence of a solution is not guaranteed because a given (or desired) response function may be physically impossible to achieve.

In some cases, it may be important to know if the solution is unique. If it is not, additional requirements to choose the best one need to be imposed. These issues are treated extensively in the inverse problems in literature wherein methods are developed to directly address the synthesis problem. On the other hand, in indirect synthesis, optimization methods are often used.

In optimization methods, we try to solve the problem to best meet the needs with respect to the desired response. For this, we need to formulate an optimization problem with an objective function, design variables and parameters, and constraints. The constraints relate to the second aspect of why direct methods of synthesis are not possible (or practically infeasible) for most problems. This is because there usually are limited resources – material, size and space constraints, manufacturability limitations, etc. – to achieve a desired response. So, even if a direct method of synthesis where the inverse differential equation can be solved is used, the constraints make the resulting solution unusable. So, the synthesis methods developed for microsystems rely on optimization theory and methods. These methods are general enough to consider microsystems synthesis at different levels of abstraction as explained in the next section.

1.19.2.2 Synthesis at Different Levels of Abstraction

Microsystems, as their name implies, are integrated systems composed of different subsystems and components as shown in **Figure 3**. A system such as a micromachined diagnostic system on a chip consists of several devices as subsystems including those for plumbing, sensing, signal conditioning and processing, digital output. Each of these devices consists of several components. For example, plumbing requires pumps, valves, fluidic channels, reservoirs, mixers, etc. The behavior of each of these is governed by different physical phenomena depending on the actuation method adopted. To practically realize each of these, there is a need for a fabrication process and a set of masks after the geometry and material are selected for each component through physical level simulations or design. So, synthesis (or for that matter, analysis too) can be viewed at different levels of abstraction.

The highest level is the system level. Here, as explained in Senturia (2001), one considers the whole system with reduced-order models for the individual subsystems. Reduced-order models are necessary to make the system-level representation simple and computationally manageable. Since microsystems consist of movable solids and liquids, they are essentially dynamic systems. So, the governing equations are often first- or second-order ordinary differential equations with time as the independent variable. Each subsystem is then represented as a simple mass-spring-damper model or its equivalent in nonmechanical domains. The parameters of the subsystems' reduced-order model depend on the detailed geometry and the properties of the material that comprise the subsystems. But we do not consider those details at this highest level of abstraction. So, synthesis at this level entails determination of the parameters of the reduced-order models based on the design requirements. The desired responses are also at the highest level of abstraction. For example, in the case of a pressure sensor, one would consider the range, sensitivity, and resolution of the sensor and the effects of changing environmental conditions on the sensor rather than the low-level responses such as the displacement of the diaphragm due to a certain value of the pressure.

The system architecture is also an important component of synthesis at this level. In the case of microsystems, it assumes great significance because many types of subsystems are tightly integrated. There is also the crucial question of integrating the mechanical and electronic elements at the chip level, package level, or not at all. Packaging considerations are also important. It is not easy to treat these issues formally at this point with regard to both analysis and synthesis. So, synthesizing the system architecture is often left to the designer's discretion, prior knowledge, and experience. When formal methods are developed for this, synthesis methods will surely follow. It is important to note, however, that system-level synthesis is certainly possible once the system architecture is decided. This is discussed further in Sections 1.19.2.3 and 1.19.3. The important point to note is that there exist two synthesis problems at the system level as shown in **Figure 4**. The first is to synthesize the architecture of the system, which at this time is largely dependent on designer's preferences, experience, intuition, overall fabrication and packaging considerations, marketing related needs, etc. After the architecture is decided and individual devices are identified, the second synthesis problem comes to the fore. The synthesis at this stage can be addressed by considering a dynamic model of the system with different subsystems (devices) represented as reduced-order models. As explained above, the characteristics of the reduced-order models need to be determined to meet the overall requirements regarding the performance of the system.

If once again the example of the micromachined diagnostic system on a chip is taken, a designer has

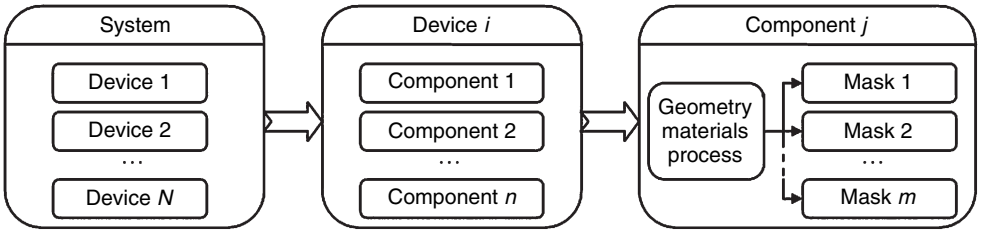


Figure 3 Hierarchical view of a microsystem.

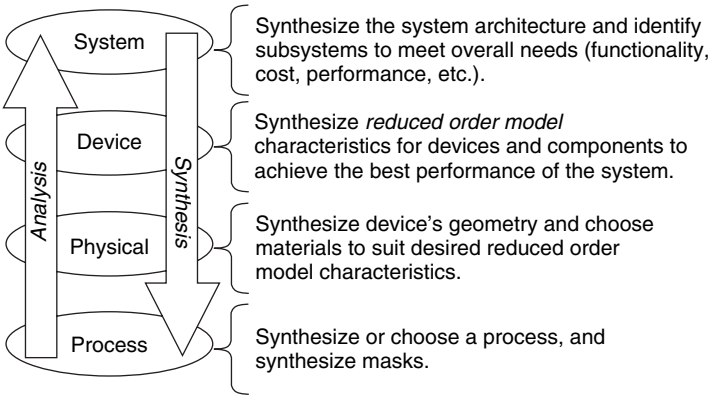


Figure 4 Synthesis problems at different levels of abstraction for a microsystem.

to first decide what devices to use and how each device would need to work. Several questions arise: How is the specimen collected? How is it pumped? How is it mixed with a reagent? What is the sensing mechanism? What are the requirements on the electronics that condition and process the sensor signal and create a quantifiable output, etc? The physical mechanisms of each of the devices that would address such needs require further consideration. Once those questions are answered, reduced-order models should be selected. If the example of a pump is taken, its configuration is to be decided first: should it be a diaphragm-based pump or something else? And then, a model and its parameters need to be identified to quantify flow rate, range and resolution of delivered volume, power consumed, etc. When a model is constructed for all the devices in this manner, synthesis can be done at the system level.

The first two levels of synthesis (see **Figure 4**) are quite difficult; systematic and automated procedures are yet to be developed for any general system. However, much progress has been made at the next two levels of abstraction. Synthesis at the third level (called physical level in **Figure 4**) has received considerable attention. Much of this chapter is concerned with this synthesis (see Sections 1.19.4.1–1.19.4.6). Here, one needs to determine the geometry and materials for each component in view of the requirements decided at the system level and a set of (usually) partial differential equations that govern the phenomena involved. After this, at the fourth level, one needs to synthesize (or choose) a process and lithography mask layouts. This level has received some attention, which is described in Section 1.19.5.

1.19.2.3 Synthesis Aided by Parametric Optimization

Optimization, as mentioned in Section 1.19.2.1, is a useful tool for synthesis. It should be noted, however, that whenever it is possible to solve the synthesis equations directly it should be done. If that is not possible, which is often the case, an optimization problem needs to be formulated. This should try to solve the synthesis equations indirectly in a least-squares sense. That is, the sum of the squares of the differences between desired and designed (or rather to be designed) quantities needs to be minimized. This minimization will be limited by resources and practical considerations. They become the constraints in the optimization problem. Another important aspect of formulating the optimization problem is to identify the design variables. This step is often called design parameterization in multidisciplinary design optimization literature. This can be done in several ways. Three methods are described here. The first is called the parametric optimization.

In parametric optimization, the component that is to be synthesized is almost determined, with the exception of a few parameters. These can be dimensions of certain parts of the components such as widths of beams, thicknesses of the diaphragms, location of supports. While it may appear simple – and it is simple as the variables involved are few – the problems posed in this manner may not always be well posed. That is, the optimization problem may not be smooth and it may not have well-characterized minima. But it is often done because the application of numerical optimization algorithms does give an improved solution. Such an improved design is often interpreted as equivalent to optimized design. For difficult problems,

this often may be the only viable remedy. For example, in an electrostatic comb actuator, if one decides to vary the number of interdigitated comb fingers, the number of beams in the suspension, and the number of fixed electrodes as design variables in the optimization problem, improved solutions are indeed obtained. This aspect has been extensively investigated (Iyer *et al.* 1998, Mukherjee *et al.* 1998, 1999) and summarized in Ananthasuresh (2003, Chap. 10).

A number of optimization algorithms are available for solving parametric optimization problems. Any standard text on optimization (e.g., Papalambros and Wilde 2000, Rao 2000) may be consulted. There are also several software implementations of the optimization algorithms. In parametric optimization, there may be continuous as well as discrete variables. This calls for special optimization techniques that can deal with this. Both deterministic and probabilistic methods may be used. Among the probabilistic methods, the use of genetic algorithms (Goldberg 1989) is on the rise. Computation time for synthesis is often a concern when probabilistic algorithms are used. It is also necessary to ensure that the solution given by an algorithm is indeed a minimum as characterized by Karush–Kuhn–Tucker conditions (Papalambros and Wilde 2000) for a constrained minimum.

1.19.2.4 Shape Optimization

Shape optimization is more general than parametric optimization. Here, the shapes of the components are varied by identifying suitable design variables. The shape of a component does play a significant role in many problems. For example, Ye and Mukherjee (1999) and Ye *et al.* (1998) have shown that by changing the shape of the comb fingers in an electrostatic comb drive actuator, different force–displacement characteristics can be obtained. The shapes of flow channels are of importance in microfluidic systems. The shapes of different elastically deforming parts are also crucial for micromachined grippers and other mechanical devices (e.g., Xu and Ananthasuresh 2003). The shapes of holes within a continuum matter in optimizing the dynamic performance of a component where the elastic and inertial effects compete with each other, as is the case with micromachined resonators.

The design parameterization for shape optimization can be done in many different ways. Splines, Bezier curves, and other smooth interpolant methods for shape optimization in terms of a few continuous variables are often good choices (Xu and Ananthasuresh 2003). Varying the coordinates of nodes in a finite element model is also done (Hetrick and Kota 1999) but it has its drawbacks as explained in Achtziger (2006). It must be noted that shape optimization assumes that the topology of the geometry of a component is already known. There are several methods for the optimization of even the topology. These are explained in the next subsection.

1.19.2.5 Topology Optimization

Microsystems components are multilayered structures made of multiple materials. Choosing the materials and determining the geometry of the components are the main tasks of synthesis at the physical level. As explained above, it can be done by parametric or shape optimization. But there is the fundamental task of choosing the most suitable topology. Consider a simple rectangular domain shown in **Figure 5**. The design domain is the space in which a component has to fit. Let there be three regions of interest at the boundaries of the design domain as indicated by R_1, R_2, and R_3. In structural situations, they will be the points of application of force, sensing ports, or mechanical anchors. In fluidic situations, they will be the entry and exit points for the fluid. In designing photonic band gap material structures, they can be the terminals of wave guides. What is the proper connectivity of these regions of interest? Clearly, as shown in **Figure 5(a)**, there are many possible ways to connect the regions even with a single intermediate point. There can be more than three such regions in different problems. More intermediate points would give a richer variety of

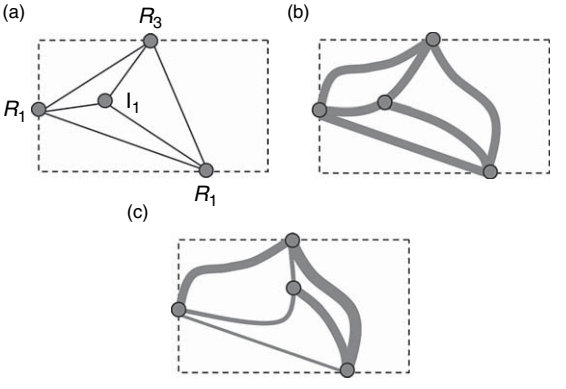


Figure 5 Topology, shape, and parametric optimization problems in schematic form. (a) Topology possibilities with one intermediate point; (b) a possible shape for the full topology; and (c) a final geometric form with different sizes including the shifting of the intermediate point.

topologies than what is shown in Figure 5. Then, the number of connections would increase dramatically and it would be impractical to explore all. This is especially true in view of the possibility that the shapes of these connections need to be explored as well. Furthermore, sizes (or parameters) of different segments of the connections also need to be determined. Thus, there are three levels of optimization. As shown in Figure 5(a)–5(c), there is topology optimization that determines the optimal connectivity, shape optimization that determines the shapes of individual connections, and then the size optimization that determines the sizes or the parameters. As the word topology entails, topology optimization also determines the number of holes in the design domain. It is a way of distributing a given amount of material within a given design domain.

Topology optimization is a powerful technique for determining the optimal distribution of material. It is a rapidly growing field with many applications (Bendsoe and Sigmund 2003). Not only the topologies but also the materials of different connections or parts of connections can be determined using topology optimization. In this method, the key idea is to parameterize the design domain so that general topological variations can be considered within the ambit of continuous optimization. Let us consider a design domain that is discretized into a finite element model – which is often the case as the phenomena governing a microsystem are modeled by partial differential equations for which the finite element analysis is arguably the most efficient numerical solution technique. Then, if we have material everywhere in the design domain, there is nothing to optimize. But then, there will always be a limited amount of material that is to be distributed in a larger domain than the material can fill. So, all elements in the finite element model cannot be filled with material because much material is not available. It then becomes a question of which ones to fill. If we denote the material state of each element by 0 (for empty) and 1 (for full), then we get a binary optimization problem. Because of the combinatorial explosion the binary parameterization gives rise to a large-scale – often computationally intractable – discrete optimization problem. On the other hand, if this material state is smoothly varied between 0 and 1, it gives rise to a continuous optimization problem. Gradient-based methods that solve such a problem are well developed and have been applied to topology optimization problems involving microsystems components. The details of topology optimization are best explained with specific examples in Section 1.19.4. We end this section by giving some general details of an optimal synthesis problem for topology optimization.

1.19.2.6 Optimal Synthesis Problem Statement

The optimal synthesis problem for topology optimization, like any other optimization problem, consists of the following essential features:

$$\begin{aligned} &\underset{d}{\text{Minimize}} \quad f(d, \mathbf{u}, \mathbf{p}) \\ &\text{Subject to} \\ &\qquad \text{Governing equations of static or dynamic equilibrium} \\ &\qquad \text{Resource constraints} \\ &\qquad \text{Performance constraints} \\ &\qquad \text{Practical constraints} \\ &\qquad \text{Bounds on the design variables: } d_l \le d \le d_u \end{aligned} \tag{4}$$

Here f is the objective function that needs to be minimized. When written analytically prior to finite element discretization of the model, it will be a functional for topology optimization problems. For shape and size optimization problems, it is most often simply a function of a finite number of design variables arranged in the form of a column vector $\mathbf{d}$. For topology optimization, the design variable $d(\mathbf{x})$ is a continuous spatial function defined at every point $\mathbf{x}$ in the design domain. This is the reason for the objective function being a functional rather than a function. The functional is most often in the form of an integral in terms of three variables indicated in eqn [4]: design variable d, state variable(s) $\mathbf{u}$, and some parameters $\mathbf{p}$. The number of state variables depends on the number of physical phenomena involved in the problem. In microsystems, as is well known, more than one phenomenon are involved. Thus, there will be more than one state variables. In a mechanical problem of designing structures and compliant (elastic) mechanisms (Ananthasuresh and Kota 1995, Howell 2001), the state variable can be the displacement, the stress, or the strain. These have to obey the static or dynamic equilibrium constraints. If the electrostatics problem is considered, charge and voltage will be the state variables. In a thermal problem, heat flux and temperature are the state variables. For each state variable, there must be a corresponding governing equation. These governing equations are included in the

optimization problem statement because the design variables directly influence the state variables, which in turn influence the objective function and the constraints. The parameters **p** are sometimes what could be the design variables for optimization and sometimes fixed constants that cannot be changed by the designer.

The resource constraints usually refer to the amount of material or space available. Performance constraints include all considerations that are important for a problem, but they are not included in the objective function. For example, in a microactuator design problem, if the objective is to maximize the stroke then one may include performance constraints on delivered force, power consumed, resolution in stroke and force, and so on. Practical constraints include manufacturability concerns and economical considerations. These are more difficult than others. One way to include them is discussed in Section 1.19.5.3. Finally, there will be lower and upper bounds on the design variables, as indicated by d_l and d_u. This problem, once posed properly, can be solved in many ways.

All numerical optimization methods need an initial guess using numerical optimization algorithms which find the optimum iteratively. This process can be loosely compared with the analysis-based iterations that a designer would do. But there is a significant difference that iterations in systematic optimal synthesis are directed by the gradient information at that instant. In trial-and-error-based iteration, it is driven by the designer's understanding of the problem.

The general problem statement shown in eqn [4] will be made specific for a few types of optimal microsystems synthesis problems discussed in Section 1.19.4. Before that, parametric optimization at the system level is described in the next section.

1.19.3 System-Level Synthesis

As noted earlier in the chapter, for most microsystems, there are two synthesis problems at the system level. The first synthesis problem of determining the system architecture is not amenable for systematic synthesis at present. The second synthesis problem of determining the characteristics of the reduced-order models of the subsystems is very possible. Here, it is illustrated with an example.

1.19.3.1 Accelerometer Example

Let us consider one of the most successful microsystems products: an accelerometer. Assume that a designer has decided to use capacitance-based sensing and feedback stabilization technique. With the performance requirements (including resolution and range of the accelerometer) in mind, if the designer wants to see whether to integrate the mechanical and electronics at the chip-level or package-level, it will be beneficial to use a schematic such as the one shown in **Figure 6**. The same schematic is helpful for system-level synthesis too. **Figure 6** shows different subsystems: mechanical subsystem consisting of an inertial mass, a suspension to hold and guide the mass, electrostatic sensing comb array, electrostatic comb array for applying the feedback voltage, damping model, etc.; electronic circuits for capacitance extraction; PID controller; and feedback voltage generation. In order to be realistic, the schematic also includes mechanical and circuit noise sources. In order to quantify the different blocks, we need a reduced-order model or an equivalent electrical circuit for all the subsystems. Once they are formulated, the performance of the system can be simulated. **Figure 7** shows a Simulink (www.mathworks.com) model of this system.

The Simulink model shows the mechanical subsystem using three parameters **K**, **M**, and **D**, which indicate the stiffness, inertia, and damping characteristics. These would be derived from the eigenmode analysis of the mechanical structure including the inertial mass, the suspension, and combs. These appear as gains in the block diagram. Only a few modes in the relevant range of frequencies are included here. This reduced-order model is based on eigenmode shapes (Ananthasuresh *et al.* 1996). The electronic circuits are also represented in **Figure 7** in the Simulink block diagram format. The model characteristics are adjustable in this model. For example, the gains in the PID controller can be tuned until the desired resolution is obtained

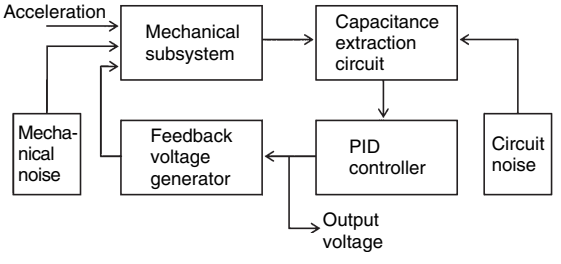


Figure 6 Block diagram of the system-level view of a feedback-stabilized micromachined accelerometer.

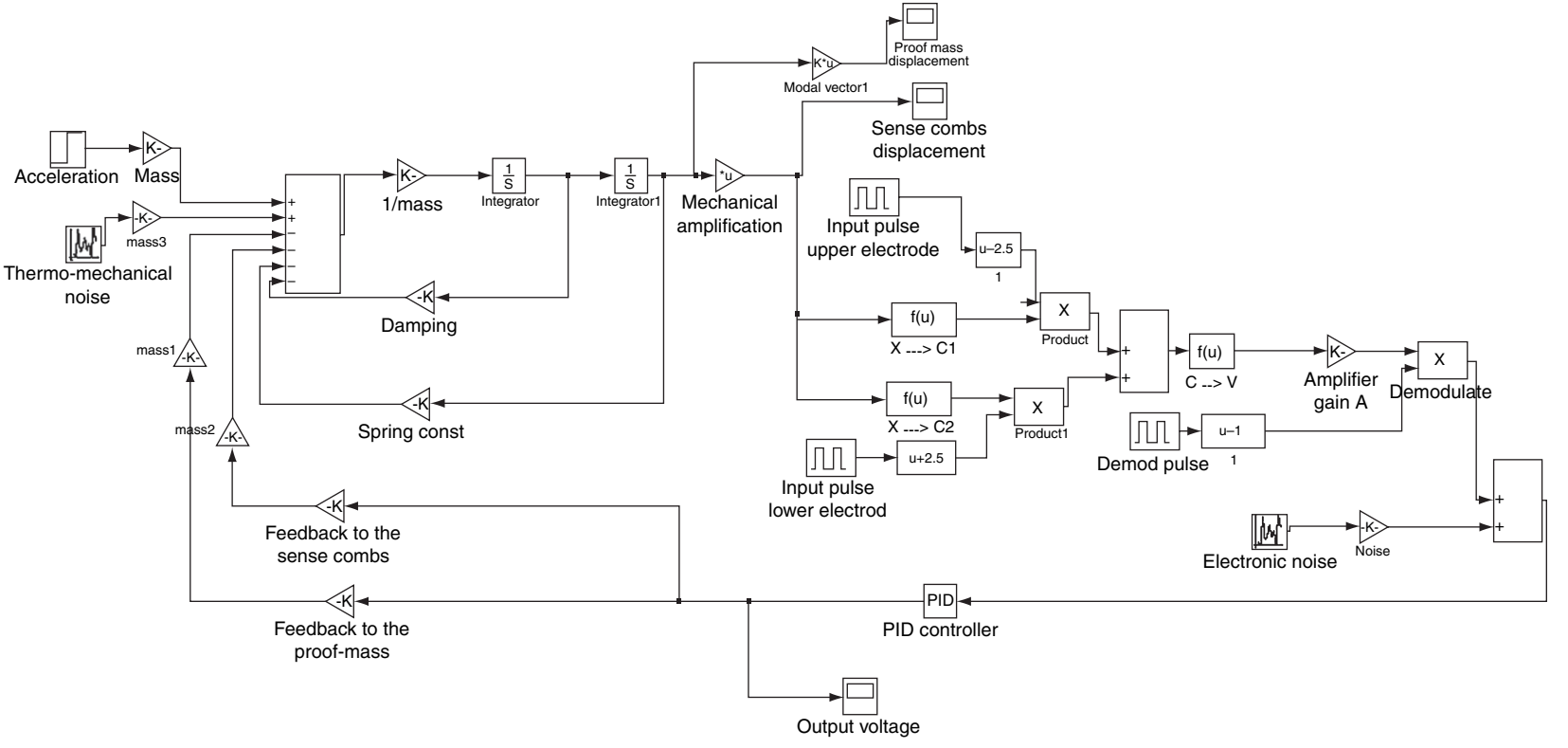


Figure 7 Simulink representation of the block diagram in **Figure 6**.

in the acceleration that is to be sensed. When this process is automated by posing it as a constrained optimization problem, it paves the way for systematic parametric optimal synthesis. The optimization procedure will determine all the parameters in the reduced-order models to meet the desired performance and other constraints.

A slightly different approach is sometimes followed wherein commonly used elements of microsystems are saved in a library. The designer can pick different instances of those elements from the library to create a system such as the one shown in **Figures 6** and **7**. The characteristics of the instantiated elements are adjusted and the system's response is assessed. This is iterated manually to obtain improved design. This approach is described in Ananthasuresh (2003, Chap. 10), Iyer *et al.* (1998), and Mukherjee *et al.* (1999). Some commercial microsystem design software programs follow this approach.

After the subsystems are synthesized (using optimization or not), the next step is to obtain the geometry of individual components that make the subsystems. For example, what physical shape of the inertial mass, suspension, and combs will give the finalized stiffness, inertia, and damping values in the system-level model? This task is described next for a few types of microsystems.

1.19.4 Synthesis of Components of Microsystems

Until now, most of the successful and well-characterized microsystems products consist of simple mechanical structural elements, viz., beams, plates, and membranes often arranged in topologically simple configurations. However, there are instances when simple topologies conceived intuitively are not adequate to meet the specifications. This is especially true when multiple physical phenomena are involved. We shall see how optimal topology synthesis can help in those situations.

1.19.4.1 Mechanical Structures and Transmissions

Movable mechanical structures are the crucial components of microsystems. They help distinguish microsystems chips from purely microelectronic integrated circuit chips. The mechanical components can be broadly classified into two groups based on whether they are intended to be movable or simply bear the loads without significant deformation. The former class, which encompasses most of the microsystems components, can be further grouped in two ways: one in which jointed rigid bodies are constrained to move in a certain way because of joints such as hinges and sliders and the other where elastic deformation of flexible bodies gives the intended motion. Systematic topology synthesis techniques are available for both types. The first type is based purely on kinematic approaches. A recent review of the state of the art in this area can be found in Antonsson and Cagan (2001, Chap. 9). The second type includes elastic mechanics. In this case, one or more of the following characteristics may be of interest:

- Stiffness
- Flexibility
- Strength
- Tracing a path or moving in a certain way
- Energy storage and absorption
- Resonance frequency
- Mode shapes
- Transient response

A microsystem designer may be interested in the overall stiffness of a mechanical component or in the stiffness for a portion of a component. In either case, this requirement can be posed as an objective function. A quantifiable criterion in this case is the strain energy stored in the elastically deformed structure. The lower the strain energy, the stiffer the structure. Hence, for maximizing stiffness, the strain energy should be minimized. The strain energy, SE, of an elastic body is given by the following equation:

$$\mathrm{SE} = \int_{\Omega} \frac{1}{2} \boldsymbol{\varepsilon}^{\mathrm{T}} \mathbf{D} \boldsymbol{\varepsilon} \, \mathrm{d}\Omega \qquad [5]$$

where $\boldsymbol{\varepsilon} = \left\{ \varepsilon_{xx} \, \varepsilon_{yy} \, \varepsilon_{zz} \, \varepsilon_{xy} \, \varepsilon_{yz} \, \varepsilon_{zx} \right\}^{\mathrm{T}}$ is the linear strain with

$$\varepsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial j} + \frac{\partial u_j}{\partial i} \right)$$

wherein the labels i and j take on x, y, and z, and $\mathbf{u} = \left\{ u_x \, u_y \, u_z \right\}^{\mathrm{T}}$ is the displacement vector. The design domain occupied by the elastic body is denoted by Ω. The symbol $\mathbf{D}$ denotes the constitutive material property relationship, which relates stress and strain. As implied in eqn [4], the objective function depends on the state variable $\mathbf{u}$ and the

design variable **D**. In order to see how the design variable appears in **D**, it is necessary to consider how topology optimization techniques continuously parameterize the design space.

In a given problem, any point in the design domain is a potential point to be occupied by the material. Let $\rho(\mathbf{x})$ be the function that decides whether the point located at $\mathbf{x}$ (in some chosen coordinate system) is occupied by the material. Let $\rho(\mathbf{x})$ take one of the two values: 1 or 0. If $\rho(\mathbf{x})$ has a value of 1 at $\mathbf{x}$ then we say that there is material at that point, and if it has a value of 0, then there is no material. The binary function $\rho(\mathbf{x})$ thus becomes the optimization variable. An optimization algorithm decides its value at each point so that an objective is minimized subject to some constraints. But a binary function is not well suited for continuous gradient-based numerical optimization. So, it is relaxed to a continuous function $\rho(\mathbf{x})$, which is bounded by 0 and 1 from below and above, respectively. As an astute reader will immediately realize, in the optimized solution, this leaves scope for some intermediate material state when $\rho(\mathbf{x})$ takes a value between 0 and 1. There are several schemes to prevent this situation of which the power law model (also known as simple isotropic material with penalty – SIMP, Bendsoe and Sigmund 1999) is widely used. In this scheme each material property is multiplied by $\rho(\mathbf{x})^{\eta}$ where η is the penalty parameter, which helps to push $\rho(\mathbf{x})$ to its bounds 0 or 1. Therefore, **D** is given by the following equation

$$\mathbf{D} = \mathbf{D}_0 \rho(\mathbf{x})^{\eta} \qquad [6]$$

where $\mathbf{D}_0$ is the material property with $\rho(\mathbf{x})$ as the design variable.

As per the problem statement in eqn [4], the governing equilibrium equation for the state variable **u** needs to be identified. If the considered problem involves only static loads, the partial differential equation with boundary conditions is given by the following equation:

$$\begin{aligned} &\nabla \cdot (\mathbf{D}\boldsymbol{\varepsilon}) + \mathbf{b} = 0 \quad \text{for}\, \Omega \\ &\sum_j (\mathbf{D}\boldsymbol{\varepsilon})_{ij}\, n_j = t_i \ \text{ on } \partial\Omega_{\mathrm{t}} \ \text{ and } \ \mathbf{u} = \mathbf{u}_0 \ \text{ on } \partial\Omega_{\mathrm{u}} \end{aligned} \qquad [7]$$

where $\mathbf{b} = \left\{ b_x\ b_y\ b_z \right\}^{\mathrm{T}}$ and $\mathbf{t} = \left\{ t_x\ t_y\ t_z \right\}^{\mathrm{T}}$ are body and surface traction forces, respectively, acting in the interior Ω and a portion of the boundary $\partial\Omega_{\mathrm{t}}$, $\mathbf{n} = \left\{ n_x\ n_y\ n_z \right\}^{\mathrm{T}}$ is the unit normal to the boundary, and $\mathbf{u}_0$ is the prescribed displacement over another portion of the boundary $\partial\Omega_{\mathrm{u}}$. The next in the list of constraints in eqn [4] is the resource constraint, which in this case is the upper bound on the volume of material used. This constraint is necessary in this problem because, without it, the entire design domain would be filled with material to give the stiffest topology. In the absence of performance-related and practical constraints, the optimal synthesis problem can be stated as follows:

$$\begin{aligned} &\text{Minimize}_{\ \rho(\mathbf{x})} \quad \mathrm{SE} = \int_{\Omega} \frac{1}{2} \boldsymbol{\varepsilon}^{\mathrm{T}} \mathbf{D} \boldsymbol{\varepsilon}\, \mathrm{d}\Omega \\ &\text{Subject to} \\ &\quad \nabla \cdot (\mathbf{D}\boldsymbol{\varepsilon}) + \mathbf{b} = 0 \,\text{for}\, \Omega \\ &\quad \sum_j (\mathbf{D}\boldsymbol{\varepsilon})_{ij}\, n_j = t_i \,\text{on}\, \partial\Omega_{\mathrm{t}} \,\text{and}\, \mathbf{u} = \mathbf{u}_0 \,\text{on}\, \partial\Omega_{\mathrm{u}} \\ &\quad \int_{\Omega} \rho \mathrm{d}\Omega - V_{\mathrm{u}} \le 0 \\ &\quad 0 \le \rho \le 1 \\ &\text{Given that } \mathbf{D} = \mathbf{D}_0 \rho^{\eta} \end{aligned} \qquad [8]$$

The methods for solving the above calculus of variations problem are discussed in books on structural optimization (Hafka and Gurdal 1989, Weinstock 1974). One popular and computationally attractive method is the optimality criteria method. In this method, the first order necessary condition of the above problem is analytically written in the following form:

$$A + \Lambda B = 0 \qquad [9]$$

where A and B are in terms of the state variable in its equilibrium configuration and is the Lagrange multiplier corresponding to the resource constraint. Then, heuristically (Bendsoe and Kukuchi 1988), iterative update formula for the design variables is set up as follows:

$$\rho^{(k+1)} = \rho^{(k)} + (A + \Lambda B)^{(k)} \qquad [10]$$

where k is the iteration number. As it can be noticed, when the iterative process converges, i.e., $A + \Lambda B = 0$, the necessary condition of eqn [9] is automatically satisfied. In practical implementation, a number of modifications and enhancements need to be made to eqn [10] and some of these are specific to the problem under consideration. Interested readers can refer to the literature (Bendsoe and Kukuchi 1988, Bendsoe and Sigmund 2003, Sigmund 2001c).

Alternatively, mathematical programming methods can be used (Hafka and Gurdal 1989, Papalambros and Wilde 2000, Rao 2000,). For this,

the problem stated in continuous form in eqn [8] can be discretized as follows:

$$\begin{aligned} &\underset{\rho}{\text{Minimize}}\ \frac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{K}\mathbf{u}^{\mathrm{T}} \\ &\text{Subject to} \\ &\quad \mathbf{K}\mathbf{u} = \mathbf{f} \\ &\quad \sum_i \rho_i v_i - V_{\mathrm{u}} \le 0 \\ &\quad 0 \le \rho_i \le 1 \end{aligned} \qquad [11]$$

where $\mathbf{u}$ and $\mathbf{f}$ are the displacement and force vectors, respectively, and $\mathbf{K}$ is the stiffness matrix. This problem posed in discrete variables, which are equal in number to the number of elements in the discretized design domain, can be solved using gradient-based numerical optimization algorithms. Details can be found in engineering optimization books (e.g., Papalambros and Wilde 2000, Rao 2000).

Figure 8 shows an example problem and its solution. As can be seen in **Figure 8(a)**, the input required from the designer is minimal: where the structure could be anchored and where the forces are applied. The obtained optimal topology, which shows black for the regions of the domain with $\rho = 1$ and white for regions with $\rho = 0$, is seen in **Figure 8(b)**. Using this, a smooth geometric form shown in **Figure 8(c)** can be obtained. The ability to give the optimal geometry with nominal designer input is the highlight of optimal topology synthesis method.

In compliant mechanisms, which are elastically deformable structures used to transmit force and motion (Ananthasuresh 1994, Ananthasuresh and Kota 1995, Howell 2001), the objective function and constraints are different. For example, if we are interested in maximum displacement at one point in the design domain when force is applied at another point, the objective function becomes the negative of the displacement at the desired point. The volume constraint can be retained as in eqn [8]. A hypothetical example is shown in **Figure 9**. The problem specification is shown in **Figure 9(a)**. This specification is nonintuitive because it is hard for a human designer to imagine what material connectivity would result in the displacement in the desired direction when the force is applied as shown. The optimal topology shown in **Figure 9(b)** provides what is desired but its geometric form is slightly complicated at first sight. However, when it is interpreted as a continuum (filled portion in **Figure 9(c)**), the flexible joint regions and relatively rigid regions can be clearly seen. Then, by imagining rigid bodies joined with hinges at the flexible joint regions, we can see that it is indeed two four-

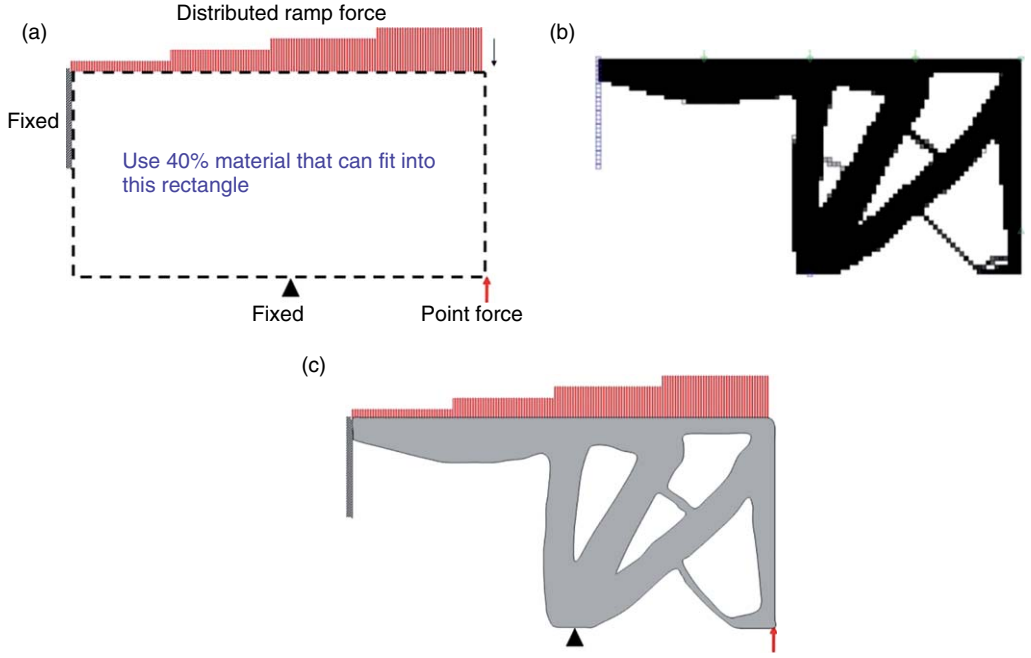


Figure 8 Optimal topology synthesis of the stiffest structure under given static loads for 40% volume constraint. (a) Problem specifications; (b) topology solution; and (c) interpreted final geometric form.

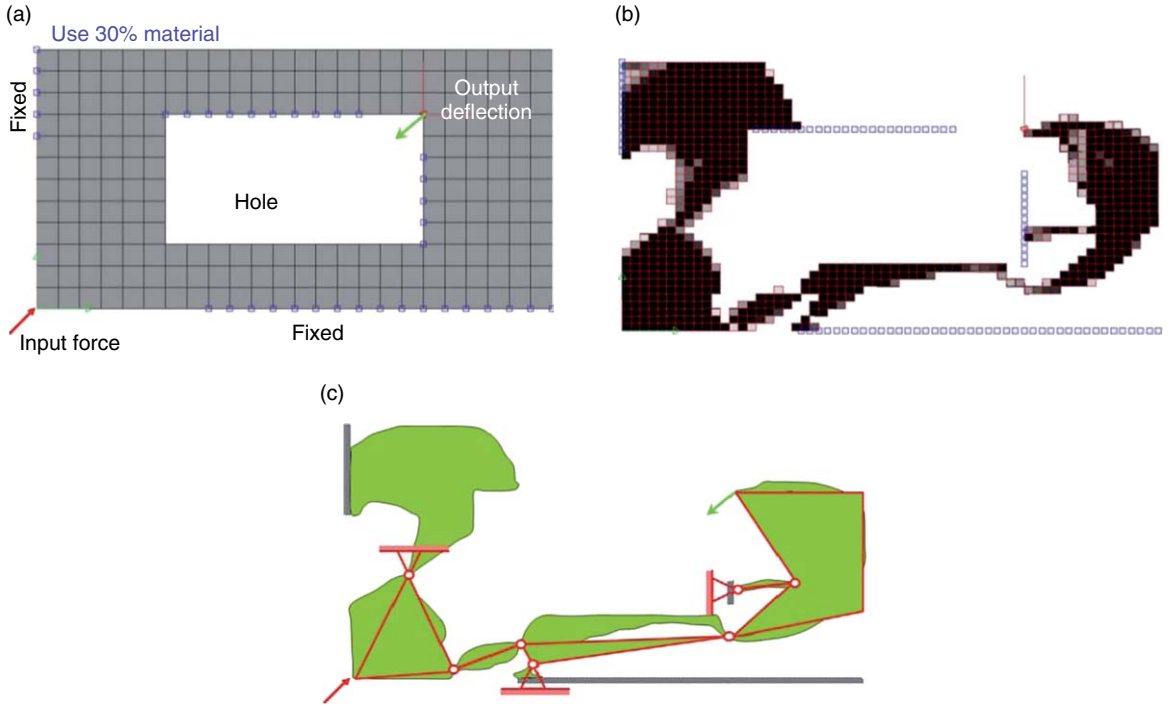


Figure 9 Optimal topology synthesis of a compliant mechanism. (a) Problem specifications; (b) optimal topology; and (c) interpretations of the topology with a smoothened continuum and rigid-body linkage.

bar linkages in tandem. Thus, optimal topology synthesis methodology can aid the designers in suggesting novel topologies that meet the given specifications.

The above two examples were solved using a Fortran program, which includes its own finite element routine and optimization algorithm (PennSyn: www.mecheng.iisc.ernet.in/~suresh/PennSyn). In a similar manner, strength considerations (Duysinx and Bendsoe 1998, Saxena and Ananthasuresh 2001a), prescribed curved paths (Pedersen *et al.* 2001, Saxena and Ananthasuresh 2001b), energy storage to modify the force–deflection characteristic of electrostatic microactuators (Pedersen *et al.* 2006), natural frequencies (Diaz and Kikuchi 1992) and mode shapes (Lai and Ananthasuresh 2002), and transient response (Ma *et al.* 1995) can be optimized. Two microsystem-related examples can be found in Ananthasuresh (2003, Chap. 2).

Arguably, the ability to generate topologies for the mechanical loads has matured much more than the current needs in the microsystems field. As noted earlier, microsystems components are usually very simple beams and diaphragms. But as microdevices get mechanically more complex, we can expect to see the usefulness of the optimal topology synthesis procedures. Next, we consider some coupled domains where similar topology synthesis procedures can be applied.

1.19.4.2 Electrostatically Actuated Structures

Historically, modeling and CAD of microsystems began with the self-consistent solution of coupled electrostatic and elastostatic equations (Senturia *et al.* 1992). In many ways, it is a new problem that arose exclusively in the microscale domain. It is also a practically significant problem as electrostatic actuation is arguably the most popular actuation used in microsystems today. It took significant focused effort to develop computationally efficient numerical techniques and software to solve any general problem of coupled electrostatic and elastostatic simulation problems. The main challenge here was that the two domains are tightly coupled: electrostatic force depends on the geometry of the conductor (or dielectric) parts of the component but the geometry changes under the influence of the electrostatic force. It is best seen with the help of the governing equations of the two domains, as shown in **Figure 10**.

The governing equation for the electrostatics is represented using an integral equation in **Figure 10**.

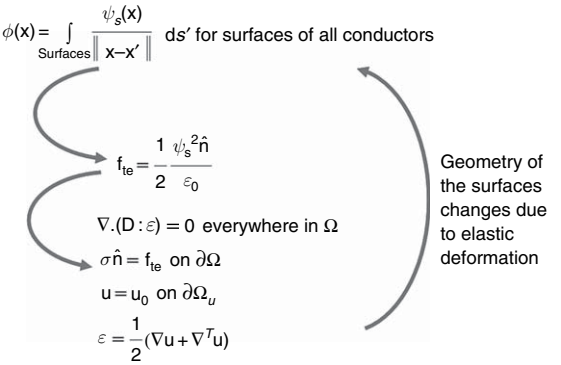


Figure 10 Coupling between electrostatic and elastostatic domains.

Its differential equation form is given by the following equation:

$$\nabla^2 \phi = 0 \qquad [12]$$

where ϕ is the electric potential (i.e., voltage) specified on the surfaces of all the conductors, but it needs to be determined in the intervening medium between the conductors. The surface charge density on all the conductors can then be found using the following expression:

$$\psi_s = \varepsilon_0 \mathbf{n} \cdot \mathbf{E} = -\varepsilon \mathbf{n} \cdot \nabla \phi \qquad [13]$$

where ε_0 is permittivity, $\mathbf{n}$ is the unit surface normal, and $\mathbf{E}$ is the electric field vector. The quantities in the elastostatic equation were explained earlier (see eqn [7]). Since the solution of one governing equation depends on that of the other, the two equations are to be solved simultaneously. Good techniques to do it are now available (e.g., Senturia *et al.* 1997). Now, let us consider the synthesis problem: how is the geometry of the conductors obtained so that they deform as desired when different electric potentials are applied on them?

The notable work in the electrostatic synthesis problem is in Ye *et al.* (1998), and is also described in detail in Ananthasuresh (2003, Chap. 3). This work focused on the shape optimization, in particular, on optimizing the shapes of the interdigitated comb fingers in an electrostatic comb linear actuator. It was shown that the nonlinearity of the stiffness of the suspension of the comb drive can be compensated by curved (rather than the usually straight) comb fingers. Linear, quadratic, and cubic force displacement characteristics were demonstrated.

Interestingly, the topology optimization of this coupled problem had not received attention until recently (Alwan and Ananthasuresh 2006b, Raulli and Maute 2005). An interesting challenge is encountered in doing so because the electrostatic force is a boundary force and the boundaries are not known and will be changing during the process of topology optimization. In Raulli (2006) and Raulli and Maute (2005), a method to track the boundaries was used to impose the forces on the boundaries properly. In Alwan and Ananthasuresh (2006a, b), a novel method that smoothly interpolates the material state from a conduct or to dielectric to void (empty space) was proposed. In this method, a leaky capacitor model and Maxwell's stress tensor were used to represent the surface electrostatic force as a body force (Alwan and Ananthasuresh 2006). This enabled this coupled problem to be cast in the standard form (see eqn [4]) for topology optimization. An example is shown in **Figure 11** for the simplest problem of designing the stiffest structure under electrostatic forces. Here, the bottom electrode was prescribed with a gap above to prevent a short circuit. The mechanical structure's domain along with the location of the mechanical anchor are given. If the structure is too close to the electrode, then there will be a large electrostatic force. On the other hand, if it is too far from the anchoring point, the structure becomes more flexible. So, a trade-off is needed between the two. As shown in **Figure 11**, it takes an arch shape, which is the stiffest for this specification. More work is under way in extending this methodology to other problems including that indicated in **Figure 2**.

1.19.4.3 Electro-thermal-compliant Actuators

The electro-thermal-compliant (ETC) microactuators, as their name implies, are governed by three energy domains: electrical, thermal, and elastic. These actuators, first proposed by Comtois and Bright (1996) and Guckel *et al.* (1992), have been generalized as versatile in-plane actuators in Moulton and Ananthasuresh (2001). The corresponding synthesis problem has been investigated in Mankame and Ananthasuresh (2004), Sigmund (2001a, b), and Yin and Ananthasuresh (2002). Even though it involves three coupled domains, this problem is simpler than the coupled electrostatic–elastostatic synthesis problem. This is because, if the temperature dependence of elastic and thermoelastic properties is neglected, the coupling in this problem is not as strong as it is in the electrostatic case. The coupling involved in this problem is explained below.

Let us consider the statement of topology synthesis problem of ETC actuators as shown in **Figure 12**. The

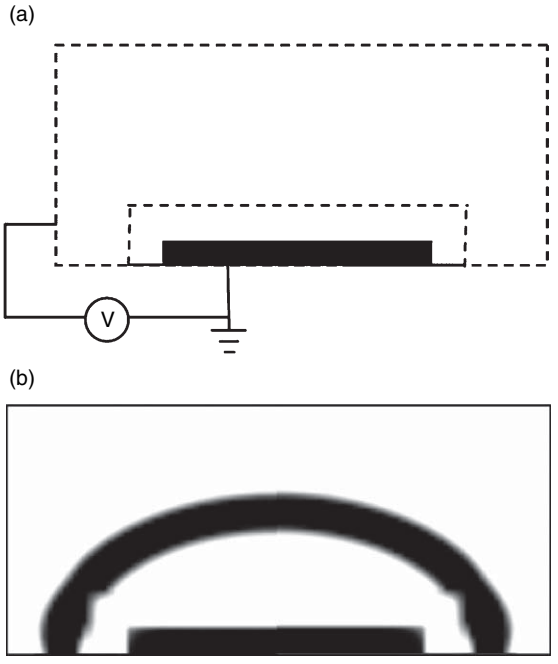


Figure 11 An example of the stiffest structure under electrostatic load. (a) Problem specification; and (b) optimal solution. The objective is to minimize the work done by the electrostatic force because it is equivalent to making the structure the stiffest. The design domain is shown by dashed lines. If the structure is too close to the bottom electrode, the force will be too large. If it is too far from the anchors (and hence from the bottom electrode), it will have a large deformation for a given force. So, the arch is a compromise between the two.

given design domain now has three sets of boundary conditions. First, there is a portion $\partial\Omega_{eg}$, which is electrically grounded, and another $\partial\Omega_{eV}$ where a potential V is applied. Second, there is a portion $\partial\Omega_{tg}$, which is thermally grounded, and a portion $\partial\Omega_{tf}$ where heat flux is specified. The heat flux can be incoming heat or outgoing heat through convection and radiation. Third, mechanical anchoring is done on $\partial\Omega_{mg}$ with external traction force on $\partial\Omega_{mt}$. There is also another portion $\partial\Omega_{md}$ where the displacement is desired in a certain way. The objective of the synthesis problem is to obtain a particular connectivity of the above different regions, perhaps with interior holes, so that upon application of the voltage, the elastic solid deforms to give the desired displacement at $\partial\Omega_{md}$.

For any geometry of the material distribution in Ω of **Figure 12**, first we need to solve the electrical current problem governed by the following equation:

$$\begin{aligned} &\nabla\cdot(k_e\nabla V)=0 \text{ in } \Omega \\ &V=0 \text{ on } \partial\Omega_{eg} \text{ and } V=V_0 \text{ on } \partial\Omega_{eV} \end{aligned} \quad [14]$$

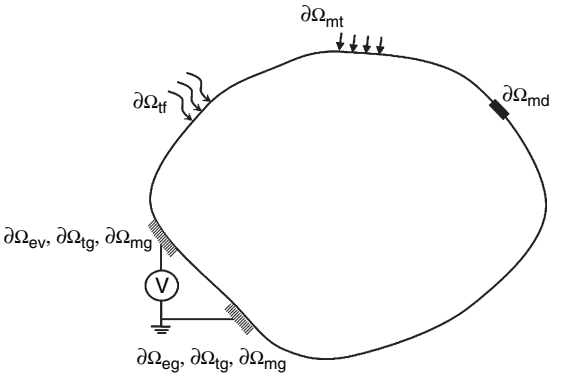


Figure 12 Synthesis problem specification for the electro-thermal-compliant (ETC) microactuator. The synthesis task entails the determination of material connectivity and its geometry within the design domain such that desired output displacement is generated at $\partial\Omega_{md}$ with V applied.

where k_e is the electrical conductivity and voltage V is the state variable. After this equation is solved for V, the Joule heat that is generated goes as a source term in the governing equations of thermal problem as shown below.

$$\begin{aligned} &\nabla\cdot(k_t\nabla T)+\nabla^T k_e\nabla V=0 \text{ in } \Omega \\ &T=T_0 \text{ on } \partial\Omega_{tg} \text{ and } \mathbf{n}\cdot(k_t\nabla T)=f_{nt} \text{ on } \partial\Omega_{tf} \end{aligned} \quad [15]$$

where k_t is the thermal conductivity, $\mathbf{n}$ is the surface normal as usual, f_{nt} is the external heat flux or the temperature-dependent heat flux, which accounts for convection and radiation, and even conduction in some cases, and temperature T is the state variable. Since k_e and k_t can vary significantly with temperature, it is preferable to solve eqns [14] and [15] simultaneously. This is a standard feature in some commercial finite element software programs (e.g., ABAQUS, www.hks.com). Coding this on one's own is also rather straightforward.

After the temperature is known everywhere in the domain, it is included to impose the thermal load as shown below in the governing equations for thermo-elastic static equilibrium

$$\begin{aligned} &\nabla\cdot\boldsymbol{\sigma}=0 \text{ everywhere in } \Omega \\ &\boldsymbol{\sigma}=\mathbf{D}:(\boldsymbol{\varepsilon}-\alpha(T-T_0)\mathbf{I}) \\ &\mathbf{u}=\mathbf{u}_0 \text{ on } \partial\Omega_{md} \\ &\boldsymbol{\varepsilon}=\frac{1}{2}(\nabla\mathbf{u}+\nabla^T\mathbf{u}) \end{aligned} \quad [16]$$

where α is the coefficient of thermal expansion, T_0 is the ambient temperature, and the rest are as defined earlier for the elastostatic equilibrium problem. Eqn

[16] can be solved after eqns [14] and [15] are solved together.

Let the objective of the topology synthesis of ETC microactuators be the maximization of the displacement at a point, which we denote by $u_{\partial\Omega_{md}}$. By using the power law (or SIMP: Solid Isotropic Microstructure with Penalty) design parameterization (see eqn [6]), all material properties can be written in terms of the design variable $\rho(\mathbf{x})$

$$\begin{aligned} k_e &= k_{e0}\rho(\mathbf{x})^\eta \\ k_t &= k_{t0}\rho(\mathbf{x})^\eta \\ \alpha &= \alpha_0\rho(\mathbf{x})^\eta \\ \mathbf{D} &= \mathbf{D}_0\rho(\mathbf{x})^\eta \end{aligned} \quad [17]$$

As before, $\rho(\mathbf{x}) = 1$ indicates the presence of material at $\mathbf{x}$ and $\rho(\mathbf{x}) = 0$ the absence of material at $\mathbf{x}$. The penalty parameter $\rho(\mathbf{x})$ helps to push η to one of its extremes in the optimal solution. Now, following eqn [4], the topology synthesis problem involving the three energy domains can be posed as follows:

$$\begin{aligned} &\underset{\rho(\mathbf{x})}{\text{Minimize}} -u_{\partial\Omega_{md}} \\ &\text{Subject to} \\ &\quad \nabla\cdot(k_e\nabla V) = 0 \text{ in } \Omega \\ &\quad V = 0 \text{ on } \partial\Omega_{eg} \text{ and } V = V_0 \text{ on } \partial\Omega_{eV} \\ &\quad \nabla\cdot(k_t\nabla T) + \nabla^T k_e \nabla V = 0 \text{ in } \Omega \\ &\quad T = T_0 \text{ on } \partial\Omega_{tg} \text{ and } \mathbf{n}\cdot(k_t\nabla T) = f_{nt} \text{ on } \partial\Omega_{tf} \\ &\quad \nabla\cdot\boldsymbol{\sigma} = 0 \text{ everywhere in } \Omega \\ &\quad \boldsymbol{\sigma} = \mathbf{D}:(\boldsymbol{\varepsilon} - \alpha(T - T_0)\mathbf{I}) \\ &\quad \mathbf{u} = \mathbf{u}_0 \text{ on } \partial\Omega_{md} \\ &\quad \boldsymbol{\varepsilon} = \frac{1}{2}\left(\nabla\mathbf{u} + \nabla^T\mathbf{u}\right) \\ &\quad \int_\Omega \rho\, d\Omega - V_u \le 0 \\ &\quad 0 \le \rho \le 1 \\ &\text{Given that } k_e = k_{e0}\rho(\mathbf{x})^\eta;\ k_t = k_{t0}\rho(\mathbf{x})^\eta; \\ &\quad \alpha = \alpha_0\rho(\mathbf{x})^\eta;\ \mathbf{D} = \mathbf{D}_0\rho(\mathbf{x})^\eta \end{aligned} \quad [18]$$

The necessary conditions for the local minimum of this problem can be written and the resulting equations can be analytically simplified to derive an update formula for $\rho(\mathbf{x})$ as in eqn [10]. While this may look intimidating at first sight, the procedure is quite simple if the rules of calculus of variations and ordinary calculus are followed. Interested readers may consult Ananthasuresh (2003), Bendsoe and Sigmund (2003), Hafka and Gurdal (1989), and Weinstock (1974) for details. This procedure can also be performed for more than one material such as silicon and silicon dioxide. An example is shown in **Figure 13** where only thermal heat flux is applied and displacement that is inward into the design domain is desired.

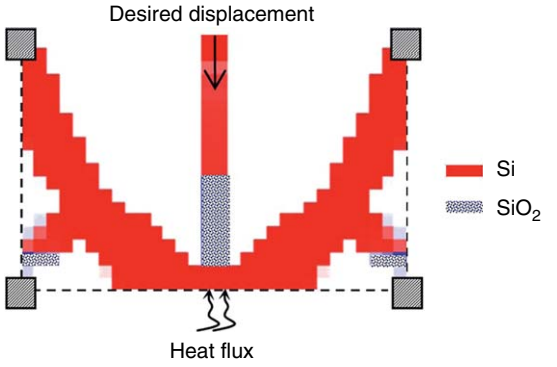


Figure 13 An example of thermoelastic synthesis problem. Here, two materials with different thermal conductivities need to be distributed in a rectangular design domain such that the midpoint of the top edge moves down, which is nonintuitive when heat flux is supplied from the bottom side as shown. The optimal solution achieves this by placing material with low conductivity between the portion that should move down and the portion where heat is supplied. This allows for a small amount of heat to pass and thus decreases the free deflection of the vertical bar in the middle.

ETC microactuators and a three degree-of-freedom planar platform, and a topology-optimized *xy*-microstage can be found in Moulton and Ananthasuresh (2001) and Sigmund (2001a), respectively.

1.19.4.4 Piezoelectric Actuators

Piezoelectric material-based actuators are widely used in micro- and nanopositioners and stages today. Piezoelectric materials provide very small displacements but large forces. So, they need to be combined with mechanical amplification mechanisms. Thus, there is a need for topology synthesis involving piezoactuators. This has been extensively investigated in Ananthasuresh (2003, Chaps. 5 and 6).

In piezoelectric materials, the coupling between the electrical and the elastic fields is inherently present at

the microscopic scale in the material's constitution. When mechanically stressed, these materials generate a net charge through electric polarization; and when an electric field is applied, they develop mechanical strain. Thus, this coupling is different from the ones discussed above in the context of electrostatic and electrothermal with the elastic field. In piezoelectric materials, the stress and strain tensors ($\boldsymbol{\sigma}_{ij}$ and ε_{kl}), and electric field and electric displacement vectors (E_k and D_i) are inherently related as follows:

$$\begin{aligned} \sigma_{ij} &= D_{ijkl}\varepsilon_{kl} - p_{kij}E_k \\ D_i &= p_{ikl}\varepsilon_{kl} + e_{ik}E_k \end{aligned} \qquad [19]$$

where D_{ijkl} is the usual elastic constitutive tensor relating stress and strain, e_{ik} is the dielectric tensor relating electric displacement and the electric field, and p_{kji} is the piezoelectric strain tensor. Based on this, the governing dynamic equilibrium equations in weak variational form under a sinusoidal excitation frequency ω for a body made of a piezoelectric material with a mass density γ can be written as given below (Eer Nisse 1967).

$$\begin{aligned} &\int_{\Omega} \left\{ \left(\varepsilon_{ij} D_{ijkl} \delta\varepsilon_{kl}\right) - \omega^2 \gamma u_i \delta u_i + (\nabla\phi)_i (\delta\varepsilon_{kl} p_{ikl}) \right\} \mathrm{d}\Omega \\ &\quad = \int_{\Omega} b_i \delta u_i \, \mathrm{d}\Omega + \int_{\partial\Omega_t} t_i \delta u_i \, \mathrm{d}\partial\Omega \\ &\int_{\Omega} \left\{ \varepsilon_{ij} p_{kij} (\delta\nabla\phi)_k - (\nabla\phi)_i \, e_{ij} (\delta\nabla\phi)_j \right\} \mathrm{d}\Omega \\ &\quad = \int_{\partial\Omega_d} (D_i n_i) \delta\phi \, \mathrm{d}\partial\Omega \end{aligned} \qquad [20]$$

where $\mathbf{E} = -\nabla\phi$ is used, Einstein's summation index notation is adopted, and state variables and their corresponding virtual quantities (variations) appear. These equations in the weak form can be used in the place of governing differential equations in eqn [4]. Then, by defining an objective function and relevant constraints, an optimization problem is formulated. In the design of the so-called flextensional actuators, a piezoelectric medium is combined with an elastic medium and the maximizing output displacement at a point in the elastic medium for the applied voltage between two points in the piezomedium is sought. The solution to this optimization problem provides a displacement amplifier that is tailored for a given piezoelectric actuator. Details of this can be found in Ananthasuresh (2003), Canfield and Frecker (2000), Silva and Kikuchi (1999), and Silva *et al.* (2000).

There have also been attempts to synthesize piezocomposites, which are made of piezomaterial and a nonpiezomaterial with a regular microstructure. In these problems, the topology of the microstructure is optimized to obtain desired piezoproperties indicated in eqn [19]. This is similar to the topology synthesis of microstructures materials discussed in Section 1.19.4.5. The optimized microstructure of piezocomposites has been manufactured using stereolithography and tested (Silva and Kikuchi 1999, Silva *et al.* 2000). This is indicative of the practical utility of topology synthesis of new piezoelectric materials with improved properties. Since thin piezoelectric films can be deposited and patterned, this provides the scope for novel uses of piezoelectric materials in microsystems.

1.19.4.5 Microstructured Materials

The compliant mechanism shown in **Figure 9** indicates that we can design a constrained elastic body that deforms in a desired manner in response to the mechanical forces applied on it. If this compliant mechanism is interpreted as a building block in the microstructure of a material and if this building block is periodically repeated to create a material, novel and unusual properties can be obtained. These are variously called periodic micromechanisms (Ananthasuresh 2003, Chap. 7), metamaterials, or simply tailor-made materials with an engineered periodic microstructure. Topology synthesis techniques involved in this are similar to the ones described above with the exception of periodic boundary conditions and computing the averaged (or homogenized) effects of a repeating building block. In mircosystems, while this idea of systematically synthesizing and fabricating the microstructure may look rather esoteric, the possibility certainly exists to enhance the performance. The piezocomposites noted in Section 1.19.4 also fall under this class of novel possibilities of combining topology synthesis with modern micromanufacturing methods.

Figure 14 shows a repeating pattern of a rod-and-hinge model. If this pattern is pulled outward in the horizontal direction, one can imagine that it will pull out in the vertical direction also. This means that this arrangement has negative Poisson's ratio (Almgren 1985). This is a consequence of the reentrant type arrangement of the rod and hinges in the building block of this repeating pattern. Now, if the building block with rods and hinges is converted to an elastic continuum – a compliant mechanism – a

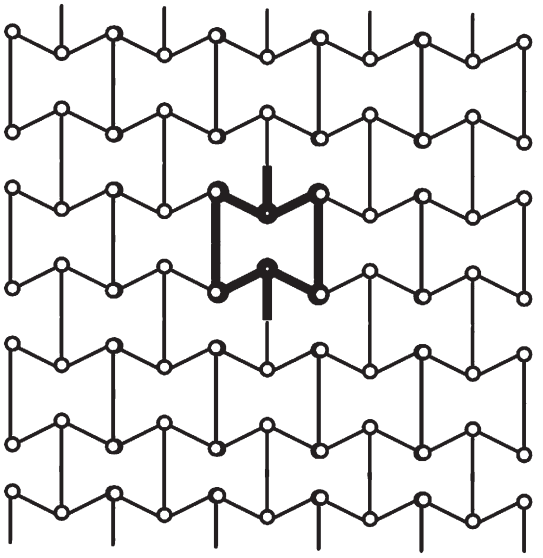


Figure 14 Rod and hinge model of a material with negative Poisson's ratio. (Source: Almgren R F 1985 An isotropic three-dimensional structure with Poisson's ratio = −1. *J. Elast.* **15(4)**, 427–30. Sigmund O 1994 Materials with prescribed constitutive parameters: An inverse homogenization problem. *Int. J. Solids Struct.* **31(17)**, 2313–29.)

microstructured material with negative Poisson's ratio is obtained. An example design, which was fabricated using single-crystal silicon and wafer bonding, is shown in **Figure 15**. This idea was proposed in Sigmund (1995) in which a numerical optimization technique, called inverse homogenization method, was also developed. Subsequently, microstructured materials with other extreme properties such as negative thermal expansion coefficients, tailored band gap characteristics, have been synthesized (Jensen and Sigmund 2006, Sigmund 1994, Sigmund and Torquato 1996). These materials are different from functionally graded materials, which have inhomogeneous material properties in a bulk structure.

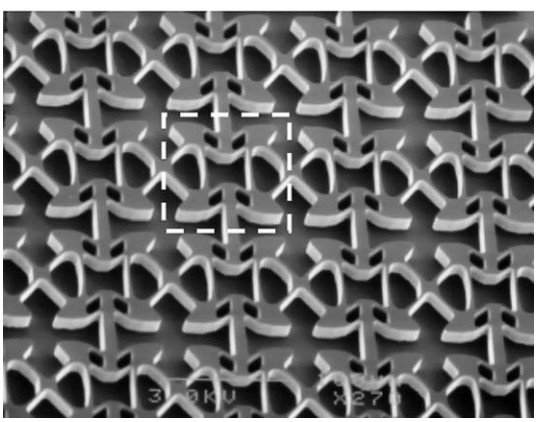


Figure 15 Silicon sheet with a repeating pattern of re-entrant type compliant mechanism building block (shown within a box of dashed lines) to give negative Poisson's ratio.

1.19.4.6 Microfluidic Components and Devices

Microfluidics is one of the most useful and also challenging areas of microsystems. With the advent of bioMEMS (microelectromechanical systems with biology applications), microfluidic devices have already been shown to have immense commercial potential in the healthcare industry. Plumbing fluids at the microscale is one of the main tasks of these devices. The performance of pumps, valves, mixers, reservoirs, etc., must be optimized by suitable geometry optimization to make the microfluidic systems more efficient and cost-effective. While the topology synthesis procedures described so far were concerned only with solids, fluids are not excluded. Methods for optimizing the topology and shape of fluidic systems are also in the offing. Representative works in this area include Evgrafov *et al.* (2006) and Klarbring *et al.* (2003). A simple example is shown in **Figure 16** wherein it is desired to obtain the optimal flow path between the inlet and the outlet to minimize the pressure drop. As one can imagine, exploring this manually involves a lot of trial-and-error attempts before the best is found. Of course, as with solids, experienced fluid mechanicians can do this intuitively. But in complex situations, systematic synthesis methods would help. Referring back to eqn [6] where material state was interpolated

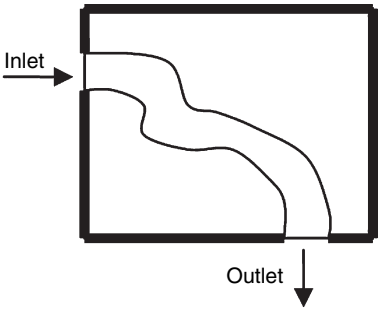


Figure 16 Topology synthesis for flow paths. What is the optimal path between the inlet and the outlet to minimize the pressure drop?

between 0 and 1 to get the absence, presence, and the intermediate state for the problem in **Figure 16**, the state of the material in the design domain varies smoothly between a liquid and a void. Governing equations for this needed special treatment that uses lattice Boltzmann modeling as can be found in Jensen and Sigmund (2006). Similar ideas of smooth transitions between solid and liquid were proposed in fluid–structure interaction topology synthesis problems (Clausen and Sigmund 2006). Significant potential exists in effectively using these techniques in microfluidic components.

1.19.5 Process and Mask Synthesis

The final step in the design of any microsystem component is to obtain the lithography mask layouts that are essential for microfabrication. But interestingly, most often microsystem designers are expected to begin with lithography layouts to initiate their design process. Commercial microsystem design software programs include a module for this and advocate beginning the design task with layout drawings. Lately, this has been changing with the introduction of a system-level abstraction with reduced-order models. In contrast, designers at the macroscale are not overly restricted in creating the geometries while designing parts. At the macroscale, process planning for machining is usually a step that follows the design process. At the microscale, despite the rapid development in microfabrication, several limitations exist, and these constrain the microsystem designers. They also do not yet have tools for the equivalent of process planning. Such tools exist for microelectronics (Hasanuzzaman and Mastrangelo 1996) but not for micromechanical components operating in different energy domains. In this section, some demonstrated works, which are very few, are mentioned, and some possibilities for a paradigm shift in microsystem design at the process level indicated in **Figure 4** are discussed.

1.19.5.1 Process Synthesis

Unlike a macrosystem designer, a microsystem designer ought to choose a fabrication process at the very beginning of the design process. In most cases, the designers use an existing process and tailor the design to fit that process. Those designers who use microsystem foundry processes have no choice in this matter. But for those who can develop and run their own processes, it is sensible to obtain the most suitable process for a given design. This implies that after the geometry of the components and its material composition are decided (preferably without being constrained by how the component is going to be made) a process would be chosen. This is currently done by human designers and microfabrication specialists. Making this systematic and even automating it would be preferable. A few methods have been developed for microelectronic processes along these lines. The investigations and methods described in Hasanuzzaman and Mastrangelo (1996) and Hasanuzzaman *et al.* (1999) are the right steps in this direction. Details can be found in Ananthasuresh (2003, Chap. 8).

1.19.5.2 Mask Synthesis

Choosing a process is not the end of the microsystem design procedure; one needs to create lithography mask layouts that would help realize the desired 3D geometry. Thus, mask synthesis entails systematic means to generate the mask layouts from optimized or finalized 3D geometry of components. There is a limited amount of literature in this regard. Mask synthesis for wet-etched single-crystal silicon is discussed in Lee and Antonsson (2000), with an overview provided in Antonsson and Cagan (2001). In this technique, a genetic algorithm was used to synthesize the desired 3D etch patterns in a single layer of silicon. The mask synthesis for surface micromachining is proposed in Ananthakrishnan *et al.* (2003) wherein a new representation scheme was developed so that the mask synthesis can be reduced to solve a system of linear equations. Details of this can be found in Ananthasuresh (2003, Chap. 9).

1.19.5.3 Synthesis with Manufacturing Constraints

If a component's geometry is determined without considering how it will be fabricated (which is what a designer would like), it might not even be possible to practically realize that geometry with a chosen process. So, it is desirable to include micromanufacturing constraints within the topology synthesis procedures. An attempt in this direction has been made recently for surface micromachining process (Alwan and Ananthasuresh 2006c). Alternatively, a feature-based modeling wherein the designer is automatically restricted from creating features that are

illegal for a chosen process has also been reported (Ananthakrishnan 2000). This enables model creation with human participation as well as automated geometry design.

Given the limited number of investigations in process and mask synthesis and significant opportunities in making the microsystem design efficient, there is a need for further progress in this area.

1.19.6 Closure

Design is an iterative process. Several analysis steps using candidate designs are necessary before a design is finalized. This is a laborious process. The designer's intuition, prior experience, and knowledge about the entity being designed guide this process. Often it is a trial-and-error process even for an experienced designer who encounters a new design task. Synthesis, on the other hand, is a systematic step of arriving at a design in less time than in trial-and-error design. If optimization is used, the systematic synthesis can also give the optimum design. In this chapter, optimal synthesis procedures that are developed for microsystems and their components have been discussed.

Four categories of synthesis problems were noted for different levels of abstraction of microsystems. The first level, where the architecture of the system and its subsystems are identified, is highly subjective and is not easily amenable for systematic synthesis. At present there is no significant literature on this to discuss it. The second level, where a system-level model is constructed with reduced-order models for the subsystems, is suited for parametric optimization. The methods for this are simple as discussed in the chapter. The third level, where the geometry of the components is determined, is the most time-consuming step and is also the most suited for systematic synthesis. The methods for this are described for six different types of problems – many involving multiple coupled energy domains. The fourth level, where the microfabrication process and lithography mask layouts are to be generated, is a practically important synthesis problem. But it has not received as much attention as it deserves. A few methods that are developed for this are presented.

The work noted in this chapter is by no means exhaustive. Several other researchers have also contributed to the synthesis of microsystems. Only selected works that fit within the scope and viewpoint of this chapter are included here.

Synthesis techniques enable designers to design faster and better. Unlike the macroworld where there are many more options for integration and manufacturing, microsystems are conveniently restricted. So, systematic synthesis is possible and beneficial for microsystems. This was the conclusion of a workshop held in 1996 in which several synthesis researchers were working in the field of microsystems (Antonsson 1996). After a decade of further development in synthesis, the microsystems industry and users are yet to significantly benefit by adopting synthesis procedures as part of their design tasks. One hopes that it will change gradually with an anticipation that designers would see synthesis as an integral part of design rather than viewing the design merely as repeated analysis (i.e., simulations) trials until a satisfactory (not necessarily optimum) design is found.

Acknowledgments

The author thanks all his former and present students and post-doctoral research associates who worked on different aspects of the work represented here.

References

ABAQUS, www.hks.com

Achtziger W 2006 Simultaneous optimization of Truss topology and geometry. Revised. In: Bendsoe M P, Olhoff N, and Sigmund O (eds.) *IUTAM Symp. Topology Design Optimization of Structures, Machines and Materials.* Springer, Berlin, pp. 413–23

Almgren R F 1985 An isotropic three-dimensional structure with Poisson's ratio $= -1$. *J. Elast.* **15(4)**, 427–30. Sigmund O 1994 Materials with prescribed constitutive parameters: An inverse homogenization problem. *Int. J. Solids Struct.* **31(17)**, 2313–29

Alwan A, Ananthasuresh G K 2006a Coupled electrostatic–elastic analysis for topology optimization using material interpolation. *J. Phys. Conf. Ser.* **34**, 264–71

Alwan A, Ananthasuresh GK 2006b Topology optimization of electrostatically actuated micromechanical structures with accurate electrostatic modeling of the interpolated material model. *CD-ROM Proc. ASME International Design Engineering and Technical Conferences*, Philadelphia, PA, USA, Sept. 10–13, 2006, paper #. DETC2006-99684

Alwan A, Ananthasuresh GK 2006c Topology optimization of micromachined structures with surface micromachining manufacturing constraints. *CD-ROM Proc. ASME International Design Engineering and Technical Conferences*, Philadelphia, PA, USA, Sept. 10–13, 2006, paper # DETC2006-99341

Ananthakrishnan V 2000 Part to art: The basis for a systematic geometric design tool for surface micromachined MEMS. Masters thesis, University of Toledo

Ananthakrishnan V, Sarma R, Ananthasuresh G K 2003 Mask synthesis for surface-micromachined MEMS. *Micromech. Microeng.* **13**, 927–41

Ananthasuresh GK 1994 Investigations on the synthesis of compliant mechanisms and a new design paradigm for micro-electro-mechanical systems. Ph.D. dissertation, University of Michigan

Ananthasuresh G K 2003 *Optimal Synthesis Methods for MEMS.* Kluwer Academic Publishers, Boston, MA

Ananthasuresh G K, Kota S 1995 Designing compliant mechanisms. *Mech. Eng.* **117(1)**, 93–6

Ananthasuresh GK, Kota S, Kikuchi N 1994 Strategies for systematic synthesis of compliant MEMS. *Proc. 1994 ASME Winter Annual Meeting*, *Dynamics Systems and Control*, Chicago, IL, Nov. 1994, DSC-Vol. 55-2, pp. 677–86

Ananthasuresh G K, Gupta R K, Senturia S D 1996 An approach to macromodeling of MEMS for nonlinear dynamic analysis. *Micro-Electro-Mechanical Systems (MEMS), 1996*, *ASME International Mechanical Engineering Congress and Exposition*, Atlanta, GA, USA, Nov. 17–22, 1996, DSC-Vol. 59, pp. 401–7

Antonsson E K (ed.) 1996 *Structured Design Methods for MEMS. Workshop Report*. CalTech, Pasadena, CA, http://www.design.caltech.edu/NSF_MEMS_Workshop

Antonsson E K, Cagan J 2001 *Formal Engineering Design Synthesis.* Cambridge University Press, Cambridge

Bangtsson E, Noreland D, Berggren M 2003 Shape optimization of an acoustic horn. *Comput. Meth. Appl. Mech. Eng.* **192**, 1533–71

Bendsoe M P, Kukuchi N 1988 Generating optimal topologies in structural design using a homogenization method. *Comput. Meth. Appl. Mech. Eng.* **71(2)**, 197–224

Bendsoe M P, Sigmund O 1999 Material interpolation schemes in topology optimization. *Arch. Appl. Mech.* **69(9–10)**, 635–54

Bendsoe M P, Sigmund O 2003 *Topology Optimization: Theory, Methods and Applications.* Springer, Berlin

Canfield S, Frecker M 2000 Topology optimization of compliant mechanical amplifier for piezoelectric actuator. *Struct. Multidisc. Optim.* **20**, 269–79

Clausen P M, Sigmund O 2006 The pressure load problem re-visited. In: Bendsoe M P, Olhoff N, and Sigmund O (eds.) *IUTAM Symp. Topology Design Optimization of Structures, Machines and Materials.* Springer, Berlin, pp. 261–7

Comtois J, Bright V 1996 Surface micromachined polysilicon thermal actuator arrays and applications. *Tech. Digest Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, June 1996, pp. 174–7

Diaz A R, Kikuchi N 1992 Solution to shape and topology eigenvalue optimization problems using a homogenization method. *Int. J. Numer. Meth. Eng.* **35**, 1487–502

Duysinx P, Bendsoe M P 1998 Topology optimization of continuum structures with local stress constraints. *Int. J. Numer. Meth. Eng.* **43**, 1453–78

Eer Nisse E P 1967 Variational method for electroelastic vibration analysis. *IEEE Trans. Sonics Ultrason.* **SU-14(4)**, 153–60

Evgrafov A, Pingen G, Maute K 2006 Topology optimization of fluid problems by the lattice Boltzmann method. In: Bendsoe M P, Olhoff N, and Sigmund O (eds.) *IUTAM Symp. Topology Design Optimization of Structures, Machines and Materials.* Springer, Berlin, pp. 559–68

Frecker M, Ananthasuresh G K, Nishiwaki N, Kikuchi N, Kota S 1997 Topological synthesis of compliant mechanisms using multi-criteria optimization. *ASME J. Mech. Des.* **119**, 238–45

Goldberg D E 1989 *Genetic Algorithms in Search, Optimization and Machine Learning* Kluwer Academic Publishers, Boston, MA

Guckel H, Klein J, Christenson T, Skrobis K, Laudon M, Lovell E G 1992 Thermomagnetic metal flexure actuators. *Tech. Digest Solid State Sensors and Actuators Workshop* p. 73 Hilton-Head Island, SC, USA, p. 73

Haftka R T, Gurdal Z 1989 *Elements of Structural Optimization.* Kluwer Academic Publishers, New York

Hasanuzzaman M, Mastrangelo C H 1996 Process compilation of thin-film microdevices. *IEEE Trans. Comput. Aided Des.* **30**, 1438–53

Hasanuzzaman M, Carlen E T, Mastrangelo C H 1999 Automatic generation of thin-film process flows – Part I: Basic algorithms. *IEEE Trans. Semicond. Manufact.* **12(1)**, 1–12

Hetrick J A, Kota S 1999 An energy formulation for parametric size and shape optimization of compliant mechanisms. *ASME J. Mech. Des.* **121**, 229–234

Howell L L 2001 *Compliant Mechanisms.* John Wiley and Sons, New York

Hung E S, Senturia S D 1999 Extending the travel range of analog-tuned electrostatic actuators. *J. Microelectromech. Syst.* **8**, 497

Iyer S, Mukherjee T, Fedder GK 1998 Multi-mode sensitive layout synthesis of microresonators. *Proc. First Int. Conf. Modeling and Simulation of Microsystems (MSM 98)*, Santa Clara, CA, USA, April 6–8, 1998

Jensen J S, Sigmund O 2006 Topology optimization of wave-propagation problems. In: Bendsoe M P, Olhoff N, and Sigmund O (eds.) *IUTAM Symp. Topology Design Optimization of Structures, Machines and Materials.* Springer, Berlin, pp. 387–90

Klarbring A, Petersson J, Torstenfelt B, Karlsson M 2003 Topology optimization of flow networks. *Comput. Meth. Appl. Mech. Eng.* **192(35–36)**, 3909–32

Lai E, Ananthasuresh G K 2002 On the design of bars and beams for desired mode shapes. *J. Sound Vib.* **254(2)**, 393–406

Lee C-Y, Antonsson E K 2000 Self-adapting vertices for mask layout synthesis. In: Laudon M and Romanowicz B (eds.) *Proc. Modeling and Simulation of Microsystems Conference,* San Diego, CA, USA, March 27–29, 2000, pp. 83–6

Ma Z-D, Kikuchi N, Cheng H-C 1995 Topological design for vibrating structures. *Comput. Meth. Appl. Mech. Eng.* **121(1–4)**, 259–280

Mankame N, Ananthasuresh G K 2004 Topology synthesis of electro-thermal-compliant mechanisms using line elements. *Struct. Multidisc. Optim.* **26**, 209–18

Moulton T, Ananthasuresh G K 2001 Design and manufacture of electro-thermal-compliant micro devices. *Sens. Actuators, Phys.* **90**, 38–48

Mukherjee T, Iyer S, Fedder G K 1998 Optimization based synthesis of microresonators. *Sens. Actuators A* **21**, 328–31

Mukherjee T, Zhou Y, Fedder GK 1999 Automated optimal synthesis of microaccelerometers. *Tech. Digest IEEE Int. Conf. Microelectromechanical Systems (MEMS 99)*, Orlando, FL, USA, Jan. 17–21, 1999, pp. 326–31

Papalambros P Y, Wilde D J 2000 *Principles of Optimal Design: Modeling and Computation.* Cambridge University Press, Cambridge

Pedersen C B W, Buhl T, Sigmund O 2001 Topology synthesis of large-displacement compliant mechanisms. *Int. J. Numer. Meth. Eng.* **50**, 2683–705

Pedersen C B W, Fleck N A, Ananthasuresh G K 2006 Design of a compliant mechanism to modify an actuator characteristic to deliver a constant output force. *J. Mech. Des.* **128(5)**, 1101–12

Rao S S 2000 *Engineering Optimization: Theory and Practice.* John Wiley and Sons, New York

Raulli M 2006 Electro-mechanical topology optimization considering non-matching meshes. In: Bendsoe M P, Olhoff N, and Sigmund O (eds.) *IUTAM Symposium on Topology Design Optimization of Structures, Machines and Materials.* Springer, Berlin, pp. 277–87

Raulli M, Maute K 2005 Topology optimization of electrostatically actuated microsystems. *Struct. Multidisc. Optim.* **30(5)**, 342–59

Saxena A, Ananthasuresh G K 2001a Topology design of compliant mechanisms with strength considerations. *Mech. Struct. Mach.* **29(2)**, 199–221

Saxena A, Ananthasuresh G K 2001b Topology synthesis of compliant mechanisms for nonlinear force-deflection and curved path specifications. *J. Mech. Des.* **123**, 33–42

Senturia S D 2001 *Microsystem Design.* Kluwer Academic Publishers, New York

Senturia S D, Harris R M, Johnson B P, Kim S, Nabors K, Shulman M A, White J K 1992 A computer-aided design system for microelectromechanical system (MEMCAD). *J. Microelectromech. Syst.* **1(1)**, 3–13

Senturia S D, Aluru N, White J 1997 Simulating the behavior of MEMS devices: Computational methods and needs. *Comput. Sci. Eng.* **4(1)**, 30–43

Sigmund O 1994 Design of material structures using topology optimization. Ph.D. thesis, Department of Mechanics, Technical University of Denmark

Sigmund O 1995 Materials with prescribed elastic properties. *Mech. Mater.* **20**, 351–68

Sigmund O 1997 On the design of compliant mechanisms using topology optimization. *Mech. Struct. Mach.* **25**, 495–526

Sigmund O 2001a Design of multiphysics actuators using topology optimization – Part I: One-material structures. *Comput. Meth. Appl. Mech. Eng.* **190**, 6577–604

Sigmund O 2001b Design of multiphysics actuators using topology optimization – Part II: Two-material structures. *Comput. Meth. Appl. Mech. Eng.* **190**, 6605–27

Sigmund O 2001c A 99 line topology optimization code written in MATLAB. *Struct. Multidisc. Optim.* **21**, 120–7

Sigmund O, Torquato S 1996 Composites with extremal thermal expansion coefficients. *Appl. Phys. Lett.* **69(21)**, 3203–5

Silva E C N, Kikuchi N 1999 Design of piezoelectric transducers using topology optimization. *J. Smart Mater. Struct.* **8**, 350–64

Silva E C N, Nishiwaki S, Kikuchi N 2000 Topology optimization design of flextensional actuators. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **17(3)**, 657–71

Weinstock R 1974 *Calculus of Variations with Applications to Physics and Engineering.* Dover Publications, New York

Xu D, Ananthasuresh G K 2003 Skeletal shape optimization of compliant mechanisms. *ASME J. Mech. Des.* **125(2)**, 253–61

Ye W, Mukherjee S 1999 Optimal design of three-dimensional MEMS with applications to electrostatic comb-drives. *Int. J. Numer. Meth. Eng.* **45**, 175–94

Ye W, Mukherjee S, MacDonald N C 1998 Optimal shape design of an electrostatic comb-drive in microelectromechanical systems. *J. Microelectromech. Syst.* **7(1)**, 16–26

Yin L, Ananthasuresh G K 2002 Novel design technique for electro-thermally actuated compliant micromechanisms. *Sens. Actuators A Phys.* **97–98**, 599–609

Yoo J, Soh H J 2005 An optimal design of magnetic actuators using topology optimization and the response surface method. *Microsyst. Technol.* **11(12)**, 1252–61

Biography

G. K. Ananthasuresh obtained his B.Tech. degree in 1989 from IIT, Madras, and Ph.D. in 1994 from the University of Michigan, Ann Arbor, MI, USA. He was a post-doctoral research associate in the Microsystems Technology Laboratories at MIT, Cambridge, MA, USA, before he joined the University of Pennsylvania's Mechanical Engineering and Applied Mechanics Department in 1996 as an assistant professor and then promoted with tenure to Associate Professor in 2002. He joined the Indian Institute of Science (IISc) in 2004. His research interests include compliant mechanisms, microelectromechanical systems (MEMS), multidisciplinary design optimization, topology optimization, microfabrication, mechanism design and kinematics, protein design, and micromanipulation. He has more than 115 publications out of which more than 45 are journal papers. He is also the author of five book-chapters and an edited book entitled *Optimal Synthesis Methods for MEMS*. He has two patents and three best paper awards at international conferences. He is the recipient of the Swarnajayanthi Fellowship of the Department of Science and Technology of India, National Science Foundations' Early Career (CAREER) award and Society of Automotive Engineers' Ralph Teetor Distinguished Educator award. He has supervised six Ph.D. theses and three MS theses at the University of Pennsylvania. He is currently advising eight students and five research project assistants at IISc.

2.01 Electrostatic Actuation

Hiroshi Toshiyoshi, Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

Glossary

Electrostatic actuation A principle to drive micro actuators using an attractive or a repulsive Coulomb force between electrically charged bodies, which is proportional to the square of applied voltage and inversely proportional to the mutual distance. It is widely used in MEMS actuators thanks to the simplicity of mechanism and its increasing significance in micron scale.

Parallel-plate actuator An ideal analytical model for electrostatic actuators that are composed of a pair of parallel electrode plates (fixed and movable) with an air gap in between.

Comb-drive actuator An electrostatic actuator with an inter-digitized electrodes of a comb-look shape for longer travel distance than that of a parallel-plate actuator.

Vertical comb-drive actuator A derivative of lateral comb-drive actuator, and is usually used to cause out-of-plane motion for tilting mirrors, for instance.

Electrostatic pull-in A characteristic phenomenon found in most electrostatic actuators, at which a movable electrode loses control due to the electrostatic attractive force that has become larger than the restoring force of suspending tethers.

Electrostatic charge-up A cause of displacement drift of electrostatic actuator, where accumulation of electrical charge on the electrodes or nearby bodies causes change in the electrostatic force acting on the movable electrode. In an extreme case of electrostatic charge-up, the movable electrode is trapped on the counter electrode to cause in-use stiction.

Differential voltage drive A method to linearize the displacement-voltage characteristic of electrostatic actuator by using a pair of electrostatic attractive forces pulling a movable body in opposite directions.

Scratch-drive actuator A micro plate with a bushing on the bottom surface of its leading edge, and used to cause long travel range on an electrically insulated surface through the repetition of electrostatic stick-slip motion.

Negative spring constant effect A phenomenon that an electrostatic actuator looks to have reduced its spring constant at an electrically biased state due to the non-linearly increasing electrostatic force over linearly increasing restoring elastic force.

2.01.1 Overview of Electrostatic Actuation

2.01.1.1 Introduction

Electrostatic actuation utilizes an attractive or a repulsive Coulomb force between electrically charged bodies. This mechanism is the most widely used in microelectromechanical systems (MEMS) microactuators today for its mechanical simplicity; a pair of electrodes with a tiny separation can be implemented in various shapes and orientation by the micromachining technique. This mechanism has good compatibility in scale with microfabrication. Electrostatic force acting on the surface drastically increases with decreasing scale, compared with the body force such as gravity and inertia force.

2.01.1.2 Classic Electrostatic Actuation

The history of electrostatic actuators dates back further than that of electromagnetic mechanism, to the early 18th century (Jefimenko 1973). The world's first electrostatic actuator was the electric bell invented by Andrew Gordon in 1742. A tiny metal

hammer was suspended with an electrically insulating thread in between a pair of metal bells that were oppositely charged. When the hammer was attracted to one of the bells, it was charged to the same polarity as the bell and then attracted to the opposite one. The shuttle process of ringing repeated as long as the electricity was supplied to the bells. Gordon also invented another type of electrostatic mechanism called an electric fly. Metal wires were bent in a windmill shape with pointed ends. When the whole windmill was electrically charged, the air molecules around the pointing ends were corona-discharged, resulting in a repulsive force that turned the mill.

Another mechanism, the spark motor, was a rotational version of the Gordon's electric bell, invented by Benjamin Franklin in 1748 (Jefimenko 1973). Glass rods with a metal cap were arranged like the spokes of a wheel and placed in between a pair of Leyden jars that were charged with opposite polarity. The metal cap that passed by the Leyden jar was positively charged by electric breakdown (spark) and was driven out by the repulsive force that turns the wheel. On the other side of the wheel, the cap was negatively charged to keep the wheel rotated. Franklin observed 0.1 W power output by using a wheel 1 m in diameter. Franklin's spark motor was modernized by Karpov *et al.* in 1960 to be a corona motor, where electrical charges were given not by the electrostatic breakdown but by the corona discharge.

Capacitor motor is a precursor of today's electrostatic micromotors. It is also known to be a variable capacitance motor, because the rotor moves in the direction to increase the electrostatic capacitance with increasing drive voltage. The first capacitor motor was developed by Karl Zipernowsky in 1889. Recently, Bollee *et al.* reported a precision-machined version of capacitor motor, as shown in **Figure 1** (Bollee 1969). When voltages were applied to the stator and the rotor, the electrostatic attractive force acted on the rotor in the direction to maximize the overlap of the electrodes. For perpetual rotating motion, pulse voltages were applied to the stators in turn. Bollee demonstrated a capacitor motor of 100 stators that generated 0.6 mW power with an operating voltage of 200 V.

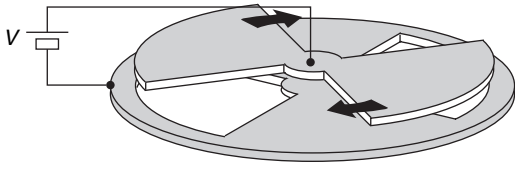


Figure 1 Capacitor motor. (Source: Bollee B 1969 Electrostatic motors. *Philips Tech. Rev*. **30(6/7)**, 178–94.)

2.01.1.3 Electrostatic Actuator in Microscale

Electrostatic motors manufactured by conventional macromachining were reported to be in the same physical scale as today's electromagnetic motors, ranging in the order of a few centimeters to several tens of centimeters, and the electrostatic gaps were made to be on the order of millimeters. In such old and large-scale mechanics, however, the energy density of electrostatic actuator was lower than that of the electromagnetic type, and high voltages (over several hundreds) were needed to produce significant mechanical work. For this reason, the electrostatic force was not a popular principle for macroscale mechanism but was used in limited applications such as dust collectors, electrostatic chuck, and in recent electronics, toner transfer mechanism in laser printers.

A landsliding change in electrostatic actuation came in 1987, when AT&T Bell Labs reported a new use of the semiconductor integrated circuit (IC) processes to develop micromechanical structures (Trimmer and Gabriel 1987). A thin-film polysilicon was used as a structural material to develop a series of free-rotating microgears 125–185 μm in diameter and 4 μm in thickness on a silicon wafer, as shown in **Figure 2**. No actuation mechanism was implemented in their first demonstration, but the scanning electron microscope (SEM) image gave a huge impetus to both electrical and mechanical engineers.

It was only a year later that an electrically driven micromotor was reported by Fan *et al.* with the University of California at Berkeley (Fan *et al.* 1988, Tai *et al.* 1989). The motor material was a thin-film doped polysilicon, the same as that of AT&T microgears. In the Berkeley motor, a cross-shaped rotor

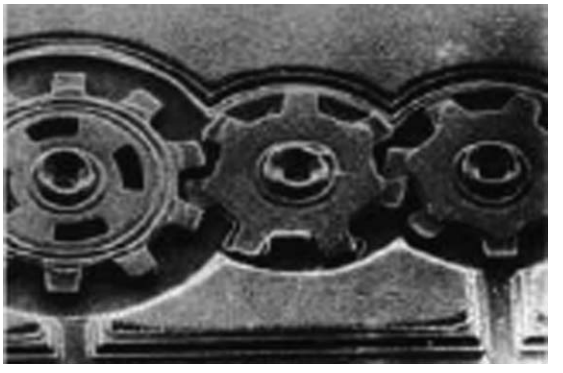


Figure 2 Microgears developed by silicon surface micromachining. (Source: Mehregany M, Gabriel K J, Trimmer W S N 1988 Integrated fabrication of polysilicon mechanisms. *IEEE Trans. ED* **35(6)**, 719–23.)

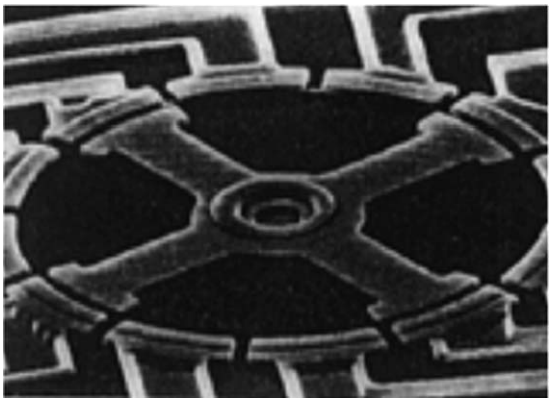


Figure 3 First demonstration of electrostatic micromotor. (Source: Tai Y C, Muller R S 1989 IC-processed electrostatic synchronous microrotors. *Sens. Actuators* **20(1/2)**, 41–8.)

(2 μm in thickness and 120 μm in diameter) was made to freely rotate about the center pivot, as shown in **Figure 3**. A total of 12 stators were placed around the rotor, with a 2 μm gap; a set of triphase voltages (300 V) was applied to the stator pads in turn to generate electrostatic force that worked to align the rotor's end to the stator. The motor was called edge-driven type because the electrostatic force acted on the rotor edges. A silicon nitride film, a solid lubricant material, was deposited on the inner wall of the rotor to reduce the friction with the center pivot (Tai *et al.* 1990). Despite such effort to reduce friction, the maximum rotation speed was limited to only 300 rpm; this speed was quite low compared with the their theoretical prediction, 120 000 rpm. The first prototype motor was also reported to work only for the first 3 min.

A solution to overcome the friction was to eliminate any sliding motion at the mechanical interface. Mehregany *et al.* (1990a, b) with MIT developed a new electrostatic micromotor. A schematic is shown in **Figure 4**. Unlike in the University of California, Berkeley micromotor, there were no projections in the rotor but a round disk. The rotor was designed to have a small looseness in the inner diameter, which allowed the rotor to move toward an electrically biased stator. When the voltage moved to the neighboring stator in turn, the rotor redirected itself to it without sliding on the center shaft, exhibiting a frictionless wobbling motion. The inner circumference of the disk (L_d) was larger than the outer circumference of the center shaft (L_c); hence, the rotation of the disk was reduced with respect to the rotation of voltage by a factor of $(L_d - L_c)/L_d$. Thanks to the frictionless motion and reduction gear mechanism, the disk could rotate at a fast spin rate of 10 000 rpm or higher for more than 1 week in the dry nitrogen environment. Mehregany also discovered that the mechanism of motor failure was associated with the native oxide grown on the surface of the polysilicon rotor; electrical conductance on the surface was degraded by the presence of native oxide, and the rotor became gradually charged up, finally resulting in electrostatic stiction to the substrate or the center post.

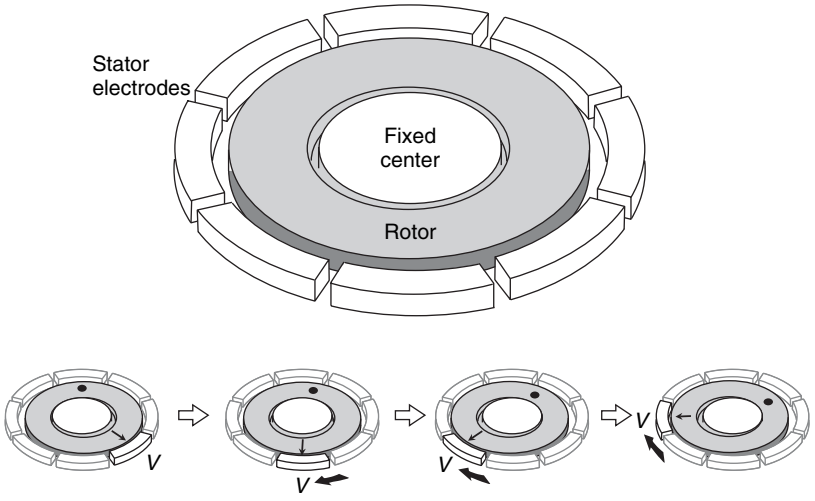


Figure 4 Electrostatic wobble motor. (Sources: Mehregany M, Bart S F, Tavrow L S, Lang J H, Senturia S D, Schlecht M F 1990a A study of three microfabricated variable-capacitance motors. *Sens. Actuators* **A21**, 173–9 and Mehregany M, Nagarkar P, Senturia S D, Lang J H 1990b Operation of microfabricated harmonic and ordinary side-drive motors. *Proc. IEEE Micro Electro Mechanical Systems Workshop*, Napa Valley, CA, USA, February 11–14, 1990, pp. 1–8.)

2.01.1.4 Suspended Micromechanism

Friction due to the sliding motion can be also eliminated by using an elastically tethered structure, but its motion is limited by the elastic restoring force of the suspensions. Mechanically movable suspended microstructures were developed even before micromotors. The earliest micromechanical structure in a cantilever shape was developed by electroplating metal on a metal oxide semiconductor (MOS) transistor substrate, as shown in **Figure 5**, reported by Nathanson and Wickstrom (1965). They used the resonant frequency of the suspended structure for bandpass filtering and obtained the signal component of a particular frequency by using the underlying MOS transistor. Various sorts of bulk silicon micromachined structures were reported in the famous review article by K. E. Petersen (1982). Most of their actuation directions were limited to the out-of-plane with respect to the substrate surface.

The first lateral-motion actuators of long traveling range was developed by Tang *et al.* (1989) with the University of California at Berkeley. **Figure 6(a)** shows a SEM image of the laterally driven electrostatic actuator. The movable structure was suspended with the folded beam structures (100 μm long each)

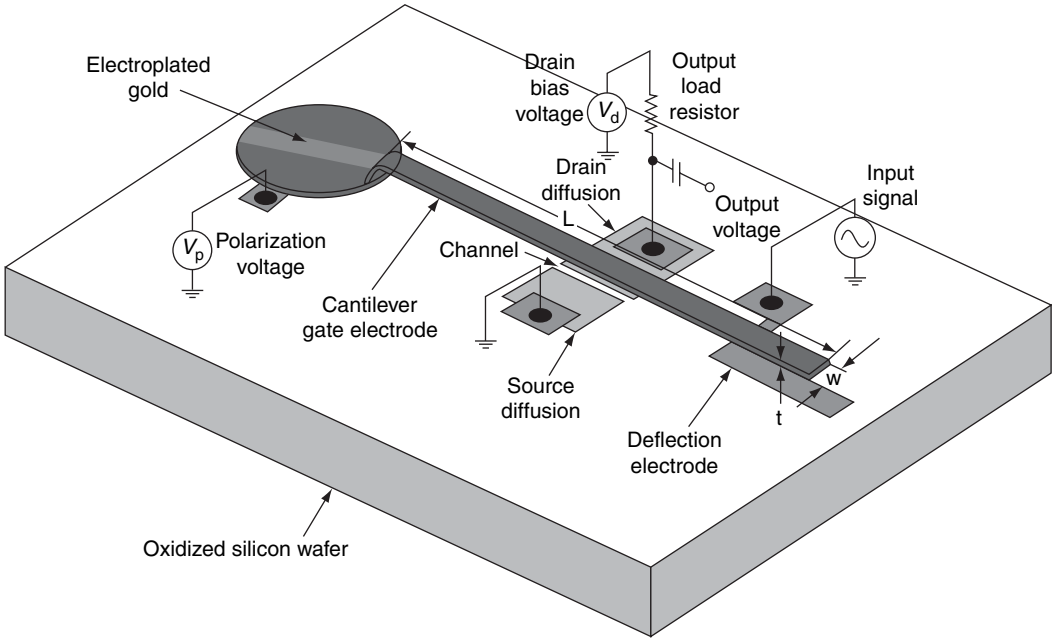


Figure 5 Resonant gate transistor used as an electromechanical bandpass filter. (Source: Nathanson H C, Wickstrom R A 1965 A resonant-gate silicon surface transistor with high-Q bandpass properties. *Appl. Phys. Lett.* **7**, 84.)

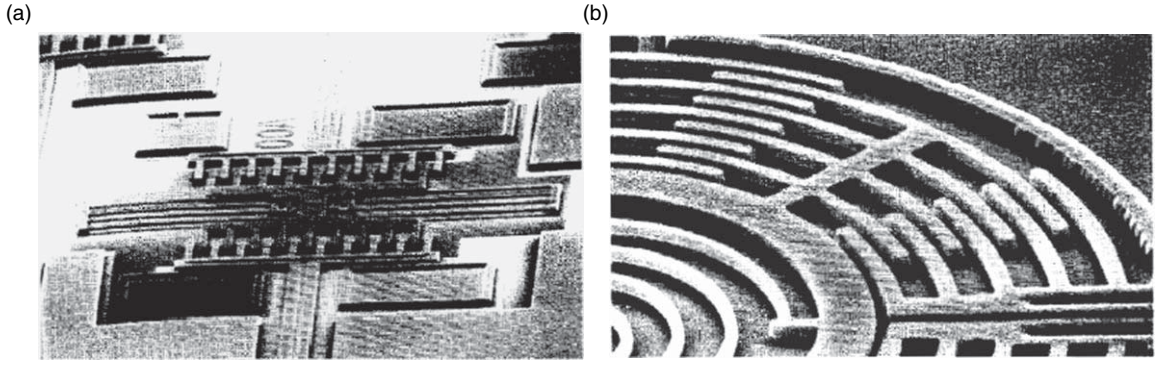


Figure 6 Electrostatic comb-drive resonator: (a) lateral and (b) rotational types. (Source: Tang W C, Nguyen T-C, Howe R T 1989 Laterally driven polysilicon resonant microstructures. *Sens. Actuators* **20**, 25–32.)

and anchored to the substrate at the middle part. The drive electrodes were designed to be interdigitated structures like comb teeth with a 2-μm gap. Mechanical stroke of 5 to 20 μm was reported with typical drive voltage ranging from 10 to 30 V.

Electrostatic force of the comb-drive actuator could be increased by decreasing the intercomb air gap. Hirano *et al.* (1992) reported a postprocess technique to make a comb-drive actuator with very small gap of 0.5 μm, which was further smaller than the resolution of the photolithography steps.

An in-plane rotational version of the comb-drive actuator, as shown in **Figure 6(b)**, was also reported by Tang *et al.* A polysilicon ring 1 mm in diameter was supported in the center part with four sets of meandering suspensions, and coaxial comb electrodes were arranged at the outer rim. The comb-drive mechanism was employed in actuators; they can also be used as physical sensors such as accelerometers, gyroscopes, and electromechanical filters.

2.01.2 Advantages of Electrostatic Actuation

2.01.2.1 Electrostatic Energy

Despite the long history of electrostatic actuation, the mechanism has been used in limited applications where no large force was required. Compared with the magnetic field, electrostatic field has poor energy density; however, it becomes comparable with that of magnetic, because the maximum electrostatic field that a solid insulating material can sustain (electrostatic breakdown field) increases with decreasing thickness of the material. For instance, the breakdown field in silicon dioxide (SiO_2) can be as high as 2 MV cm^{-1}, which corresponds to an energy of a 1.3-T magnetic field. The electrostatic attractive force (pressure) at this field strength is calculated to be almost 0.65 MPa (nearly 6.5 atm). Breakdown voltage (or electric field) further increases in gas. According to Paschen's law, breakdown voltage in gas is known to have a minimum value, no matter how small the electrostatic gap is tuned (Dakin *et al.* 1974, Paschen 1889). For instance, the minimum voltage to start breakdown in air is 330 V; assuming a 1-μm gap, the voltage corresponds to a static field as large as 3 MV cm^{-1}. On the contrary, magnetic field is limited by the saturation field strength of material; typical value is on the order of 1 T.

Besides the large electrostatic energy density in microscale, simplicity of mechanism is an advantage of electrostatic actuation. Compared with electromagnetic actuators that require coils of rather complicated topology, electrostatic actuators can be developed simply by using a pair of electrodes. Consequently, electrostatic actuators can be made light in mass, being more compatible with high-frequency operation.

2.01.2.2 Compatibility with MEMS

From an IC fabrication point of view, metal or semiconductor electrodes sitting on an insulating material (such as silicon dioxide) is a standard combination of materials that could be fabricated by the photolithography and etching processes. Thanks to the fact that electrostatic actuator is driven by voltage (charge) rather than by electrical current, power consumption is expected to be extremely small, and hence, small footprint ICs can be used as a driver. In fact, Digital Mirror Device (DMD) of Texas Instruments is electrostatically controlled by an integrated static random access memory (SRAM) of the complementary metal oxide semiconductor (CMOS) technology (Van Kessel *et al.* 1998). Electrostatic position feedback control is also used in the MEMS accelerometers and optical MEMS mirrors of Analog Devices Inc. (Core *et al.* 1993, Roessig *et al.* 2002).

2.01.3 Analytical Model for Electrostatic Actuation

2.01.3.1 Gap-closing Parallel Plate Mechanism

The actuation principle of an electrostatic actuator is understood by using a simple parallel plate model as shown in **Figure 7**. The drive electrode at the bottom is usually anchored to the substrate and electrically biased. On the contrary, the movable electrode on the top is usually suspended with an elastic hinge, through which it is electrically grounded to the same potential as the substrate. The suspended structures can easily be developed by the surface micromachining with a sacrificially removed material (Bustillo *et al.* 1998). Because of the electrically grounded movable electrode and the nature of the Coulomb force, the applied voltage always works to attract the electrodes regardless of the voltage polarity. Electrostatic repulsive force is available only when permanent electric charges exist on the electrode; the principle is called electret actuation, which will be discussed in Section 2.01.9.

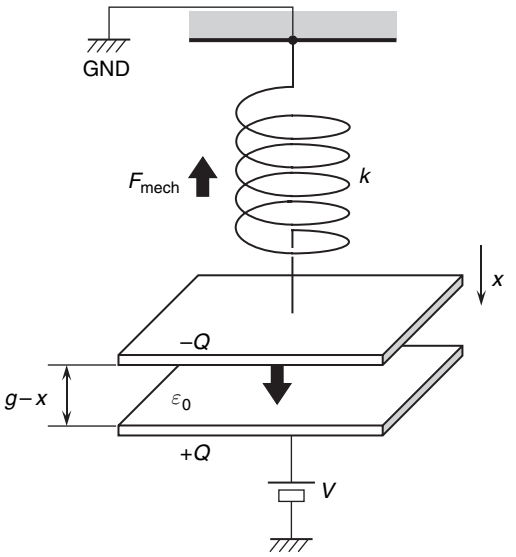


Figure 7 Analytical model for gap-closing parallel plate.

When the voltage is applied to the fixed electrode, the electrostatic attractive force moves the electrode to increase the electric capacitance. The spring suspending the movable electrode deflects until the restoring force reaches equilibrium with the electrostatic force. The electric capacitance C between the plates can be simply modeled as

$$C = \varepsilon_0 \frac{S}{g-x} \qquad [1]$$

where ε_0 is the dielectric constant of vacuum ($8.85 \times 10^{-12}\,\mathrm{F\,m^{-1}}$), S the area of the electrodes, g the initial gap between the electrodes, and x the displacement of the movable electrode toward the fixed one. Note that eqn [1] does not take the field concentration on the plate edge into account, but a contribution from a small fraction area S out from infinitely large plates is assumed. However, when the field concentration is considered, a good approximation for the capacitance is obtained by replacing the area S with a value one would get if the plates were extended artificially by a width equal to three-eighths of the gap (Feynman *et al.* 1964).

In the presence of an applied voltage V, the electrostatic attraction force can be calculated (see Section 2.01.12) as

$$F_{elec} = \frac{1}{2}\frac{dC}{dx}V^2 = \frac{1}{2}\varepsilon_0 \frac{S}{(g-x)^2}V^2 \qquad [2]$$

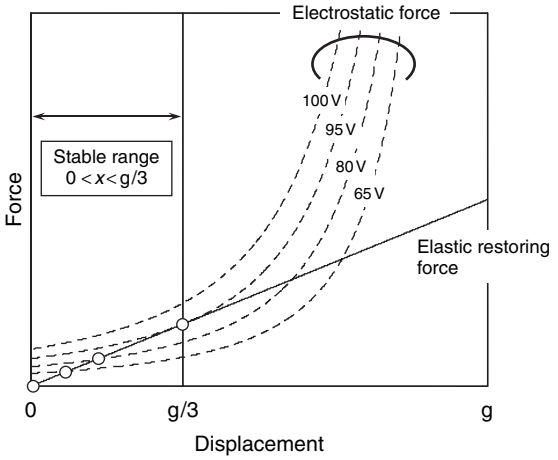


Figure 8 Comparison of electrostatic force and elastic restoring force.

On the contrary, the restoring force by the spring constant k is described simply by Hooke's law as

$$F_{mech} = k\,x \qquad [3]$$

Neglecting the effect of gravity, we equate eqns [2] and [3] to find the actuator displacement at the equilibrium condition for a given voltage as

$$k\,x = \frac{1}{2}\varepsilon_0 \frac{S}{(g-x)^2}V^2 \qquad [4]$$

Figure 8 plots the electrostatic force curve, eqn [2], and mechanical restoring force, eqn [4], as a function of displacement x. The parameters used in the calculation are $100\,\mu m \times 100\,\mu m$ for area S, $3\,\mu m$ for g, and $100\,\mathrm{N\,m^{-1}}$ for k. The intersection of the curves represents the equilibrium displacement at the given voltage. At voltage zero, the curves find a solution at $x = 0$. With increasing voltage, the static force curve moves upward, and the section moves toward greater values of x, implying that the electrostatic gap becomes small with increasing voltage. At some point where the electrostatic voltage exceeds the mechanical restoring force, the movable electrode is brought into contact with the counter-electrode to totally close the gap. This phenomenon is called electrostatic pull-in and is typically found in most electrostatic actuators. In the plot, the pull-in occurs at the point where the electrostatic force curve takes the restoring force line as a tangent.

Figure 9 plots the numerically computed values of x as a function of voltage V, using the same parameter set. Before the pull-in voltage (95 V), the plate displacement increases almost proportionally to

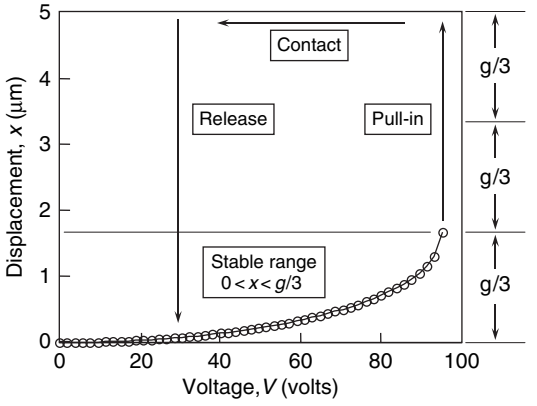


Figure 9 Parallel plate displacement as a function of applied voltage.

the square of voltage, V^2; in this range, the displacement can be stably controlled with negligible hysteresis. When the drive voltage reaches the pull-in voltage, the displacement leaps to close the gap. When the voltage is reduced, on the contrary, the plate remains in contact with the counterelectrode as long as the electrostatic attractive force is greater than the mechanical restoring force. The movable plate is released from the counterelectrode, when the electrostatic force becomes smaller than the restoring force.

2.01.3.2 Electrostatic Pull-in

It should be noted that the electrostatic pull-in occurs when the movable electrode comes to 2/3 of the initial gap. This phenomenon can be understood by rewriting eqn [4] in the following form:

$$\frac{1}{2}\varepsilon_0 \frac{k}{S} V^2 = x(g-x)^2 \qquad [5]$$

The left-hand side of the equation is a constant value determined by the voltage V, whereas the right-hand side varies by the geometrical factors x and g. In **Figure 10**, the pull-in point is shown by the local peak of the curve, where the first-order derivative is null:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left\{x(g-x)^2\right\} = (g-x)(g-3x) = 0 \qquad [6]$$

Therefore, the pull-in displacement is analytically found to be $x_{\text{pull-in}} = 1/3(g)$. In terms of the electrode gap, it corresponds to 2/3 of the initial gap. By substituting this value into eqn [5], we would obtain the pull-in voltage:

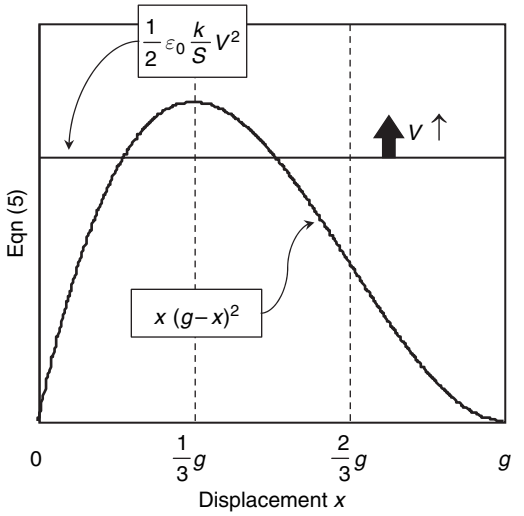


Figure 10 Pull-in instability.

$$V_{\text{pull-in}} = \sqrt{\frac{8\,k\,g^3}{27\,\varepsilon_0\,S}} \qquad [7]$$

The pull-in voltage depends on the geometrical parameters of the actuator, namely, the gap, the electrode area, and the spring shape, whereas the pull-in displacement is determined by the initial gap only.

The displacement after pull-in cannot be calculated by eqn [5], but the arrow in **Figure 9** is inserted by the *a priori* assumption that the displacement is limited by the initial electrostatic gap. Electrostatic actuators are usually designed to have a mechanical stopper somewhere on the way to the counterelectrode to avoid electrical short circuit upon the pull-in contact. In such a case, the maximum displacement should be limited by the position of the stopper. **Figure 11**

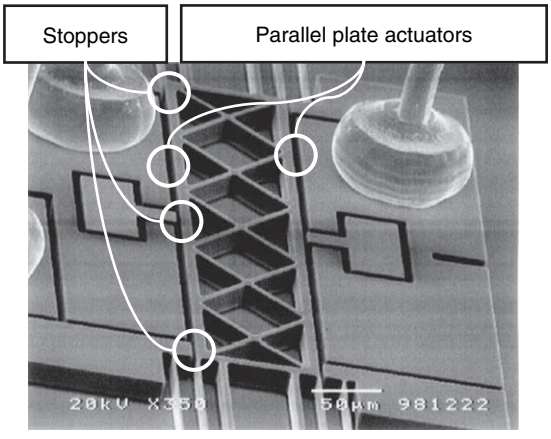


Figure 11 Laterally driven electrostatic actuator with integrated stoppers.

shows an example of parallel plate actuator whose motion is limited by the stoppers located on both the sides. Unlike the theoretical model shown in Figure 7, the structure moves laterally, but it is still well described by the parallel plate model. The SEM image was a part of a microactuator ladder structure made of a 20-μm-thick silicon-on-insulator (SOI) wafer. The movable structure seen in the middle (a meshed structure) was 500 μm long and 20 μm high (in the SOI thickness), and with a 10-μm-wide initial gap. The stoppers were embedded in the drive electrode (anchored to the substrate), and the stopper tips were designed to extend out by 5 μm. In this model, the stoppers are not electrically grounded because of the small area for wire bonding; ideally speaking, they should be grounded to avoid electrostatic charge-up problems.

2.01.3.3 Extending Travel Range

Applications may require long travel distance of the parallel plate actuators. Using a larger value for the initial gap is a solution; however, it comes with a penalty of increased drive voltage. In such a case, one may use the electrostatic comb-drive actuators, which will be discussed in Section 2.01.4. Leverage mechanism is a mechanical solution to extend the travel range. Kwon *et al.* (2006a) converted a rotational motion induced by an electrostatic actuator to produce a relatively large lateral stroke, as also discussed in Section 2.01.4.5. Takeshima *et al.* (1991) adopted a pantograph structure to magnify a small displacement into a large diagonal displacement.

Another solution to extend the travel range is to use a fast feedback control circuit, which enables operation beyond the pull-in displacement without changing the actuator design. J. Chen *et al.* (2004) used a voltage slightly higher than the pull-in voltage to the actuator when the displacement was small and linearly reduced the voltage as the actuator approached the desired displacement. Nadal-Guardia *et al.* (2002) used a current-source feedback control to drive an electrostatic diaphragm actuator and experimentally observed a controlled displacement beyond the pull-in displacement. Using a fixed capacitor in series with the electrostatic actuator is also helpful to delay the pull-in effect (Chan and Dutton 2000, Seeger and Crary 1997). The fixed capacitor and the actuator form a voltage divider. When the actuator plate is about to trip, the actuator capacitance increases, resulting in a reduced voltage across the actuator plates. However, the drawback of this method is that voltage required for operation becomes higher than the pull-in voltage under the open-loop control.

2.01.3.4 Spring Constant of Bending Beam

According to the beam theory (Gere and Timoshenko 1997, Timoshenko and Goodier 1970), beam displacement u under distributed force $q(x)$, as shown in Figure 12, is modeled simply by the following equation:

$$\frac{d^4u}{dx^4} = \frac{q(x)}{EI} \quad [8]$$

where E is the Young's modulus (160 GPa for single-crystal silicon along the <110> direction and 130 G for <100>) and I the second moment of area. When the beam section is rectangular with width w (lateral) and height h (vertical) and is forced to bend in the direction of the thickness (out of plane), the moment of inertia is written as

$$I = \frac{w h^3}{12} \quad [9]$$

When the beam is bent laterally in the direction of the width, the moment of inertia is

$$I = \frac{w^3 h}{12}$$

We use the bending beam model to derive a spring constant of a cantilever with a fixed end and a free tip. When a concentrated force P is applied downward on the cantilever tip at $x = l$, as shown in Figure 13, the beam also experiences a repulsive force in the upward direction from the supporting wall at $x = 0$. Eqn [8] is rewritten as

$$EI\frac{d^4u}{dx^4} = P\delta(x) + P\delta(x-l) \quad [10]$$

where $\delta(x-a)$ is the delta function that is defined as follows:

$$\delta(x-a) = \begin{cases} 1\,(x=a) \\ 0\,(x\neq a) \end{cases} \text{ and } \int_{-\infty}^{+\infty} \delta(x-a) = 1 \quad [11]$$

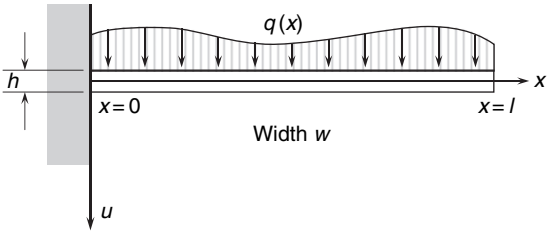


Figure 12 Bending beam under distributed force.

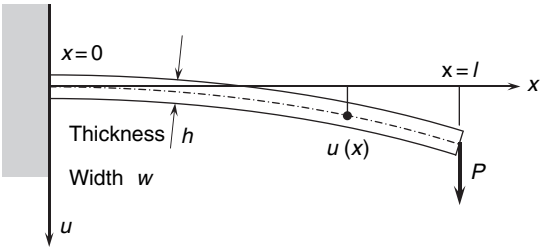


Figure 13 Fixed-end cantilever.

Note that the force and the displacement are defined to be positive in the downward direction. By integrating eqn [10] with respect to x, one would obtain the series of differential equations as follows:

$$EI\frac{d^3u}{dx^3} = -P + C_1 = -V_x \,(\text{shearing force}) \qquad [12]$$

$$EI\frac{d^2u}{dx^2} = (-P + C_1)x + C_2 = -M_x \,(\text{bending moment}) \qquad [13]$$

$$EI\frac{du}{dx} = \frac{1}{2}(-P + C_1)x^2 + C_2x + C_3 \qquad [14]$$

and

$$EI\,u(x) = \frac{1}{6}(-P + C_1)x^3 + \frac{C_2}{2}x^2 + C_3x + C_4 \qquad [15]$$

By using the boundary conditions that $u = 0$ and $du/dx = 0$ at $x = 0$ in eqns [14] and [15], respectively, we find $C_3 = C_4 = 0$. Because the shearing force $EI\left(d^3u/dx^3\right)$ is equal to $-P$ everywhere in the beam, $C_1 = 0$. In addition, no moment is given to the beam's free end, $EI\frac{d^2u}{dx^2}\big|_{x=l}$ and thus $C_2 = P/l$. Therefore, the equation for the cantilever's deflection is written as

$$u(x) = \frac{P}{6EI}\left(3lx^2 - x^3\right) \qquad [16]$$

At the cantilever tip, the displacement is calculated to be $u(l) = Pl^3/3EI$. By substituting this in Hooke's law, one would obtain the spring constant k to be $3EI/l^3$.

A similar procedure can be used to derive the spring constant for various kinds of boundary conditions. Some representative beams that are frequently used in MEMS actuators are listed in **Figure 14**.

2.01.3.5 Electrostatic Actuation of Distributed Gap System

The analytical model presented in the previous section was good for a suspended mass model, where electrostatic force acts only on a limited place such as the cantilever tip. When voltage is applied to the

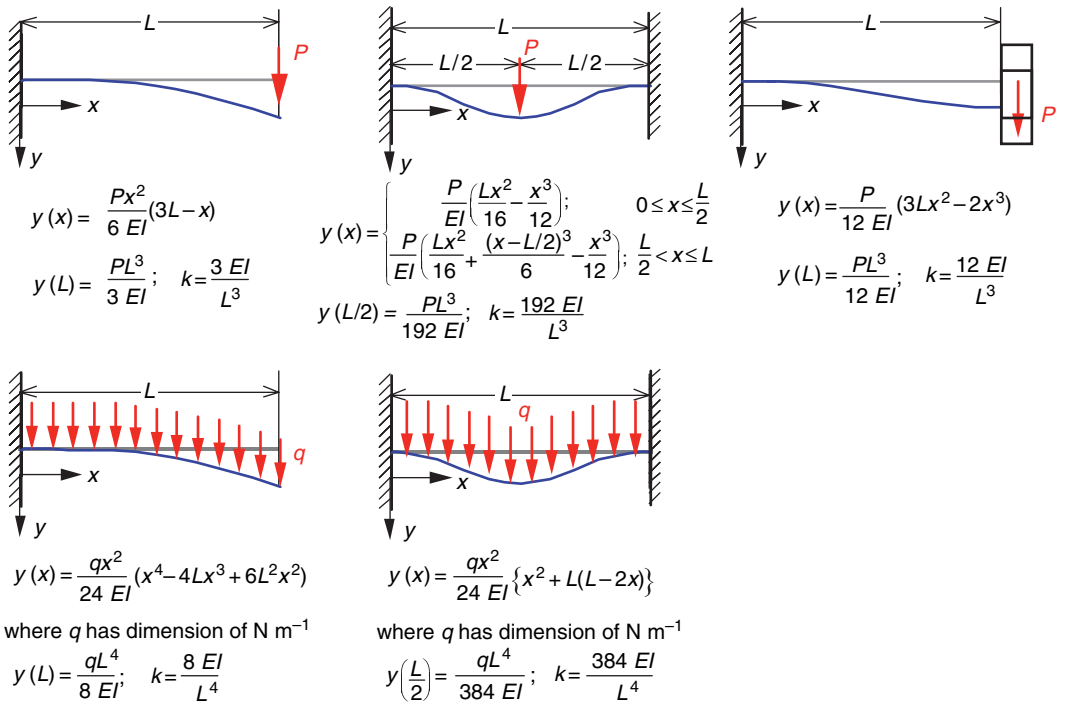


Figure 14 Spring constant of suspending beams of various boundary conditions.

entire cantilever or bridge, on the contrary, one should take the distributed force into account. Fortunately, the analytical model for such systems is not at all complicated in form, but it has already been presented by eqn [8]. Assuming a cantilever or bridge of width w suspended at the initial height g from the substrate surface and that it is biased to a voltage V with respect to the substrate, we only replace the distributed pressure $q(x)$ with the electrostatic pressure, and hence, eqn [8] is written as

$$EI\frac{d^4u}{dx^4} = \frac{1}{2}\varepsilon_0\frac{w}{g-u(x)}V^2 \qquad [17]$$

Despite the simplicity of the form, analytical solution is difficult to present. Readers are advised to use a numerical simulation module (such as the Runge–Kutta method) from commercially available mathematical software.

Figure 15 shows the results of numerical simulation using Mathcad™ (www.mathsoft.com); a simulation program is also shown in **Figure 16**. The simulation was performed with a small voltage and repeated by raising the voltage by small increments until the computation failed to converge, where pull-in instability was defined. The profile shown in **Figure 15** is of the last one before such pull-in.

The distributed gap system can also be applied to an actuator with a tapered electrostatic gap, as shown in **Figure 17** (Saruta *et al.* 2002). The maximum displacement can be made large by designing an electrode that has a wider gap near the end of an electrostatic cantilever or in the center part of a bridge.

While most finite element method (FEM)-type MEMS simulators on electrostatic force take a long time (more than an hour, depending on the complexity of structure) to finish calculation at one voltage, the analytical model presented was finished instantly. However, the analytical model discussed here holds good only for an actuator before electrostatic pull-in. The boundary condition changes after a cantilever or a bridge is brought into contact. Post pull-in simulation was studied by Gorthi *et al.* (2006).

2.01.4 Comb-drive Actuators

2.01.4.1 Sliding Plate Model

The electrostatic attractive force acting on the parallel plate actuator rapidly increases as the gap is closed, leading to quite a large nonlinear displacement–voltage characteristic. It also suffers from the pull-in instability that limits the maximum stable displacement within one-third of the initial gap. On the contrary, the comb-drive mechanism, invented by W. Tang (Tang *et al.* 1989), linearly increases the overlap between the comb fingers, resulting in an extended stable range of displacement as well as margin to the pull-in instability.

Figure 18 shows an example of the comb-drive mechanism made by the bulk silicon micromachining on an SOI wafer; the device was developed to control the submicron gap between the silicon tips, where a tunneling current flowed (Toshiyoshi *et al.* 1999). The square frame with many release holes was suspended with four pairs of folded beam suspensions. On the inner right-hand side of the frame, a comb-drive mechanism was integrated to move the frame in the direction to close the tip-to-tip distance. The bar connecting the folding points of the suspensions was used to suppress the in-plane rotation of the frame, by which the comb-to-comb short circuit was avoided. For designing the suspension rigidity, one may refer to **Figure 14** to find $k = 12EI/l^3$ for one cantilever with its tip rotation restricted. Because the folded suspension shown in **Figure 18** consists of four parallel pairs of two suspensions in series, the total spring constant is $k_{\text{total}} = 4 \times (1/2) \times k = 24EI/l^3$.

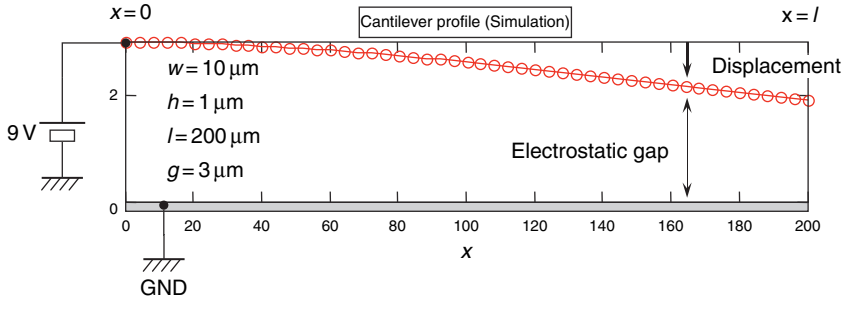


Figure 15 Simulation result of electrostatic cantilever.

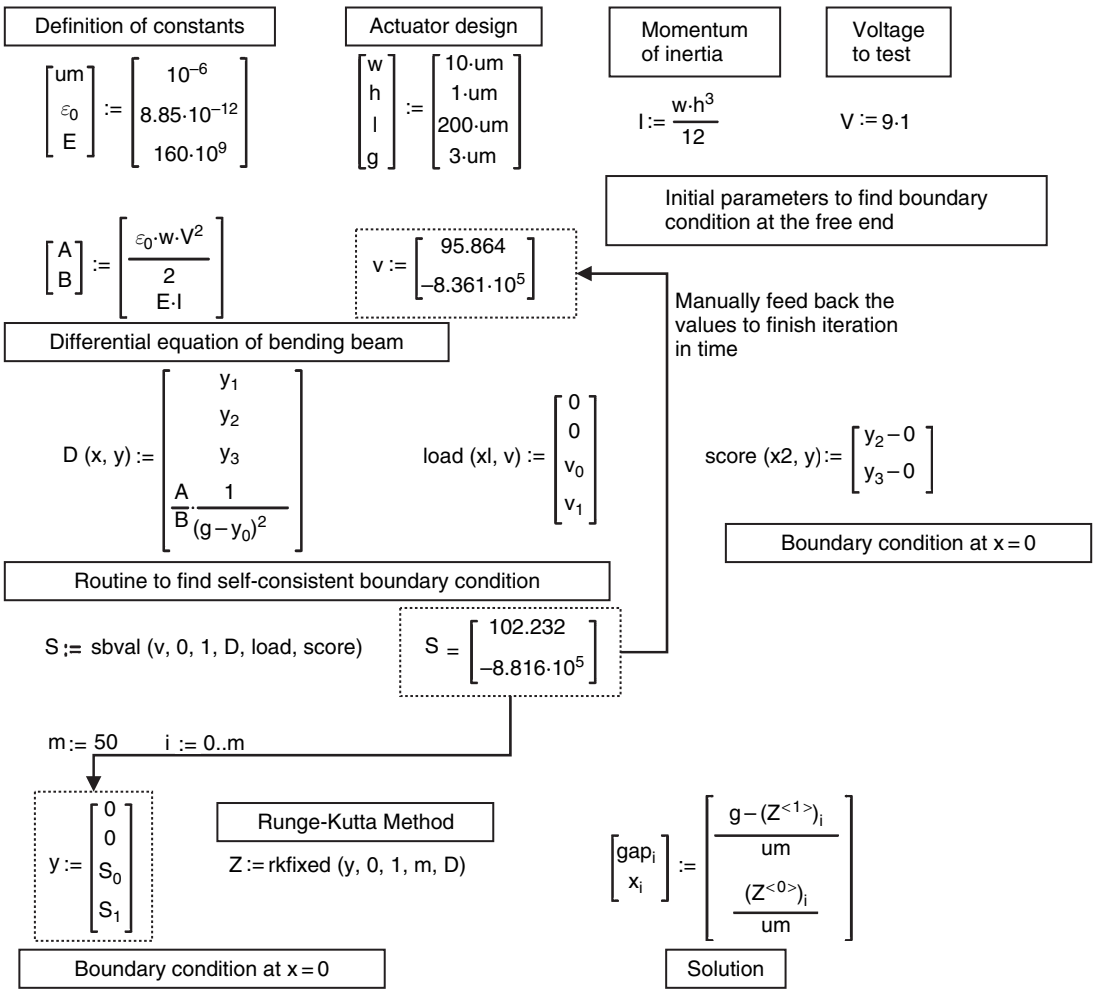


Figure 16 MATHCAD™ program for electrostatic cantilever.

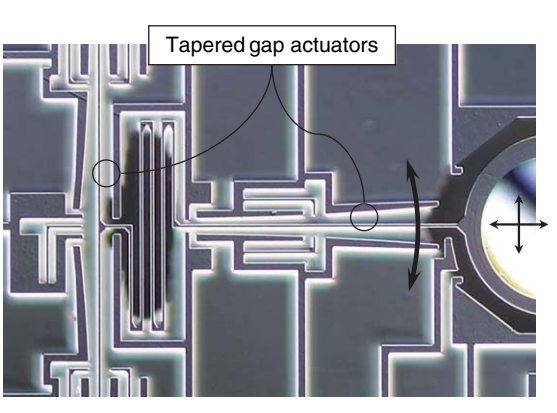


Figure 17 Electrostatic actuator with tapered gap.

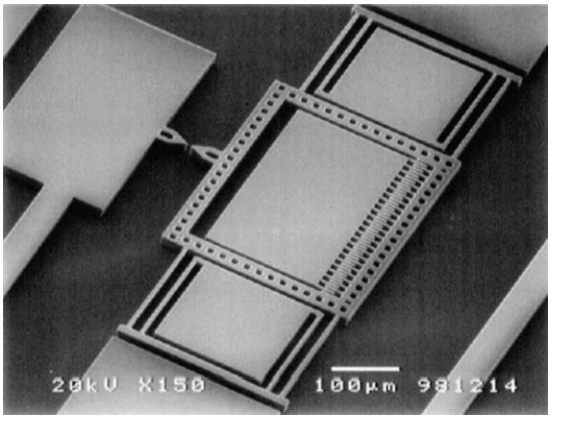


Figure 18 Electrostatic comb-drive actuator.

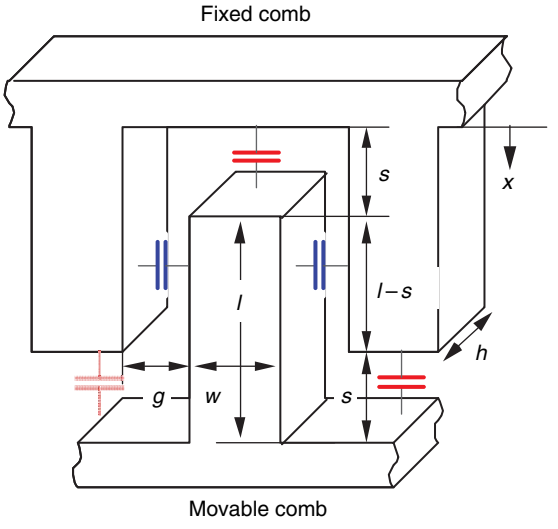


Figure 19 Analytical model (unit cell) for electrostatic comb-drive mechasnism.

Figure 19 illustrates a unit cell of the comb-drive electrodes of comb length l, width w, gap on the side g, gap on the comb tip s, and comb height h. The initial overlap of the opposing tips is written as $l = s$ in this model. The tips are usually made without having overlap ($l = s$), such that the effective gap g, after the comb is engaged in operation, can be made smaller than the photolithography resolution, enabling electrostatic actuation at lower voltages. Having no overlap is also beneficial for extending the travel distance of the movable comb.

When the upper comb in Figure 19 is displaced downward by travel distance x, the capacitor made at the comb tip is written by using the gap-closing parallel plate model:

$$C_{\text{tip}} = \varepsilon_0 \frac{w \times h}{s - x} \qquad [18]$$

On the contrary, the capacitor on the comb side can also be represented in a similar manner but with a sliding motion:

$$C_{\text{side}} = \varepsilon_0 \frac{h(l - s + x)}{g} \qquad [19]$$

The capacitor in the unit cell contains contribution from two side walls and two tips; thus,

$$C_{\text{unit}} = 2\left(C_{\text{tip}} + C_{\text{side}}\right) = 2\varepsilon_0 \left(\frac{w \times h}{s - x} + \frac{h(l - s + x)}{g}\right) \qquad [20]$$

Assuming that the comb fingers are arranged in the span of W, the number of fingers is calculated to be $N = W/2(w + g)$. Therefore, the total capacitor of an actuator is given as

$$C_{\text{unit}} = \varepsilon_0 \left(\frac{w \times h}{s - x} + \frac{h(l - s + x)}{g}\right) \frac{W}{w + g} \qquad [21]$$

For calculating the electrostatic force of the comb-drive mechanism, one may substitute eqn [21] into eqn [2]; hence,

$$F_{\text{unit}} = \frac{1}{2}\varepsilon_0 \left(\frac{w \times h}{(s - x)^2} + \frac{h}{g}\right) \frac{W}{w + g} V^2 \qquad [22]$$

It should be noted that the fringing effect (field concentration) has been neglected in this first-order approximation model. The effect of the substrate (ground plane) is also not taken into account.

Figure 20 shows the capacitance C_{unit} as a function of comb position x, calculated by using the following parameters: $h = 3\,\mu\text{m}$, $g = 1\,\mu\text{m}$, $w = 2\,\mu\text{m}$, $s = l = 6\,\mu\text{m}$, and $W = 100\,\mu\text{m}$. In a small displacement range of up to $4\,\mu\text{m}$, the capacitance increases almost linearly with the displacement, thanks to the sliding component C_{side}. When the comb tip comes to close the gap, the capacitance rapidly increases by the contribution from the gap-closing effect of C_{tip}.

Recalling the fact that the electrostatic force is proportional to the first-order derivative of the capacitance with respect to the position x, the linear range in the plot implies that the comb-drive mechanism is rather free from the pull-in effect. In an extreme case with a deep trench, $s >> l$, it is also possible to eliminate the effect of the capacitor at the tip, C_{tip}, and the electrostatic force becomes

$$F_{\text{unit}} = \frac{1}{2}\varepsilon_0 \frac{h}{g} \frac{W}{w + g} V^2 \qquad [23]$$

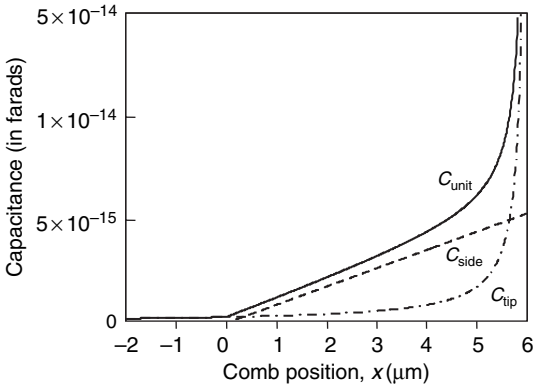


Figure 20 Capacitance of comb-drive as a function of electrode position.

Eqn [23] implies that the electrostatic force is independent of the comb's geometry, namely, the overlap of the comb electrodes, but it is proportional to the square of voltage. It should also be mentioned that using a small gap g does not lead to the limited travel distance but allows us to operate with lower voltages.

2.01.4.2 Electrical Ground Potential

A voltage may directly be applied to a suspended structure that is intended to move in the direction normal to the substrate surface, provided that the substrate is electrically grounded. A diaphragm structure for the electrically tunable Fabry–Perot interferometer is one such application (Ford *et al.* 1998). However, suspended structures designed to move in the lateral direction should be electrically grounded to avoid the vertical motion. In such a case, drive voltages are applied to the fixed electrodes.

2.01.4.3 Linearization by Differential Voltage Drive

An interesting feature of the comb-drive mechanism is that the displacement can be linearized in part with respect to the drive voltage. **Figure 21** illustrates the comb resonator developed by Tang *et al.* (1989) (shown in **Figure 6**). The movable structure is electrically grounded, and the electrodes on the left- and the right-hand side are biased to V_L and V_R, respectively. When a common DC bias voltage V_{bias} is superposed with a differential voltage v, one would write

$$V_L = V_{bias} - v \quad [24]$$

$$V_R = V_{bias} + v \quad [25]$$

where the differential voltage v is controlled to be smaller than the bias voltage ($|v| < |V_{bias}|$).

Neglecting the contribution from the comb tips, we write the force acting on each side of the actuator by using eqn [23]:

$$F_L = A(V_{bias} - v)^2 \quad [26]$$

$$F_R = A(V_{bias} + v)^2 \quad [27]$$

where $A = \varepsilon_0(bW/2g(w+g))$ is a geometrical factor commonly applied to both sides. Therefore, the net force acting on the movable plate is

$$\begin{aligned} F_{net} &= F_R - F_L \\ &= A(V_{bias} + v)^2 - A(V_{bias} - v)^2 \\ &= 4A \cdot V_{bias} \cdot v \end{aligned} \quad [28]$$

Thus, the force is found to have been linearized in terms of the differential voltage v. Using the same parameter set used to plot **Figure 20**, we examined the effect at different bias voltages as shown in **Figure 22**. It is clearly seen that the displacement is proportional to the differential voltage under the condition $|v| < |V_{bias}|$. Eqn [28] also implies that the bias voltage V_{bias} is used as a gain to tune the sensitivity to the differential voltage v. **Figure 22** also shows the validity of the gain control without degrading the linearity. Parallel plate actuators can also be linearized in the same manner, but the linear range is small due to the asymmetry of the geometrical factor A.

The differential voltage scheme shown in **Figure 21** requires complicated driver circuits with two AC voltage sources superposed on a DC bias voltage by using, for instance, analog operational amplifiers. In fact, an equivalent function can be made simply by using only one AC voltage source as

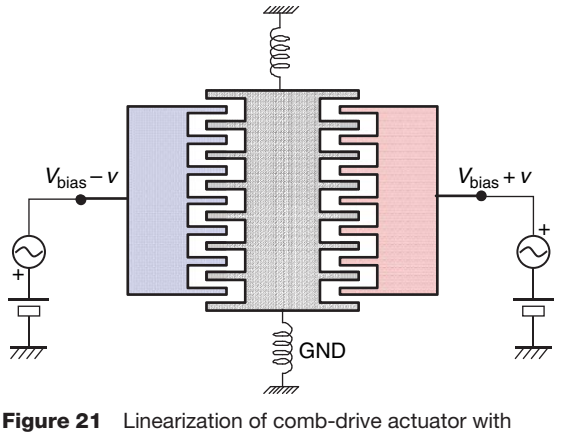


Figure 21 Linearization of comb-drive actuator with differential voltages.

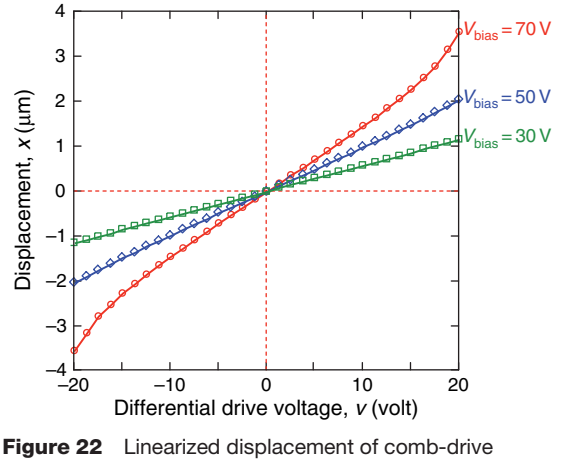


Figure 22 Linearized displacement of comb-drive mechanism.

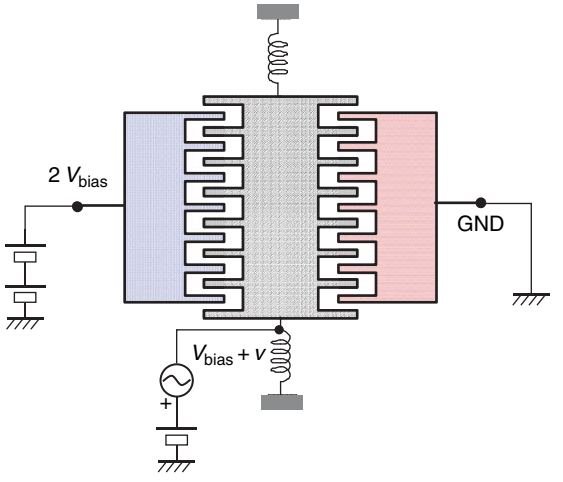


Figure 23 Differential voltage drive with only one AC voltage source.

shown in **Figure 23**, where voltages of $2V_{bias}$, $V_{bias} + v$, and 0 are applied to the left-hand-side, the movable, and the right-hand-side electrode, respectively. Thanks to the fact that electrostatic attractive force is independent of the voltage polarity, the net drive voltage on each side of the movable electrode is

$$\text{Left: } |(V_{bias} + v) - 2V_{bias}| = V_{bias} - v \text{ and} \quad [29]$$

$$\text{Right: } |0 - (V_{bias} + v)| = V_{bias} + v \quad [30]$$

Hence, the new method is found to be equivalent to **Figure 21** in terms of electrostatic force. When the suspended structure is in a close vicinity to the substrate surface, it would also be attracted toward the substrate because of the applied bias voltage. To avoid the problem, an equipotential $V_{bias} + v$ should be given to the substrate surface or a shield layer inserted in between the substrate and the suspended structure. The latter solution can be made by using a polysilicon surface micromachining technique with more than two polysilicon layers.

2.01.4.4 Bistable Digital Actuation Model

The comb-drive mechanism was originally developed to improve the linearity in displacement. As discussed in the analytical study, straight comb fingers in the deep trench would be helpful for that purpose. On the contrary, the nonlinearity can also be intentionally designed by altering the shape of the comb finger into a hammer-head style as shown in **Figure 24(a)** (Kwon *et al.* 2005). Kwon *et al.* with GIST reported the hammer-head comb drive to realize a bistable motion that could be mechanically latched with the aid of the buckling of suspension. A schematic of the actuator is shown in **Figure 24(b)**. The movable comb electrode was designed to have no overlap with the counterelectrode at the rest position. Once a pulse voltage was applied, the movable comb was pulled inside the counter-comb, and it remained engaged without voltage, thanks to the pressing force induced by the buckled suspension. To disengage, a pulse voltage is applied again to the counterelectrode to pull out the comb. Unlike the standard comb design shown in the previous sections, the hammer-head comb uses the electrostatic force acting on the corners.

2.01.4.5 Rotational Comb-drive Actuators

Combination of the comb-drive mechanism with a lever can be used to generate a relatively large stroke in microactuators. Kwon *et al.* reported a rotational comb-drive mechanism with coaxial comb fingers developed by the bulk silicon micromachining on

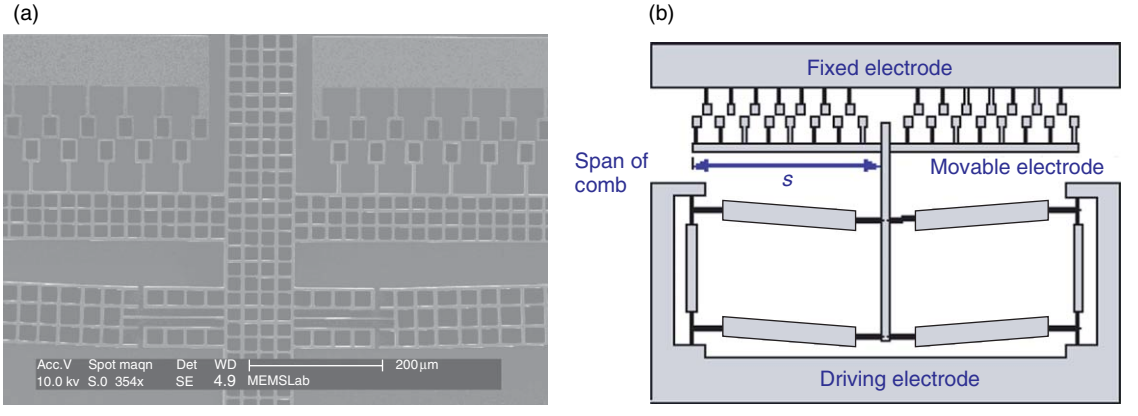


Figure 24 Bistable comb-drive mechanism with hammer-head electrode. (Source: Kwon H N, Hwang I-H, Lee J-H 2005 A pulse-operating electrostatic microactuator for bi-stable latching. *J. Micromech. Microeng.* **15**, 1511–16.)

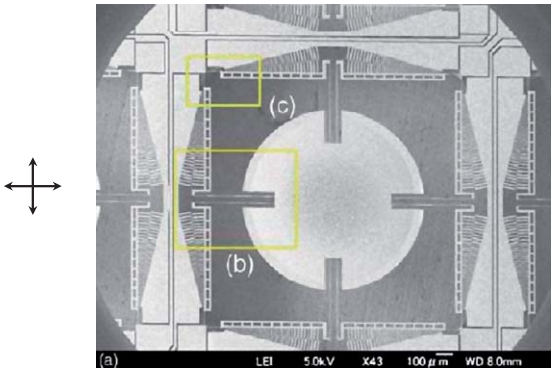


Figure 25 Coaxial comb-drive mechanism to convert angular motion into lateral stroke. (Source: Kwon H N, Lee J-H, Takahashi K, Toshiyoshi H 2006a Micro XY stages with spider-leg actuator for two-dimensional optical scanning. *Sens. Actuators* **A130-131**, 468–77.)

an SOI wafer, as shown in **Figure 25** (Kwon *et al.* 2006a). They used a pair of opposing coaxial comb drive and converted the rotational motion into a lateral motion by using the flexures. Maximum displacement of 55 μm was obtained by a DC voltage of 40 V. They used a rigid bar for holding the comb fingers and a thin flexure (1 μm wide and 90 μm long) for the rotation support; thanks to the rigidity contrast, the comb electrodes were operated stably despite the small electrostatic gap of 2.5 μm. Kwon used an orthogonal set of such unit to develop a micro *XY* stage (Kwon *et al.* 2006a) with an integrated silicon lens. They demonstrated a spatial light modulation of infrared by using the $f-\theta$ lens scanner mechanism.

2.01.4.6 Micro *XY* Stage with Comb-drive Mechanism

An *XY* stage with a nested comb-drive mechanism, as shown in **Figure 26** was reported by Takahashi *et al.* (2007, 2006a). They used a double-gimbal structure of an SOI wafer to accommodate two orthogonal sets of comb-drive mechanisms to realize a cross-talk-free *XY* motion. Thanks to their new device layout technique that split the electrostatic elements (electrodes) and elastic elements (suspensions) into the SOI and the recessed substrate, respectively, the device was relatively made in a small footprint compared with the previously reported devices (Kim *et al.* 2003, Kwon and Lee 2002, Kwon *et al.* 2006b, Sarajlic *et al.* 2005). For electrical interconnection to the inner stage, they used the elastic suspension as well as an additional small tether of negligible added rigidity, as

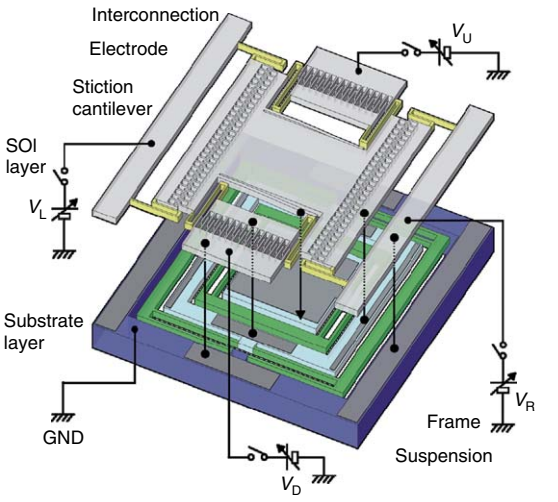


Figure 26 *XY* stage with a nested pair of comb-drive actuators. (Sources: Takahashi K, Mita M, Fujita H, Toshiyoshi H 2006a A high fill-factor comb-driven XY-stage with topological layer switch architecture. *IEICE Electron. Express* **3(9)**, 197–202 and Takahashi T, Mita M, Motohara K, Kobayashi N, Kashikawa N, Fujita H, Toshiyoshi H 2006b Electrostatically addressable gatefold micro-shutter arrays for astronomical infrared spectrograph. *Proc. Asia Pacific Conf. Transducers, Micro-Nano Technology*, Marina-Mandarin Hotel, Singapore, June 25–28, 2006.)

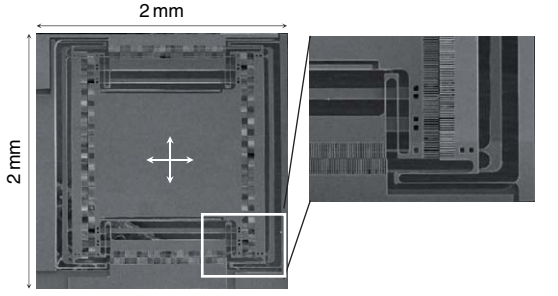


Figure 27 Schematic of the *XY*-stage combs. (Sources: Takahashi K, Mita M, Fujita H, Toshiyoshi H 2006a A high fill-factor comb-driven XY-stage with topological layer switch architecture. *IEICE Electron. Express* **3(9)**, 197–202 and Takahashi T, Mita M, Motohara K, Kobayashi N, Kashikawa N, Fujita H, Toshiyoshi H 2006b Electrostatically addressable gatefold micro-shutter arrays for astronomical infrared spectrograph. *Proc. Asia Pacific Conf. Transducers, Micro-Nano Technology, Marina-Mandarin Hotel, Singapore, June 25–28, 2006.)*

shown in **Figure 27**. With a 2 mm × 2 mm device footprint, they observed maximum 19 μm in the *X*-direction and 23 μm in the *Y*-direction with a voltage of 110 V_{DC}.

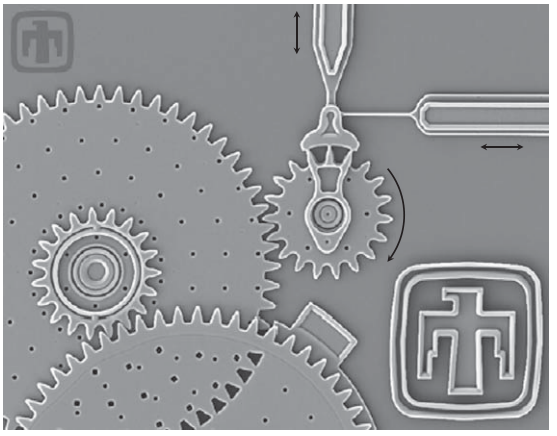


Figure 28 Micromechanism to convert lateral stroke into rotational motion. (Courtesy: Sandia National Laboratories, SUMMiT™ Technologies, www.mems.sandia.gov.)

2.01.4.7 Conversion Mechanism

Sandia National Laboratories reported a micromechanism to convert the stroke of comb-drive actuators into a perpetual rotational motion, as shown in **Figure 28**. A pair of comb-drive actuators was arranged in the orthogonal direction. Their stroke was intentionally preset to be out of phase by 90°, such that they worked as a two-cylinder engine to rotate a tiny gear that functioned as a crankshaft. The rotation of the gear was transferred to a larger gear that worked as a rotation stage. The mechanism can also be used in the linear racks of long stroke and in the 3D self-assembly. They developed a surface micromachining technique of five polysilicon layers called SUMMiT V (Sandia Ultra-planar, Multi-Level MEMS Technology V) to produce the complex mechanisms (Sniegowski and Garcia 1996, http://mems.sandia.gov/).

2.01.5 Torsion Mirror Actuators

2.01.5.1 Gap-closing Rotational Motion

Rotational motion is used for, for instance, torsion mirrors that spatially modulate light beams in free space. An electrostatic version of the torsion mirror was reported in the review article by Petersen (1982). The mirror was operated at the resonant frequency for large scan angle as a galvano mirror. In the 1990s or later, the torsion mirror mechanism was commonly used to statically scan the light beams in free space for an optical switching application (Aksyuk *et al.* 2003, Brosnihan *et al.* 2003, Dadap *et al.* 2003, Toshiyoshi and Fujita 1996, Toshiyoshi *et al.* 2001) and for image projection display (Van Kessel *et al.* 1998).

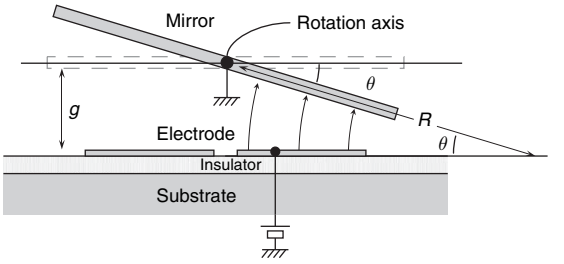


Figure 29 Analytical model for electrostatic torsion mirror actuator.

An analytical model for the torsion mirror actuator is illustrated in **Figure 29**. Usually, the mirror plate is suspended at the center axis with a pair of torsion bars, and drive electrodes are located underneath, on the counter-substrate. The drive electrodes are made shorter in width than the mirror to avoid short circuit upon the mirror's pull-in contact. A simple analytical model for the electrostatic torsion mirror was presented by Toshiyoshi *et al.* (Toshiyoshi and Fujita 1996, Toshiyoshi *et al.* 2001), by representing the electric field with the coaxial arcs, neglecting the field concentration on the mirror edges. For more detailed analysis, one should use the numerical simulation software with an electromechanical co-solver module (www.conventor.com/, http://intellisense-software.com/). When the mirror of width W and length L (taken normal to the page) is rotated to the counterelectrode by angle θ, the distance between the arc center and the mirror axis is

$$R = \frac{g}{\sin\theta} \qquad [31]$$

where g is the elevation height of the mirror axis measured from the substrate surface. Assuming that the electrostatic field in between the mirror plate and the counterelectrode is represented by an arc, the length of the arc located at position x on the mirror is $(R-x)\theta$. Therefore, the electrostatic field strength is

$$E(x) = \frac{V}{(R-x)\theta} = \frac{V}{\left(\frac{g}{\sin\theta} - x\right)\theta} \qquad [32]$$

where V is the applied voltage. The equation is also valid when $\theta < 0$ (mirror angle of elevation). At an extreme case of $\theta \to 0$, the field is equivalent to that of a parallel plate actuator.

The electrostatic pressure acting on the mirror plate, $\varepsilon_0 E^2/2$, is integrated in the mirror width to produce an electrostatic torque.

$$T_{\text{elec}}(V) = \int_0^W \frac{1}{2}\varepsilon_0 \frac{V^2}{\left[\left(\frac{g}{\sin\theta} - x\right)\theta\right]^2} L\, x\, \mathrm{d}x$$
$$= \frac{\varepsilon_0 L V^2}{2\theta^2}\left\{\frac{W\sin\theta}{g - W\sin\theta} + \log\left(1 - \frac{W}{g}\sin\theta\right)\right\} \quad [33]$$

For $|\theta| << 1$, we take an expanded form

$$T_{\text{elec}}(V) = \frac{\varepsilon_0 L V^2}{2\theta^2}\left\{\frac{1}{2}\left(\frac{W\sin\theta}{g}\right)^2 + \frac{2}{3}\left(\frac{W\sin\theta}{g}\right)^3\right\} + O(\theta^2) \quad [34]$$

to avoid the divergence in numerical calculation.

On the contrary, the restoring torque of the torsion beams of a rectangular cross section is given as:

$$T_{\text{mech}} = 2 \times \frac{G\, a\, b^3}{3\, l}\theta\left\{1 - \frac{192}{\pi^5}\frac{b}{a}\tanh\left(\frac{\pi a}{2b}\right)\right\} \quad [35]$$

where $a \times b$ is the section area of the beam, taking $a > b$, and G is the shear stress of rigidity (62.1 GPa for silicon). By numerically computing the equation $T_{\text{elec}}(V) = T_{\text{mech}}$, one will obtain the mirror angle as a function of drive voltage. As torsion mirror is a variation of parallel plate mechanism, the motion is restricted by the electrostatic pull-in effect. For a small angle design, the maximum mirror angle is limited by the pull-in contact at θ/g. Like a parallel plate mechanism, coincidentally, the pull-in contact is nearly 1/3 of the contact angle, which has been experimentally observed (Isamoto *et al.* 2004).

A large torsion angle upward of 90° can also be possible by using flexible torsion bars. Takahashi *et al.* reported a thin single-crystal silicon beam only 0.5 μm thick, 2 μm wide, and 500 μm long to support a 500 μm × 1000 μm plate that tilted to almost 90°, as shown in **Figure 30**. They used the sidewall of the through-hole underneath the torsion mirror plate and manufactured an electrically addressable microshutter array for an infrared spectroscope (Takahashi *et al.* 2006b).

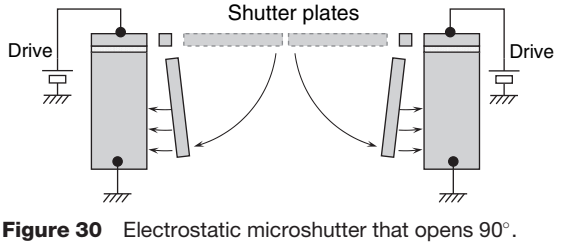


Figure 30 Electrostatic microshutter that opens 90°.

2.01.5.2 Torsional Motion Combined with Vertical Motion

When the suspensions have small rigidity in the up/down motion of the plate, the torsional motion is accompanied by the vertical piston motion of the plate. By taking advantage of the mechanical linearity in a small displacement range, we simply use simultaneous equations that take the vertical and the rotational motion into account. For the torsional motion, eqns [33] and [35] are modified to be

$$\begin{cases} T_{\text{elec}}(\theta, z, V) = \displaystyle\int_0^W \frac{1}{2}\varepsilon_0 \frac{V^2}{\left[\left(\frac{g-z}{\sin\theta} - x\right)\theta\right]^2} L\, x\, \mathrm{d}x \\ T_{\text{mech}}(\theta) = 2 \times \dfrac{G\, a\, b^3}{3\, l}\theta\left\{1 - \dfrac{192}{\pi^5}\dfrac{b}{a}\tanh\left(\dfrac{\pi a}{2b}\right)\right\} \end{cases} \quad [36]$$

The plate height has been changed from g to $g - z$, where z is the mean downward displacement of the plate. The force balance in the vertical direction should be simultaneously considered. The electrostatic force can be obtained by integrating the cosine component of the electrostatic pressure acting normal onto the plate.

$$\begin{cases} F_{\text{elec}}(\theta, z, V) = \displaystyle\int_0^W \frac{1}{2}\varepsilon_0 \frac{V^2}{\left[\left(\frac{g-z}{\sin\theta} - x\right)\theta\right]^2} L\cos\theta\, \mathrm{d}x, \\ F_z(z) = k_z\, z \end{cases} \quad [37]$$

where k_z is the rigidity of the suspension in the vertical direction.

By simultaneously solving eqns [36] and [37], one would obtain the angle θ and the downward displacement z at an equilibrium point under a given voltage. In a small angle operation (~1°), the analytical model agreed fairly well with the experimental results, as reported by Isamoto *et al.* (2004).

2.01.5.3 Double-gimbal Torsion Mirror

An analytical model that was extended to a 2D torsion mirror was reported by Toshiyoshi *et al.* (Toshiyoshi and Fujita 1996). **Figure 31** shows the SEM image of the 2D scanner made by the silicon surface micromachining technique MUMPs (www.memscap.com/). The mirror plate (400 μm × 400 μm in area and 1.5 μm in thickness) was suspended by the double-gimbal structure that consisted of two pairs of torsion bars of 2 μm wide, 1.5 μm thick, and 200 μm long. The micromirror and the frame were suspended above the

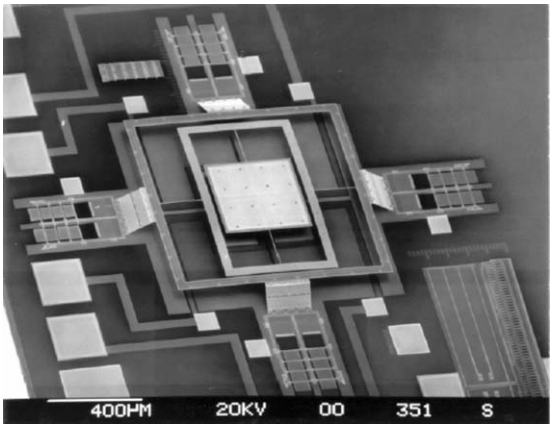


Figure 31 Double-gimbal electrostatic torsion mirror. (Source: Toshiyoshi H, Piyawattanametha W, Chan C-T, Wu M C 2001 Linearization of electrostatically actuated surface micromachined 2-D optical scanner. *J. Miroelectromech. Syst.* **10(2)**, 205–14.)

substrate by an air gap of 72 µm, using the microelevator self-assembly (MESA) technique developed by Li *et al.* (Fan *et al.* 1997). All the suspended structures were electrically grounded, and the scan angles were electrically controlled by using four voltages applied to the quadrant electrodes underneath the mirror. In recent study, Tsai and Wu (2005) reported a gimballess 2D scanner by using the Sandia National Laboratories' SUMMiT V Process.

For analyzing the 2D tilting motion, the coordinate conversion is used to find a position on the mirror after 2D rotation. The origin of the coordinate XYZ is set at the center of the mirror, and it is rotated with the mirror around the Y-axis by angle ϕ, leaving a new coordinate frame $X'Y'Z'$ at the initial place. After the ϕ rotation, A point (x, y, z) on the mirror is transferred to a new coordinate (x', y', z') in the frame $X'Y'Z'$. The new coordinate is written by using a transfer matrix of rotation as

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{M}_\phi \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad [38]$$

For the θ scan around the X axis, we use

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{M}_\theta \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad [39]$$

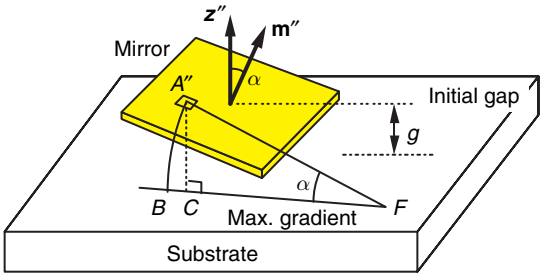


Figure 32 2D torsion mirror after rotation.

Due to the rotation mechanism shown in Figure 32, we first use the ϕ rotation and then θ rotation, i.e., the transfer matrix would be

$$\mathbf{M} = \mathbf{M}_\theta \mathbf{M}_\phi = \begin{pmatrix} \cos\phi & 0 & \sin\phi \\ \sin\theta\sin\phi & \cos\theta & -\sin\theta\cos\phi \\ -\cos\theta\sin\phi & \sin\theta & \cos\theta\cos\phi \end{pmatrix} \quad [40]$$

By using the transfer matrix [40], we calculate an element of small area (dS) on the mirror at point $A = (x, y, 0)$ and its normal vector $\mathbf{m} = (0, 0, 1)$ to be transferred to

$$A'' = (x\cos\phi, x\sin\theta\sin\phi + y\cos\theta, -x\cos\theta\sin\phi + y\sin\theta) \quad [41]$$

and

$$\mathrm{m}'' = (\sin\phi, -\sin\theta\cos\phi, \cos\theta\cos\phi) \quad [42]$$

respectively. The maximum slope angle α of the mirror measured with respect to the substrate surface is calculated by the inner product of $\mathbf{m}''$ and the substrate normal $\mathbf{z}'' = (0, 0, 1)$, and thus,

$$\cos\alpha = \mathbf{m}'' \cdot \mathbf{z}'' = \cos\theta\cos\phi \quad [43]$$

Therefore, angle α is calculated to be

$$\alpha = \cos^{-1}(\cos\theta\cos\phi) \quad [44]$$

We assume an electric field in the shape of arc, as discussed in the previous section for the 1D torsion mirror. As shown in Figure 32, the arc length at point A on the mirror is calculated to be

$$\widehat{A} = \frac{g + A''_z - z}{\sin\alpha}\alpha \quad [45]$$

where g is the initial height of the mirror, z the downward displacement caused by the mean electrostatic force, and A''_z the Z-component of point A on the mirror.

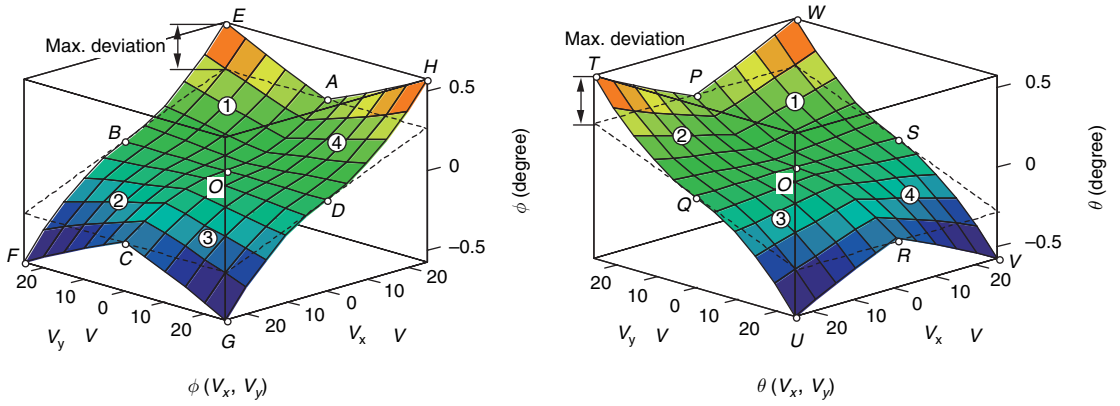


Figure 33 Simulated results of angle–voltage characteristics of the 2D torsion mirror.

Therefore, the electric field is

$$E = \frac{V}{\widehat{A}} = \frac{V}{\dfrac{g + A''_z - z}{\sin\alpha}\cdot\alpha} = \frac{\sin\alpha}{\alpha}\cdot\frac{V}{g - x - x\cos\theta\sin\phi + y\sin\theta} \quad [46]$$

The electrostatic force action on the element is written as

$$dF = \frac{1}{2}\varepsilon_0 E^2\, dS = \frac{1}{2}\varepsilon_0\left(\frac{\sin\alpha}{\alpha}\cdot\frac{V}{g - x - x\cos\theta\sin\phi + y\sin\theta}\right)^2 dS \quad [47]$$

Hence, the electrostatic torques for ϕ- and θ-scan are

$$T_\phi(\phi,\theta,V) = \iint_S \frac{1}{2}\varepsilon_0 x \times\left(\frac{\sin\alpha}{\alpha}\cdot\frac{V}{g - x - x\cos\theta\sin\phi + y\sin\theta}\right)^2 dS \quad [48]$$

and

$$T_\theta(\phi,\theta,V) = \iint_S \frac{1}{2}\varepsilon_0(-y) \times\left(\frac{\sin\alpha}{\alpha}\cdot\frac{V}{g - x - x\cos\theta\sin\phi + y\sin\theta}\right)^2 dS \quad [49]$$

respectively. The negative sign of $(-y)$ in the second equation is due to the definition of angle θ. Contribution from the four electrodes should be summed to calculate the net electrostatic torque. The range of integration for each voltage contribution is taken to be equal to the corresponding electrode area, based on the assumption of the arc-shaped electric field. The electrostatic torques are compared with elastic restoring torque to find equilibrium angles (ϕ, θ) at a given voltage, as shown in **Figure 33**. Toshiyoshi *et al.* (2001) also deal with the vertical motion $-Z$ of the mirror plate. They also demonstrated that the differential voltage scheme worked with the 2D torsion mirror to linearize the angles with respect to the voltage sets.

2.01.6 Vertical Comb-drive Actuators

2.01.6.1 Out-of-plane Motion

Careful measurement of a comb-drive actuator reveals a small out-of-plane (up-and-down) motion along with a large lateral stroke. The parasitic out-of-plane motion is actually due to the presence of the substrate plane that is set at the same potential as the suspended structure, resulting in an asymmetric electric field distribution across the substrate plane. As **Figure 34** illustrates, the bottom side of the suspended structure does not have electric field but there is an electric field on the top surfaces, which causes a small pull-up motion.

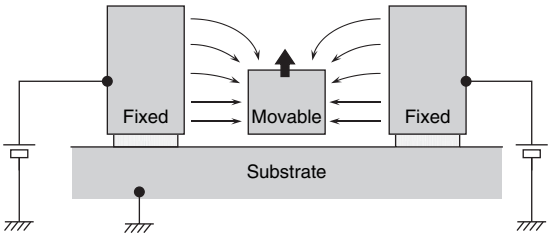


Figure 34 Cross-sectional view of comb-drive mechanism with ground plane.

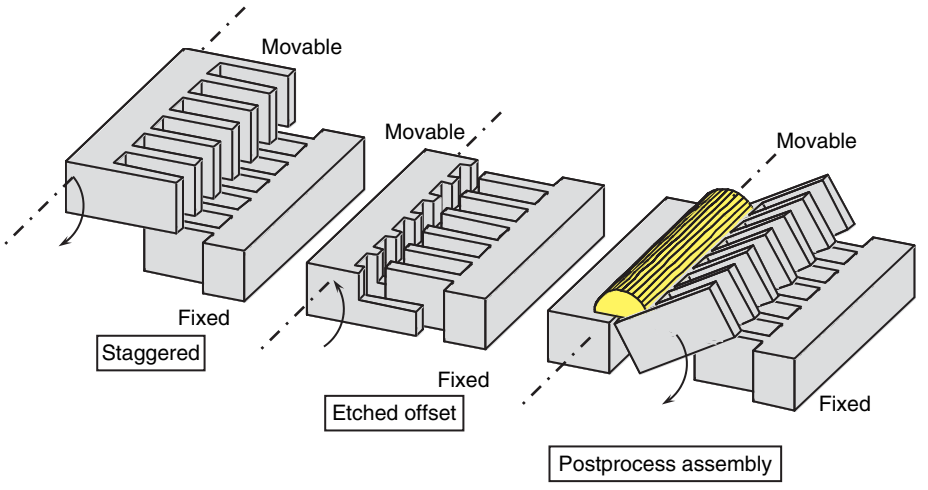


Figure 35 Vertical comb-drive actuation electrodes.

The lateral comb drive was later modified to a vertical comb-drive mechanism that effectively produces large out-of-plane rotational motion with a given device footprint. Several designs of vertical comb-drive actuators were reported, as listed in **Figure 35**. A staggered comb layout is an early version of vertical comb-drive actuators, using the top and the bottom silicon layers of an SOI wafer, for instance, to develop the comb electrodes (Conant 2003). The structure is made by the double-side photolithography and deep reactive ion etching (DRIE) process of silicon, and therefore, the electrostatic operation results are sensitive to the misalignment between the top and the bottom electrode patterns. Optical scanners with staggered combs, as shown in **Figure 36**, were developed by Fujitsu Co. for a free space optical crossconnect application (Yano *et al.* 2005). They thinned down the carrier wafer (the bottom part) of an SOI wafer to 100 μm thickness so that the actuators can be produced by using the silicon DRIE technique. Double-step etching of silicon can also be used to make vertical comb electrodes with height offset. Zicker *et al.* used a combination of the aluminum and photoresist masks to make a height offset (Zicker *et al.* 2002). Patterson *et al.* (2002) and Piyawattanametha *et al.* (2005) reported a postprocess assembly technique for comb-drive offset. They used the surface tension force of thermally reflow photoresist to lift one of the comb pair. Kim *et al.* with the University of California at Berkeley used an accelerated silicon creep at an elevated temperature (900°C) to make a comb offset (Kim *et al.* 2005). Process-stiction of an SOI microstructure onto the substrate can also be used (Isamoto *et al.* 2005), which is easy and reproducible compared with other methods, thanks to the precisely controlled BOX (buried oxide) thickness of a bonded SOI wafer.

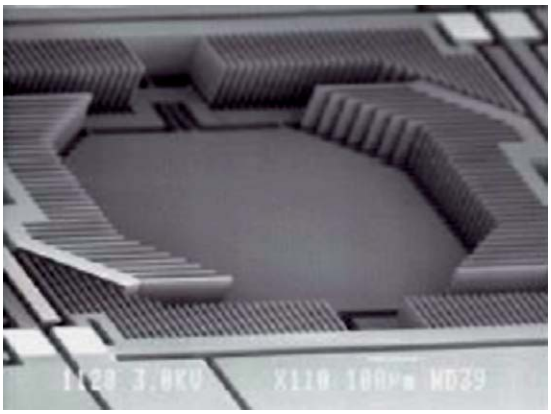


Figure 36 An example of microelectromechanical systems (MEMS) devices equipped with vertical comb-drive mechanism. (Source: Yano M, Yamagishi F, Tsuda T 2005 Optical MEMS for photonic switching – compact, stable optical crossconnect switches for simple, fast, flexible wavelength applications in recent photonic networks. *IEEE J. Sel. Top. Quantum Electron.* **11(2)**, 383–94.)

2.01.6.2 Analytical Model for Vertical-comb Drive

The electromechanical structure of a vertical-comb-driven torsion mirror is complicated compared with that of a simple parallel plate mechanism, and it is suggested that readers use a FEM software that is adapted to the MEMS technology for accuracy. Nevertheless, analytical model presented is useful to obtain a first-order picture of the actuator's

electromechanical behavior. In this section, we study the staggered comb of a uniform comb height by using a simplified model shown in **Figure 37**.

The upper electrodes are fixed, while the lower electrode is suspended to move in the vertical direction Y. The comb displacement y is measured by the position of the movable comb's top surface, taking the origin $y = 0$ aligned at the bottom position of the fixed combs. Assuming that the electrical capacitance C is represented by the overlap area between the electrodes, i.e., by neglecting the field concentration on the comb corners, one would write the capacitance (on both the sides of the comb) as

$$C(y) = \begin{cases} 0 \ (y < 0) \\ 2\varepsilon_0 \dfrac{y}{g} \ (0 < y < h) \\ 2\,\varepsilon_0 \dfrac{2h - y}{g} \ (0 < y < 2h) \\ 0 \ (y < 2h) \end{cases} \qquad [50]$$

where g is the lateral gap between the combs and h the height of the comb finger. Note that we have taken the unit length of the comb (normal to the page surface) to calculate the capacitance. **Figure 37** plots the capacitance of the comb as a function of the movable comb position (solid curve). In the same plot, a numerically calculated capacitance is shown (dashed curve), which has a finite value of capacitance even when the movable comb is not engaged between the fixed ones and has smooth connection in this range. However, the capacitance slope is found to be in parallel with that of the analytical model in the range of $0 < y < 2h$. Remembering that the electrostatic force is proportional to the derivative of capacitance, the analytical model is expected to be in good agreement with the actual device.

The torque contributed by a small fraction of such comb electrodes located at a distance x measured from the rotation axis is represented by

$$\frac{1}{2}\frac{\mathrm{d}C(y)}{\mathrm{d}y} V^2 x \, \mathrm{d}x$$

where $\mathrm{d}x$ is the comb length measured in the direction normal to the page surface. We integrate the contribution along the comb finger and multiply it with the comb number N to obtain the electrostatic torque:

$$T_{\mathrm{elec}}(\theta, V) = N \int_{x_0}^{x_0 + t} \frac{1}{2}\frac{\mathrm{d}C(x\theta)}{\mathrm{d}y} V^2 x \, \mathrm{d}x \qquad [51]$$

where the local comb position y has been replaced with the product of the comb angle θ and the position of the section x, i.e., $y = x\theta$.

The torque is compared with the mechanical restoring torque, eqn [35], to obtain the angle as a function of applied voltage.

Figure 38 shows a typical simulation result of a torsion mirror with vertical comb-drive actuators. The angle is saturated at high voltage, because the mean electrostatic capacitance takes its maximum value when the movable comb electrode is fully engaged in the fixed electrodes, and therefore, a thicker layer should be used to make taller electrodes to tilt the comb electrodes more. For achieving a large scan angle at resonance, Yoda *et al.* (2005) reported a two-mass-two-spring system to magnify the small amplitude a vertical-comb plate to make a large oscillation of the scanning mirror.

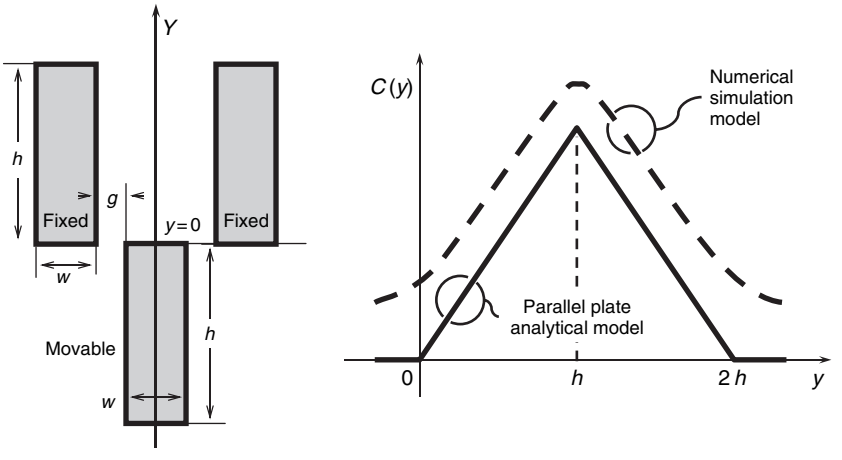


Figure 37 Simplified analytical model for staggered comb drive.

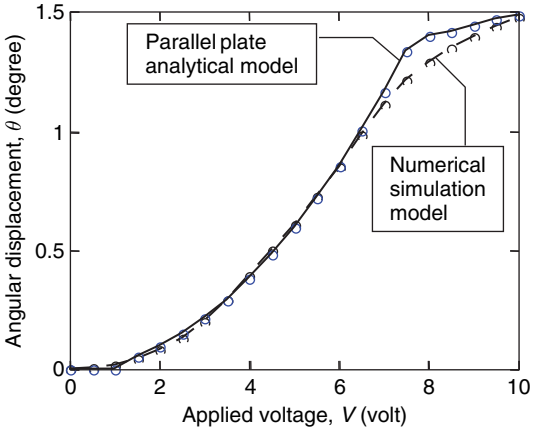


Figure 38 Simulation result of vertical-comb mirror angle as a function of drive voltage.

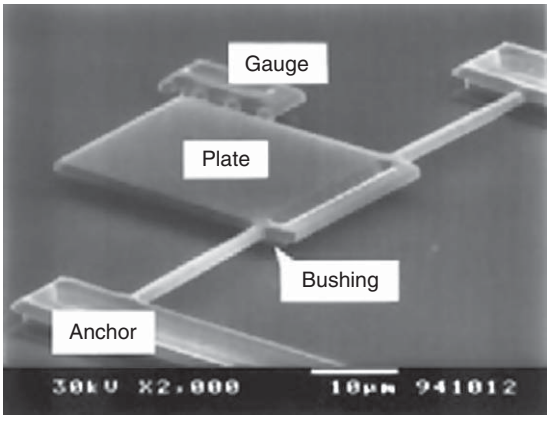


Figure 39 Scratch-drive actuator (SDA). (Source: Akiyama T, Shono K 1993 Controlled stepwise motion in polysilicon microstructures. *J. Microelectromech. Syst.* **2(3)**, 106–10.)

Eqn [51] suggests that the electrostatic torque can be made large by using many comb fingers in a given area, i.e., by increasing the comb finger density. Designing a thin comb finger and a small air gap will increase the density. Small air gaps also help increase the electrostatic force. However, one should be warned that the small-gap comb electrodes are likely to fall into the lateral pull-in instability. A suspended structure experiences an in-plane electrostatic torque rather than an out-of-plane torque and makes the comb fingers trip to the other, resulting in mechanical stop and electrical short circuit. To overcome this problem, Fujitsu Co. used V-shaped suspensions by combining two straight beams with a relatively large opening on the mirror side and a small pointing root on the fixed anchor (Yano *et al.* 2005). The V-shaped suspensions were found to have large lateral rigidity and small rotational rigidity. They used such optical scanner array to manufacture a fiber-optic crossconnect of 128 input ports and 128 output ports(Yano *et al.* 2005).

2.01.7 Scratch-drive Actuators

2.01.7.1 Tethered Linear Actuation

Most MEMS actuators are made to be tethered to the substrate via elastic suspensions, which is useful to minimize the friction force associated with the mechanical motion. However, friction was also found to be a useful tool to make a powerful microactuation mechanism. Akiyama and Shono (1993) invented a novel electrostatic linear actuator called a scratch-drive actuator (SDA) that stepped forward with a pulse voltage like an inchworm.

The actuator is made with a polysilicon plate of typical 100 μm × 100 μm in area and 1–2 μm in thickness, as shown in **Figure 39**. The leading edge of the plate has a bushing on the bottom, placed on a silicon substrate with a thin layer of silicon nitride (insulator) in between. Drive principle is a repetition of the following steps: (1) When a voltage (typical 20–100 V) was applied to the plate through the polysilicon tether, the actuator plate deflected downward to the substrate by the electrostatic attractive force. Due to the strong friction between the actuator plate and the substrate, the tail of the plate was pinned down to the substrate, while the point-contact bushing on the leading edge was forced to slide forward with a small increment in a 10- to 30-nm range. (2) When the voltage dropped to zero, the actuator plate was released from the substrate and it bounced upward. At this moment, the plate was dragged forward to the bushing because the rotation pivot was at the bushing contact. (3) The actuator plate was immediately pulled down again by the next voltage peak to keep the position. The repetition of Steps (1) to (3) enables the actuator plate to move forward. The SDA in **Figure 39** showed a travel distance of 50 μm. Traveling speed was controlled by tuning the frequency of the pulse voltage in the range of 1 Hz to 100 kHz.

Higher output force is possible by making an array of SDAs in one sheet. Minotti *et al.* integrated more than 1000 pieces of SDA plates in a polysilicon sheet of 2 mm × 3 mm, as shown in **Figure 40**. Assuming a 100-μN output from a single SDA plate, the total

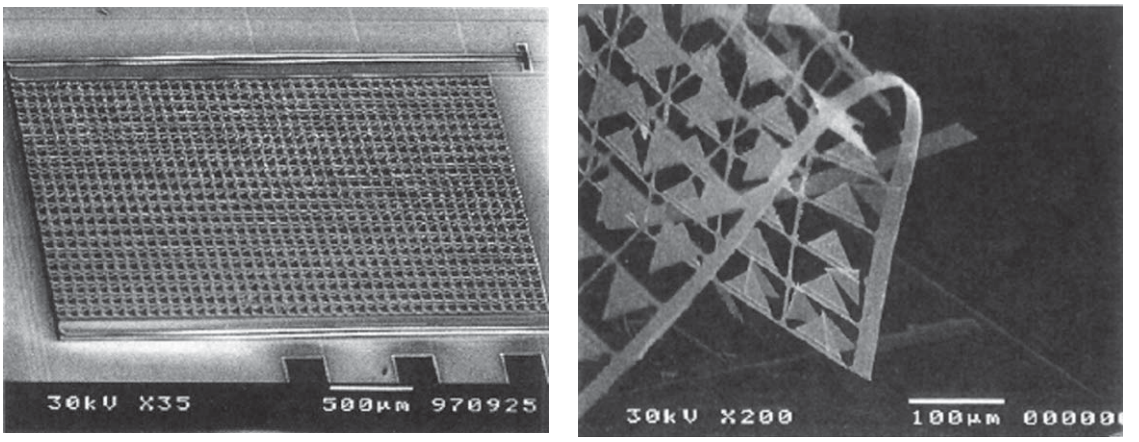


Figure 40 Sheet of arrayed scratch-drive actuators (SDAs). (Source: Minotti P, Langlet P, Bourbon G, and Masuzawa T 1998 Towards new cylindrical electrostatic micromotors using tubular combinations of arrayed direct drive actuators. *Jpn. J. Appl. Phys.* 37 part 2: L622–5.)

output force was estimated to be 0.1 N or 10 g; this implies that the SDA can hold a load mass of 10^6 times its own weight.

2.01.7.2 Self-assembly Using SDA

The output force of the SDA was found to be stronger than other electrostatic microactuators. Akiyama also demonstrated a lift-up motion of a micropiece of quartz (~1 mm) by using a single SDA of 60 µm × 75 µm operated at 110 V (Akiyama and Shono 1993). From this experiment, he estimated the actuator force output to be more than 50 µN. SDA was found to be a useful built-in actuator to self-assemble 3D microstructures after sacrificial release. **Figure 41** shows an example of such 3D microstructure of polysilicon that was elastically deformed by the SDA array (Akiyama *et al.* 1997).

In Akiyama's first demonstration, the SDA plate was elastically tethered by the polysilicon hinges, which were also used as an electrical interconnection to the SDA plate. However, the travel distance of the SDA plate was limited by the elastic restoring force of the springs. A nontethered version of SDAs was reported by Li *et al.* with the University of California at Los Angeles (UCLA) in 1997 (Fan *et al.* 1997). They used the polysilicon surface micromachining of multiuser MEMS processes (MUMPs) to develop the chain of SDA plates that were mechanically guided by the stapled rails (Fan *et al.* 1997), also integrated with a 3D hinged structure (Pister *et al.* 1992). They put a microball lens of 300 µm in diameter in the socket plate and jacked it upward by more than 100 µm using the force of SDA modules, as shown in **Figure 42**. Both vertical *Z*-direction and lateral *X*- and *Y*-directions were controlled by the SDA arrays. Thanks to the multilayered structure of the MUMPs, the drive voltage was applied through

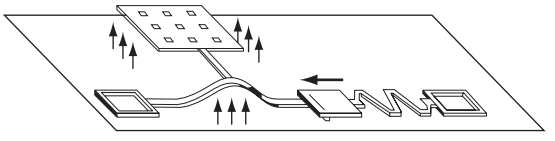


Figure 41 Self-assembled 3D microstructure using scratch-drive actuator (SDA). (Source: Akiyama T, Collard D, Fujita H 1997 Scratch drive actuator with mechanical links for self-assembly of three-dimensional MEMS. *J. Microelectromech. Syst.* **6**, 10–17.)

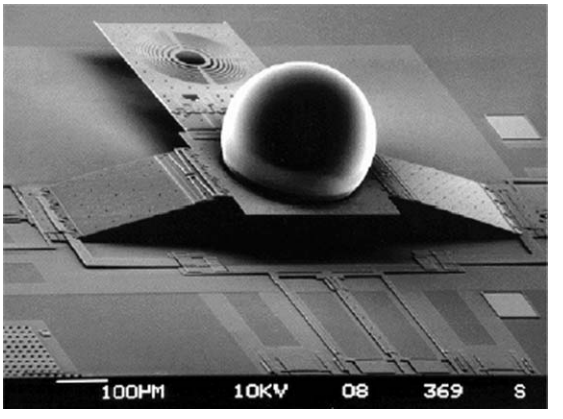


Figure 42 Self-assembled polysilicon hinged structure using scratch-drive actuator (SDA) arrays. (Source: Fan L, Wu M C, Choquette K, Crawford M 1997 Self-assembled microfabricated XYZ stages for optical scanning, alignment. *Proc. Int. Conf. Solid-State Sensors and Actuators (Transducers 97)*, Chicago, IL, USA, June 16–19, 1997, pp. 166–22.)

the underlying thin polysilicon layer (called poly-0), keeping the SDA plate and the rest of the structure at the electrical ground. Li *et al.* also developed a bidirectional SDA module by connecting a pair of oppositely directed SDAs; micropatterns of photoresist were used to mechanically bind the SDAs and also to electrically decouple them.

2.01.8 Electrostatic Impact-drive Actuators

2.01.8.1 Inchworm Actuation

The prototype model of micromechanical impact-drive actuator, presented by Higuchi *et al.* with the University of Tokyo in 1987, also used the friction to keep the position and to generate thrust force (Higuchi *et al.* 1988). They used a pair of metal blocks (several centimeters large) combined with a piezoelectric PbZrTi (lead zirconium titanate, PZT) actuator and demonstrated step motion by using pulse and ramp voltages. The step motion was found to be in nanometric range, and thus, they could use the mechanism to position the probe tip of, for instance, atomic force or tunneling microscope.

2.01.8.2 MEMS Inchworm

A MEMS version of impact-drive actuator, an electrostatic impact-drive mechanism (EIM), was reported by Mita *et al.* (2003) with the University of Tokyo in 2003. **Figure 43** illustrates the actuation principle of the EIM, where a silicon hammer is suspended with an elastic spring in a square shell made by the silicon micromachining. The mass was designed to be a movable electrode in an electrostatic actuator, and the wall on the opposing side was used as a counterelectrode. The pulse voltage applied to the counterelectrode attracted the mass by the electrostatic pull-in mechanism. The kinetic energy gained by the electrostatically accelerated mass was transferred to the shell by collision. When the impact force exceeded the friction force, the entire shell box moved forward until all the kinetic energy was consumed. The cycle was repeated by modestly retracting the mass by releasing the voltage.

Figure 44 shows a SEM image of the microfabricated EIM actuator. The silicon hammer in the middle had two masses on both sides, and the elastic suspensions were accommodated in the notched part in the middle. The drive electrodes were located on

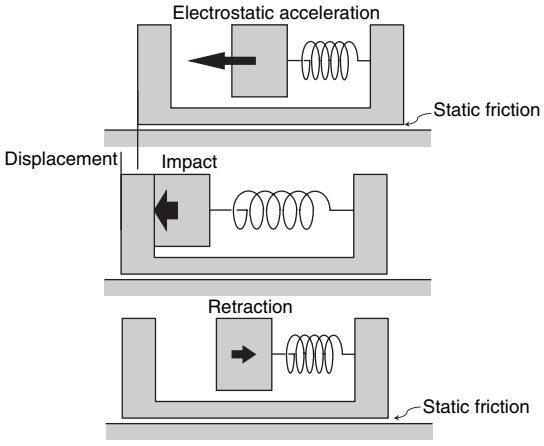


Figure 43 Operation steps of an electrostatic impact-drive mechanism (EIM). (Source: Mita M, Arai M, Tensaka S, Kobayashi D, Fujita H 2003 A micromachined impact microactuator driven by electrostatic force. *J. Microelectromech. Syst.* **12**, 37–41.)

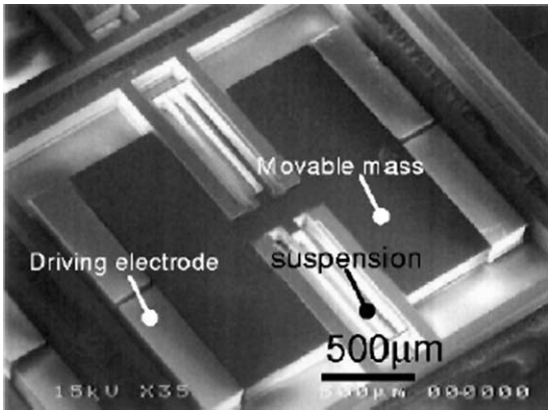


Figure 44 Scanning electron microscopic (SEM) image of microelectromechanical systems (MEMS) impact-drive actuator. (Source: Mita M, Arai M, Tensaka S, Kobayashi D, Fujita H 2003 A micromachined impact microactuator driven by electrostatic force. *J. Microelectromech. Syst.* **12**, 37–41.)

both sides of the hammer, by which bidirectional motion could be achieved. The stopper for the hammer was located in the middle part, rather than the hammer edge, in order to have a large electrode area for higher driving force. The device was made by bulk micromachining a silicon wafer with DRIE. The entire mechanism was hermetically sealed by the wafer bonding technique, and the electrical interconnection was made by using thin gold thread of wire bonding. The impact actuation was found to be powerful that gives unlimited travel distance regardless of the tethering wires. With a sinusoidal voltage of 100 V_p at 200 Hz, the module traveled at 2.7 $\mu m\, s^{-1}$,

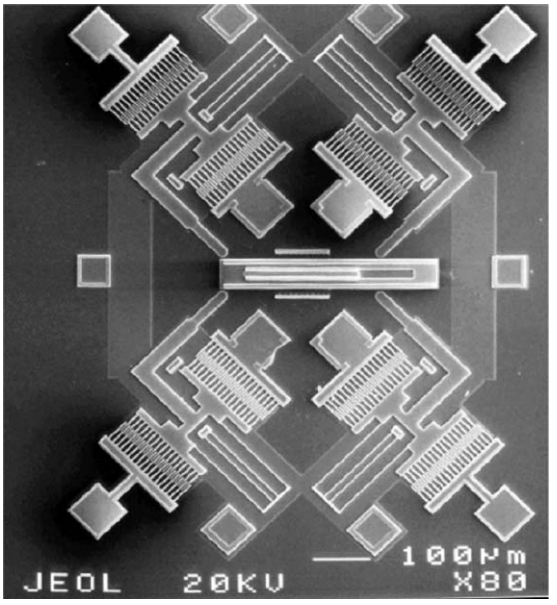


Figure 45 Micromechanical sliders using impact drivers. (Source: Daneman M J, Tien N C, Solgaard O, Pisano A P, Lau K Y, Muller R S 1996 Linear microvibromotor for positioning optical components. *J. Microelectromech. Syst.* **5(3)**, 159–65.)

with minimum step of 13 nm. The authors expected wireless control of the EIM by integrating a power transmitting module in the package.

The impact-drive mechanism was also used to develop a micromechanism of a slider. Pisano *et al.* with the University of California at Berkeley developed a bidirectional linear slider by using four sets of comb-drive impact actuators with a hammer head, as shown in **Figure 45** (Daneman *et al.* 1996). They demonstrated a slider speed of $1\,\mathrm{mm\,s^{-1}}$ with a 0.27-μm step by applying pulse voltages of 12.5 V at 7 kHz. Pai *et al.* used a similar mechanism but with electrothermal actuators to compose an optical fiber switch (Pai and Tien 1997). They used the microhinge technology to integrate a vertical reflector sitting on the slider. The mirror was driven in and out to change the optical path in a small free space between four optical fiber facets.

2.01.9 Electret Actuators

2.01.9.1 Use of Repulsive Coulomb Force

All the electric actuators discussed in the previous sections utilize the attractive Coulomb force between oppositely charged bodies. Electrostatic repulsive force can also be seen in, for instance, an electroscope, which is a simple device consisting of a metal piece sticking out of a glass jar with thin leaves of metal hanging inside. When the external metal piece is touched by a charged body, the leaves spread apart by the Coulomb force. However, the charge on the metal body cannot be fixed permanently, and a relatively large voltage is needed for charging.

2.01.9.2 Electret by Corona Discharge

True charge can be permanently trapped in dielectric material by the corona discharge or by the poling process at temperature higher than glass transition point. The first electret motor was developed in a macroscale by Gubkin in 1961 (Jefimenko 1973). The rotor consisted of a pair of electrets that had been polarized in the opposite directions, and the rotating position was electrostatically controlled by using the metal stator electrodes. Later, Jefimenko *et al.* reported another type of electret motor by using the electret material on the stators and by biasing the rotor (Jefimenko 1973). In the micromechanical scale, Wood *et al.* (1987) used a negatively charged plastic film on an optical fiber to control the position by the electrostatic force.

2.01.10 Film Actuators

2.01.10.1 Electrostatic Induction

Electrostatic induction has been known for a long time. Riccardo Arno invented a macroscale induction motor in 1892, by using the polarization in a dielectric material. A dielectric rotor (an ebonite rod of 8 cm in diameter and 18 cm in length) placed in a rotating electric field built up the polarization by electrostatic induction. Due to the rotation of the electric field and the dielectric response delay of the material, the polarization direction had a small angular offset with respect to the turning electric field, causing a torque on the rotor. Arno used a high voltage of 3.8 kV to demonstrate a rotation speed of 250 rpm and output torque on the order of 10^{-4} Nm.

2.01.10.2 Flexible Film Actuators

The charge trapped in a dielectric material disappears by discharge, and hence, the electromechanical performance of an electret actuator may drift with aging. Egawa *et al.* reported a highly reproducible electrostatic repulsive force by using an instantaneously induced charge on a highly resistive film (Egawa and Higuchi 1990).

Figure 46 illustrates the side view and the step motion of the film actuator. The stator was made of a flexible printed circuit board (PCB) with triphase

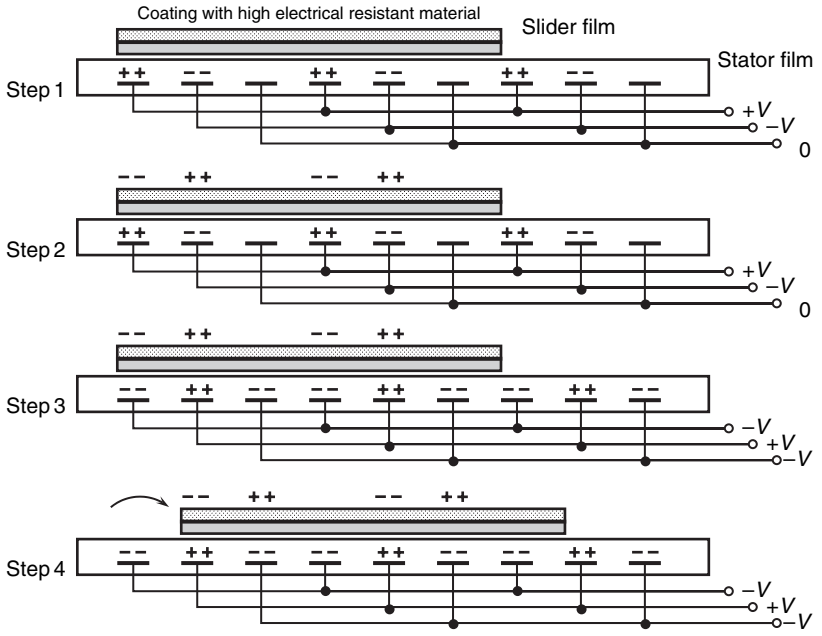


Figure 46 Operation steps of film actuator using electrostatically induction. (Source: Egawa S, Higuchi T 1990 Multi-layered electrostatic film actuator. *Proc. IEEE Micro Electro Mechanical Systems Workshop*, Napa Valley, CA, USA, February 11–14, 1990, pp. 166–71.)

electrodes embedded inside. The slider is a piece of plastic film coated with a material of high electric resistance. In the operation Step 1, the electrodes in the stator were biased with voltage. The distributed voltages induced charge on the slider film in Step 2, causing electrostatic attractive force. Due to the high resistance in the coated film, the charge was immobilized for a short time even when the driving voltages were partially flipped as shown in Step 3. At this step, the slider film experienced repulsive force in the vertical direction. At the same time, the film was pulled to the right by the electrostatic attractive force of the distributed charge. As a result, the slider film levitated for a short time to make a gliding inchworm step in the lateral direction. No accurate alignment or positioning was needed to set the slider film, because all the electrode patterns were made in the stator film.

The same authors also reported another type of electrostatic film actuator; different from the previous one, both the films (stator and slider) were made of patterned flexible PCB. Similar sliding motion was observed by using a set of sinusoidal voltage waves of different phases. The advantage of the second film actuator was that it could be made large in area to increase the output force. Furthermore, the films were stacked up to multiply the force in a given area. Egawa developed a stacked film actuator of 50 pairs of sheets and demonstrated 300 N output with a drive voltage of 2 kV. Thanks to the mechanical flexibility of the film, the actuator could be used on a curved surface such as a pillar. They also demonstrated rotational types of electrostatic film actuator.

2.01.11 Pitfalls in Electrostatic Actuation and Hints

2.01.11.1 Displacement–Response Trade-off

Displacement is not a sole parameter to describe the performance of an electrostatic actuator, but the fundamental resonant frequency plays an important role when the actuator is used for high-speed applications such as optical switches and radio frequency waveguide switches. Representing an electromechanical actuator by using a quadratic oscillating system with a mass m and a spring constant k, one would write the fundamental resonant frequency f_0 as

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \qquad [52]$$

Substituting another form of the equation, $k = 4\pi^2 m f_0^2$, into eqn [4], which gives the

equilibrium point of electrostatic force and elastic restoring force, one will obtain a governing equation of a parallel plate actuator:

$$x(g-x)^2 f_0^2 = \frac{\varepsilon_0}{8\pi^2}\frac{S}{m}V^2 \qquad [53]$$

The equation reveals that the electrostatic parallel plate has a built-in trade-off relation between the resonant frequency (speed) and the displacement (Toshiyoshi *et al.* 2002). For a larger displacement, one may need to use a soft spring, but the resonant frequency is lowered. For a fast response of actuator, on the contrary, one would use a hard spring, sacrificing the displacement. We sometimes encounter a difficulty in designing actuators when both speed and displacement are sought for.

MEMS designers are usually given a limited device footprint, material combination, and maximum allowable voltage for operation when they design actuators to meet the device requirements. Prototype devices would be developed in a trial-and-error manner by using a couple of different dimensions for, for instance, the suspension width on a photomask, because the width is a parameter that could easily be altered without causing significant layout change in the device master plan. But changing the suspension width alone may not be enough to improve the performance. **Figure 47** plots a displacement–frequency curve calculated by using the governing eqn [53], taking the suspension width as a plotting parameter. Given a voltage and a gap, one may not be able to escape from the governing curve or to approach the target performance region.

Eqn [53] suggests three parameters to improve the performance, namely, m, g, and V (Toshiyoshi *et al.* 2002). The mass m of the movable part should be designed to be light to improve the response. From a density point of view, silicon is a light material ($\rho = 2.33 \times 10^3\,\mathrm{kg\,m^{-3}}$), compared with metals such as plated nickel ($\rho = 7.2 \times 10^3\,\mathrm{kg\,m^{-3}}$) or even aluminum ($\rho = 2.7 \times 10^3\,\mathrm{kg\,m^{-3}}$). Weight saving by using a honeycomb structure would also be effective (Patterson *et al.* 2000). Using a small electrostatic gap g is helpful to increase the output force of the actuator. For this reason, narrow trench etching by silicon DRIE is highly appreciated for comb-drive mechanisms. High voltage should also be permitted to drive the actuator, because the electrostatic force is proportional to the square of voltage. As an advanced MEMS approach, several works on high-voltage driver circuits integrated with microactuators have been reported (Brosnihan *et al.* 2003, Takahashi *et al.* 2005).

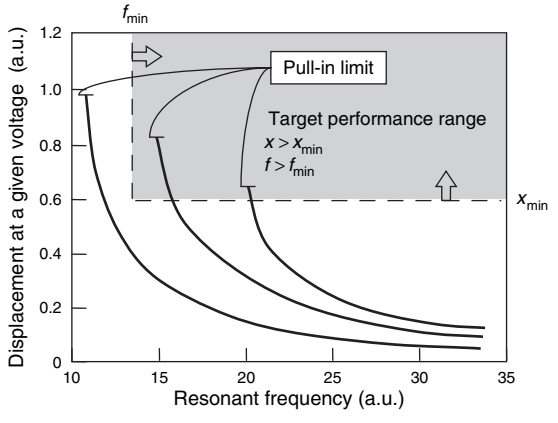


Figure 47 Trade-off relation between resonant frequency and displacement at a given voltage.

2.01.11.2 Doubled-frequency Response

Due to the nature of attractive Coulomb force, electrostatic actuators exhibit static displacement in the gap-closing direction regardless of the polarity of applied voltage, which is usually seen in the frequency range fairly lower than the actuator's resonant frequency, as shown in **Figure 48(a)**. When the input voltage does not have a DC component, the actuator responds to both positive and negative peak voltages, and it seems to have doubled the frequency with respect to that of the input voltage. One should be aware of this nonlinearity when measuring the frequency response of electrostatic actuators. For a better understanding of the actuator's response by suppressing the higher modes, the drive voltage should be accompanied with a DC offset, as shown in **Figure 48(b)**. Note that the AC component of the drive voltage should be small compared with the DC component.

Figure 49 shows the results of numerical simulation of a quadratic oscillation system with an electrostatic force by using the Runge–Kutta method. At a frequency lower than the resonant frequency, the actuator's displacement was limited in the $+X$ range. When the frequency was set to the resonance, the displacement oscillated in both $+X$ and $-X$ directions at the input frequency; also a small offset toward $+X$ was observed in the response. At a frequency higher than the resonance, the amplitude becomes small but a small offset displacement remains in the level.

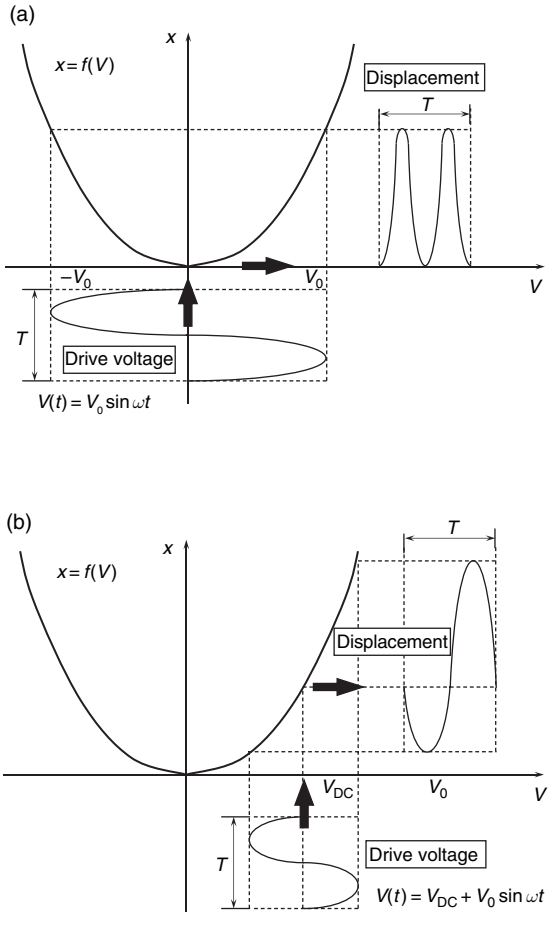


Figure 48 Plot of the transfer function of electrostatic actuator.

2.01.11.3 Pseudostatic Operation at Frequency Higher than Resonance

Thanks to the offset displacement seen in the high-frequency range, electrostatic actuators can be operated by the amplitude-modulated (AM) voltage, as if it were driven by a DC voltage of the same peak value. **Figure 50(a)** and **50(b)** compares the numerically calculated responses to a step-function voltage and an AM voltage, respectively. The AM voltage was designed to have an envelope curve identical to the step function, and its carrier frequency was set to a frequency 100 times higher than the actuator's resonance. The simulation results clearly show that the responses were almost equal to each other, reflecting the envelope shape. The AM voltage drive requires a complicated driver circuit and more power to transmit. Nevertheless, it promises a benefit that no electric charge builds up on the electrodes, and therefore, the actuator's displacement is less likely to drift with time.

2.01.11.4 Negative Spring Constant Effect

Another characteristic of electrostatic actuator is the bias-voltage dependence of its resonant frequency, which tends to shift toward lower frequencies with increase of the bias voltage. **Figure 51** is the experimental result of frequency response of an electrostatic torsion mirror that had a gap-closing electrostatic mechanism between the SOI microstructure and the silicon substrate. The mirror was originally developed for a variable optical attenuator (VOA) of relatively low drive voltages (~5 V) (Isamoto *et al.* 2004). When the mirror was operated at a 1 V AC with increasing bias voltages from 2 to 5 V, the resonant peak was found to shift toward the lower frequency from 750 to 650 Hz (13% down). In a practical scene of the optical fiber system, the lower shift of the resonance results in degraded temporal response in the switching operation of the VOA, in particular at the rising edge of drive voltage.

The phenomenon is known to be a negative spring constant effect or electrostatic soft spring effect (described by the Duffing's equation), which associates with the nonlinear characteristic of the electrostatic actuation. In addition to the elastic restoring force, $k\,x$, the actuator plate is pulled in the opposite direction by the electrostatic attractive force

$$\frac{1}{2}\varepsilon_0 \frac{S}{(g-x)^2} V^2$$

Even though the bias voltage is fixed at a constant value, the static force increases its strength as the displacement x approaches the pull-in displacement. Therefore, the actuator plate at a biased position looks as if it is suspended with a spring of reduced rigidity, resulting in the shift of resonance toward the lower frequency.

The negative spring constant effect becomes a problem when a fast response of the actuator is sought for at relatively large displacement, or when one measures the resonant frequency in a physical sensor. The benefit of the negative spring constant effect is that it could be used to electrically tune the mechanical resonance of a MEMS oscillator.

Hysteresis is also observed in the frequency response of electrostatic actuator, when it oscillates at a large amplitude. **Figure 52** shows the frequency response curve. Due to the negative spring constant

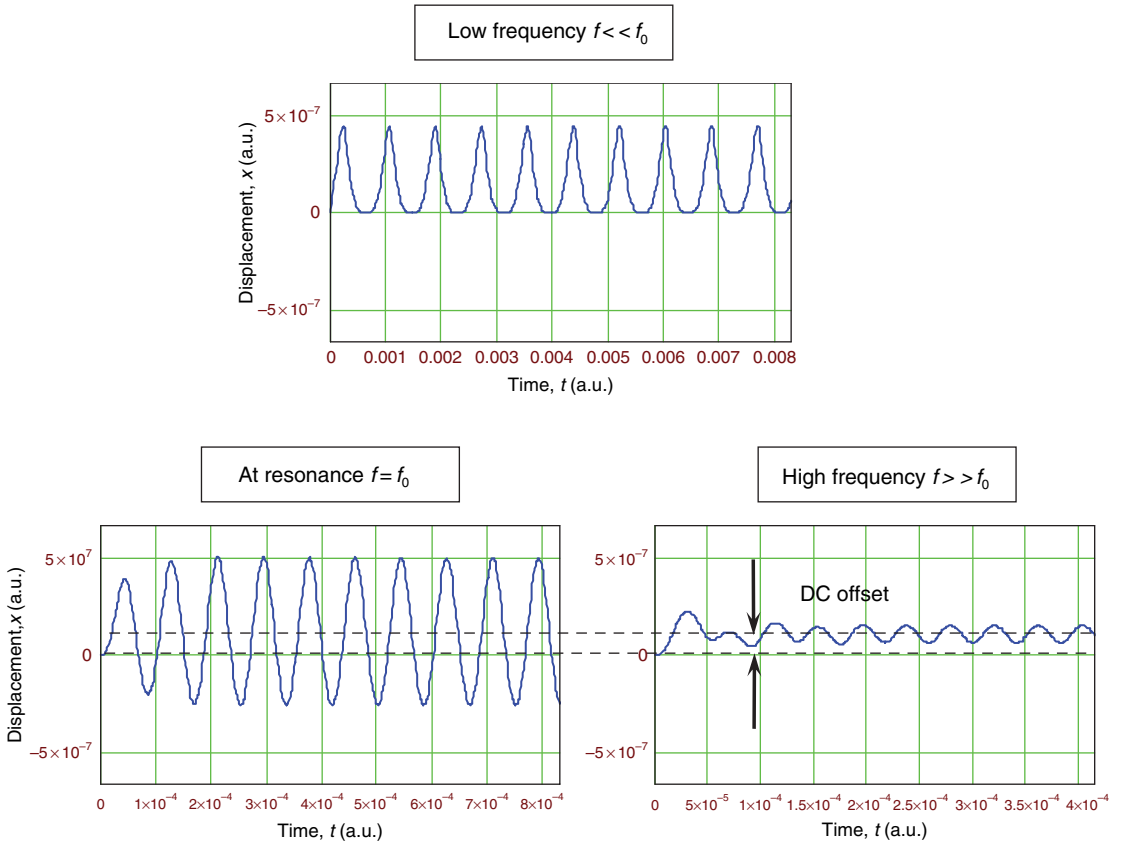


Figure 49 Response of electrostatic actuators in different frequency ranges.

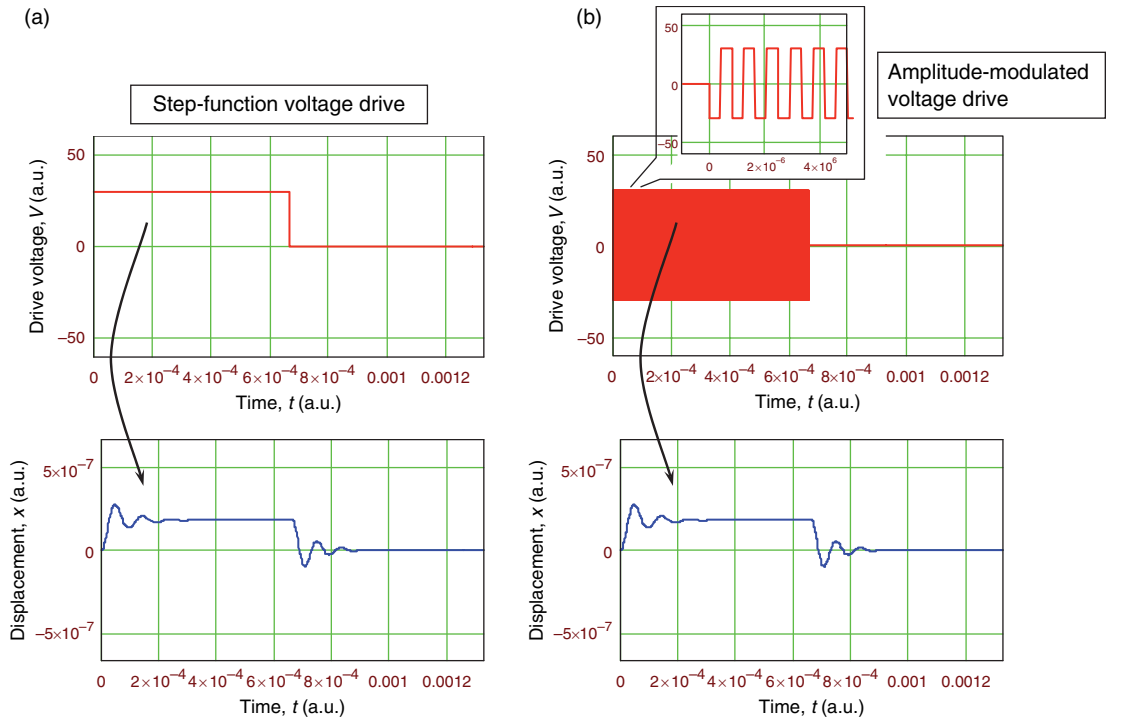


Figure 50 DC voltage operation and amplitude-modulated voltage operation.

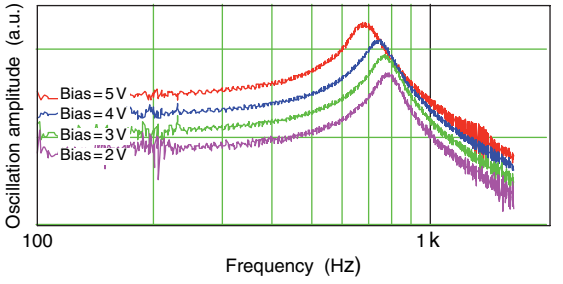


Figure 51 Shift of resonant frequency due to the negative spring constant effect of electrostatic actuation.

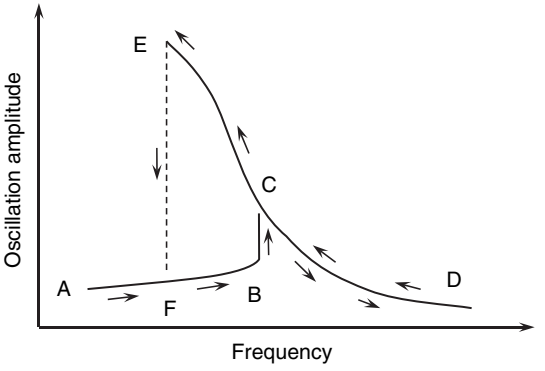


Figure 52 Hysteresis in frequency response of electrostatic actuator.

effect, the response peak inclines toward the lower frequency side. Oscillation amplitude becomes large when the excitation frequency is changed from point A to point B, and the amplitude is observed to jump to point C, after which it becomes small again, with the increase in the frequency to point D. When the frequency is decreased from point D, on the contrary, the amplitude keeps growing large until it collapses at point E. Therefore, when measuring the frequency response by manually tuning the frequency of a function generator, readers are advised to sweep up and down the frequency to see whether the actuator has a large difference in the amplitude.

2.01.11.5 Shortening Settling Time of Step Response

Damped oscillation as shown in Figure 53 is usually observed in an electrostatic actuator when it is driven with a step-function voltage. The duration of oscillation is long when the actuator has high quality factor (high Q), i.e., low loss. High-Q oscillation system is suitable for, for instance, a galvano scanner that spatially modulates a light beam with a large angle and

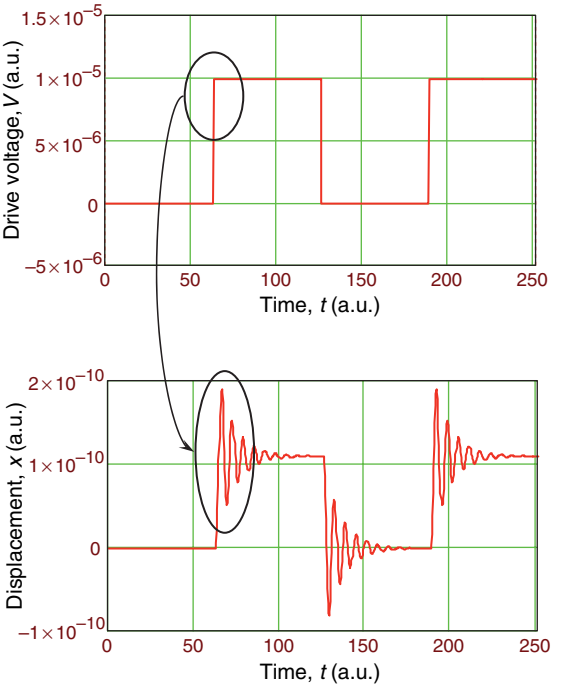


Figure 53 Damped oscillation observed in step response.

with a low voltage. However, relatively long settling time of a high-Q system becomes a problem when the actuator is used for optical switches, for instance. Long-term drift of actuator displacement can be controlled by using a feedback system that monitors the displacement, but the feedback system is not always fast enough to give correction to the transient response in time.

If the resonant frequency of the system (ω_0) is known, a simple feed-forward control can be used to suppress the damped oscillation. Figure 54 is a close-up view of the leading edge of the step response in Figure 53. Thanks to the fact that the damped oscillation is well described by $A + B\exp(-t/\tau)\sin\omega_0 t$, the oscillation would nearly diminish when it is superposed with another step response with a phase delay of π/ω_0, as shown by the dashed curve. This gives an insight that the damped oscillation can be eliminated by driving the actuator with superposed step-function voltage with a phase delay. Figure 54 is an example of feed-forward voltage for minimum settling time that was found in a trial-and-error manner by using a numerical simulation. Thanks to the superposed step voltages, the leading edge became almost free from damped oscillation (the falling edge required another preshaped voltage, which is not shown here). Note that the oscillation frequency depends on the electric

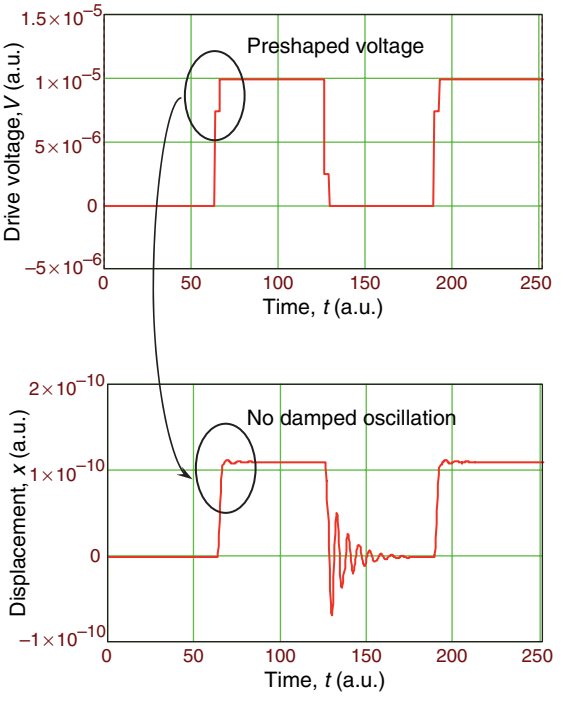


Figure 54 Response to a preshaped step-function voltage.

bias voltage due to the negative spring constant effect, and therefore, the feed-forward control system may require a look-up table of parameters.

2.01.11.6 Polarity-dependent Displacement

Electrostatic actuators made of semiconductor silicon sometimes exhibit peculiar behavior such as asymmetric displacement even when they are driven with an AC voltage without any offset. Figure 55 is an

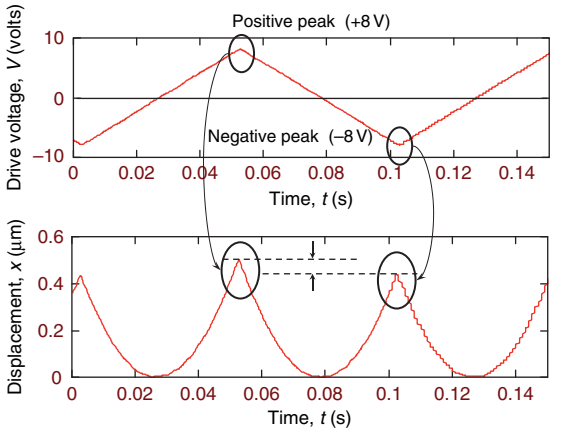


Figure 55 Asymmetric electrostatic displacement.

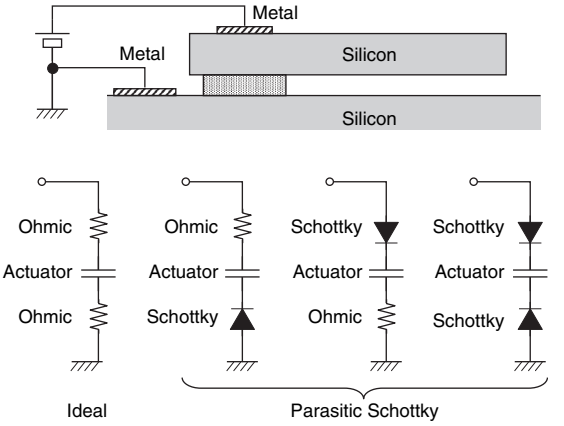


Figure 56 Schottky contact to cause asymmetry in electrostatic displacement.

example of such asymmetric displacement, which becomes notable when the drive voltage is as small as a few volts. This phenomenon is observed randomly on chips sampled out from a large number of devices made on a single piece of wafer at a time. In this figure, the peak at the positive input voltage is shown larger, but the polarity may switch on a different chip.

The asymmetry is most plausibly explained by using the Schottky diode made at the interface of the silicon and the metal, as illustrated in Figure 56. Actuators made of silicon usually require metallization for better electrical conductivity, and aluminum or a combination of chromium and gold is usually used for good adhesion to silicon and gold bonding wires. However, it is sometimes observed that the metal is not well processed to make an ohmic contact with silicon but makes an electric potential barrier. In such a case, the actual voltage drop across the actuator electrodes becomes smaller than the applied voltage by the potential barrier, and the electrostatic displacement is produced to be small. The randomness in the polarity is explained by the chance that one of the electrodes (the SOI structure or the substrate, for instance) is not made in ohmic but in Schottky.

2.01.11.7 Electrostatic Drift

Electrostatic actuators made by the surface micromachining may be found to have a drift in their displacement, as shown in Figure 57, even though they are operated at a constant drive voltage from a controlled voltage source. The drift is a serious problem for an application such as analog

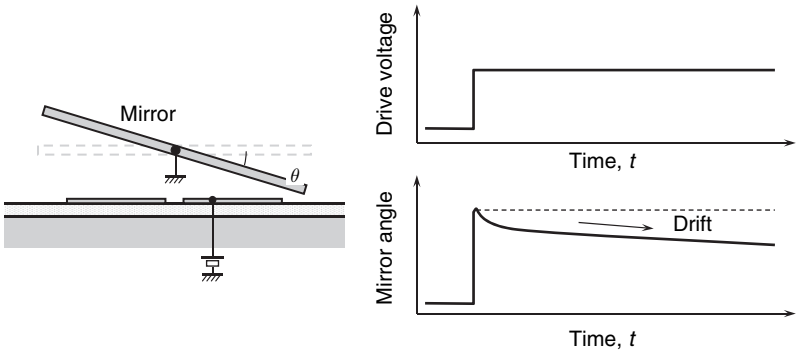


Figure 57 Example of electrostatic drift in displacement.

controlled optical switch, where the mirror position should be maintained at a fixed angle for a long time with an analog controlled voltage. The drift is usually observed in the tendency that the displacement (mirror angle) becomes smaller with aging on the order of minutes to days.

Most electrostatic drift can be suppressed by reducing the risk of charge buildup around the movable structure by, for instance, removing all the electrically floating patterns. It is also helpful to minimize the area of electrical insulator that is exposed to the movable element, as illustrated in **Figure 58**. In the silicon surface micromachining technology, thin-film silicon nitride is used as an insulator on the substrate, thanks to its chemical tolerance to hydrofluoric acid. However, silicon nitride is also known to be a material that easily captures charge (Wibbeler *et al.* 1998), possibly resulting in the drift of electrostatic performance. For a torsion mirror application, for instance, the gap between the electrodes sitting on the nitride layer should be made as small as possible.

Electrostatic drift is also seen in a device that is free from a nitride insulator. One of the curves shown in **Figure 59** is a drift observed in a silicon bulk micromachined actuator; it was originally developed as an electrostatic torsion mirror for a fiber optic attenuator (Isamoto *et al.* 2004). The electrostatic mirror was made of the active layer of an SOI wafer, and it was suspended on the substrate with an air gap in between. Operating voltage was applied to the suspended mirror with respect to the substrate at the ground level. The aforementioned plot was observed in the mirror chip, whose substrate ground was taken by the gold wire bonded on the top surface, leaving the bottom surface at an electrical floating level. **Figure 59** also shows another curve of less fluctuation; it was observed in the mirror chip of the identical design but with an evaporated metal layer on the backside of the chip. Thanks to the conductivity of the deposited metal, the electrical potential in the silicon substrate was flattened, and the voltage across the air gap was explained to have been stabilized.

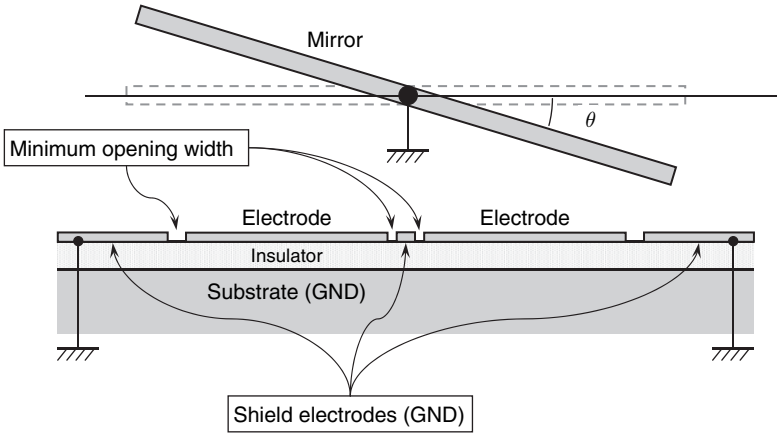


Figure 58 Minimum area of insulator for less risk of electrostatic drift.

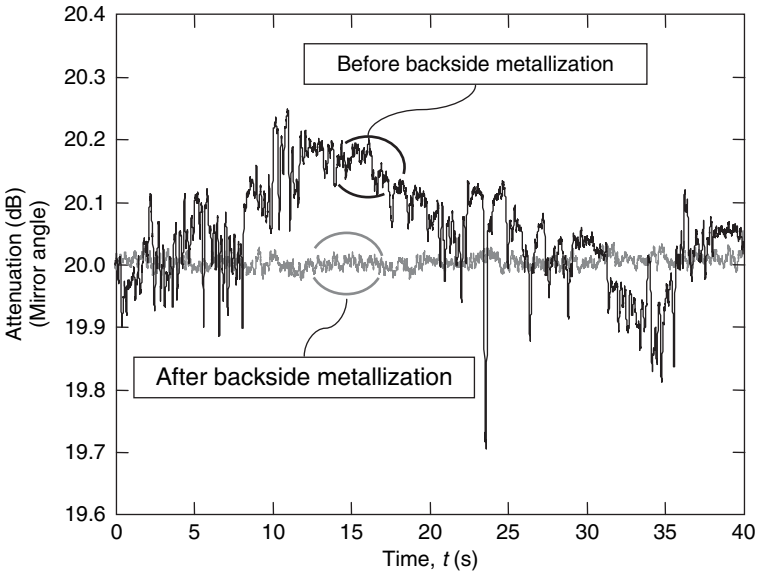


Figure 59 Short-term electrostatic drift.

2.01.11.8 MEMS Digital-to-analog Conversion Mechanism of Displacement

Toshiyoshi *et al.* (2000) reported a micromechanism to produce a displacement that was free from electrostatic fluctuation or drift. The mechanism is an *N*-stage network of ladder-connected suspensions, where an array of electrostatic microactuators was attached, as shown in **Figure 60**. The relative rigidities of the suspensions were designed to be either C or 2C such that it copied the function of digital-to-analog

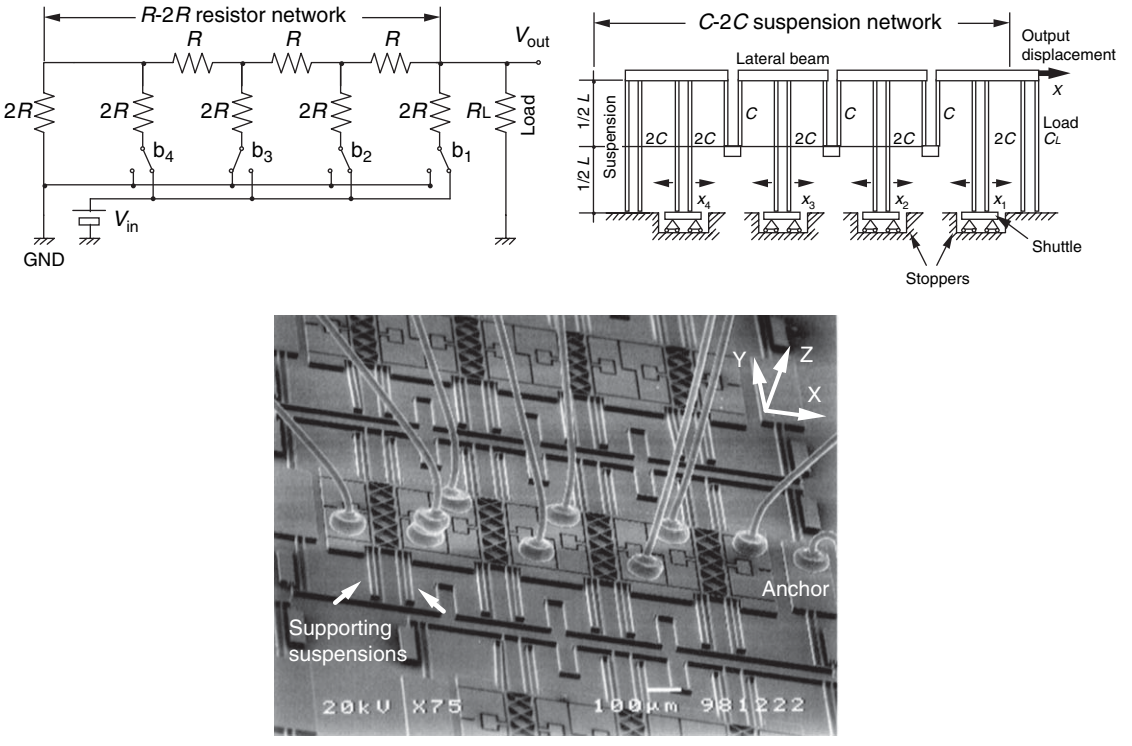


Figure 60 Microelectromechanical digital-to-analog converter of displacement (MEMDAC).

conversion of voltage using the R-2R resistor network. Each actuator in the network was an electrostatic shuttle moving back and forth between the drive electrodes and was operated by the corresponding digit (0 or 1) of input code. The N-bits of local displacement accumulate in the suspension network to synthesize an analog output displacement, which was proportional to the analog value coded with the N-bit input. Due to the analogy of electrical digital-to-analog converter, the mechanism was called microelectromechanical digital-to-analog converter (MEMDAC). The output displacement was free from the electrostatic drift or voltage fluctuation because the stroke of every shuttle was clipped by mechanical stoppers. They demonstrated a total stroke of 5.8 μm with an incremental step of 0.38 μm with a 4-bit MEMDAC. An electrothermal version of MEMDAC was also reported by Yie *et al.* (1999).

2.01.12 Electrostatic

2.01.12.1 Electrostatic Energy

Electric capacitance C between oppositely charged material bodies is defined by the amount of charge $\pm Q$ in Coulombs and the potential difference V in volts as follows:

$$Q = CV \qquad [54]$$

The electrostatic energy stored in the capacitor increased by the work that is done to pump up a small charge $\mathrm{d}Q$ against a potential difference $V = Q/C$. Integrating the work $(Q/C)\mathrm{d}Q$ until the bodies are charged to Q, the total work $((Q/C)\mathrm{d}Q$ until the bodies are charged to Q, the total work (energy) is written as

$$W = \int_0^Q \frac{Q}{C}\mathrm{d}Q = \frac{1}{2}\frac{Q^2}{C} \qquad [55]$$

Rewriting this equation in terms of voltage, we will obtain the stored energy as

$$U = \frac{1}{2}CV^2 \qquad [56]$$

2.01.12.2 Electrostatic Capacitance of Parallel Plates

Assuming a pair of conductive plates of infinite area placed in parallel with a gap g, we artificially cut out a small fraction of area S, where the electrostatic flux is uniformly distributed with its direction normal to the surface. Applying Gauss's law

$$\int_S \mathrm{div}\,\mathbf{E}\,\mathrm{d}S = \frac{1}{\varepsilon_0}\int_V \rho\,\mathrm{d}V$$

to the closed surface containing the small upper plate, we obtain the field strength E_{upper} to be $Q/(2S\varepsilon_0)$. In the same manner, the field strength originated from the lower plate is written to be $E_{\mathrm{lower}} = Q/(2S\varepsilon_0)$ in the same direction, and therefore, the net field strength is $E_{\mathrm{upper}} + E_{\mathrm{lower}} = Q/(S\varepsilon_0)$. The potential difference V can be obtained by multiplying the field strength by the plate distance g, and thus, $V = Qg/(S\varepsilon_0)$. Substituting this form in eqn [54], the capacitance of the parallel plates is written as

$$C = \varepsilon_0\frac{S}{g} \qquad [57]$$

2.01.12.3 Electrostatic Force

The principle of virtual work is used to derive the electrostatic force of electrostatic actuator. We assume a suspended electrostatic actuator that has come to a displacement x, counterbalancing with the electrostatic force F under a constant drive voltage V. When an imaginary small displacement Δx occurs on the movable electrode toward the direction to increase the capacitance by ΔC, the electrostatic energy stored in the capacitor increases by $1/2(\ \Delta CV^2)$. At the same time, the movable electrode gives a work $F\Delta x$ to external. The energy for these two components is contributed by the voltage source, which has given the capacitor the additional amount of charge ΔQ against the potential difference V. Hence, the energy balance is written as

$$\frac{1}{2}\Delta CV^2 + F\Delta x = \Delta QV \qquad [58]$$

Using $\Delta Q = \Delta CV$ for this equation, one obtains

$$\begin{aligned}\frac{1}{2}\Delta CV^2 + F\Delta x &= \Delta CV^2 \\ F\Delta x &= \frac{1}{2}\Delta CV^2\end{aligned} \qquad [59]$$

By taking the limit of $\Delta x \to 0$, the output force of an electrostatic actuator driven by a constant voltage is written as

$$F = \lim_{\Delta x \to 0}\frac{1}{2}\frac{\Delta C}{\Delta x}V^2 = \frac{1}{2}\frac{\mathrm{d}C}{\mathrm{d}x}V^2 \qquad [60]$$

References

Akiyama T, Shono K 1993 Controlled stepwise motion in polysilicon microstructures. *J. Microelectromech. Syst.* **2(3)**, 106–10

Akiyama T, Collard D, Fujita H 1997 Scratch drive actuator with mechanical links for self-assembly of three-dimensional MEMS. *J. Microelectromech. Syst.* **6**, 10–17

Aksyuk V A, Pardo F, Carr D, Greywall D, Chan H B, Simon M E, Gasparyan A, Shea H, Lifton V, Bolle C, Arney S, Frahm R, Paczkowski M, Haueis M, Ryf R, Neilson D T, Kim J, Giles C R, Bishop D 2003 Beam-steering micromirrors for large optical cross-connects. *J. Lightwave Tech.* **21(3)**, 634–42

Bollee B 1969 Electrostatic motors. *Philips Tech. Rev.* **30(6/7)**, 178–94

Brosnihan T J, Brown S A, Brogan A, Gormley C S, Collins D J, Sherman S J, Lemkin M, Polce N A, Davis M S 2003 Optical iMEMS: A fabrication process for MEMS optical switches with integrated on-chip electronics. *Dig. Tech. Papers 12th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers 03)*, Boston, MA, USA, June 8–12, 2003, pp. 1638–42

Bustillo J M, Howe R T, Muller R S 1998 Surface micromachining for microelectromechanical systems. *Proc. IEEE* **86(8)**, 1552–74

Chan E K, Dutton R W 2000 Electrostatic micromechanical actuator with extended range of travel. *J. Microelectromech. Syst.* **9**, 321–8

Chen J, Weingartner W, Azarov A, Giles R C 2004 Tilt-angle stabilization of electrostatically actuated micromechanical mirrors beyond the pull-in point. *J. Microelectromech. Syst.* **13(6)**, 988–97

Conant R A 2003 *Micromachined Mirrors*. Kluwer Academic Publishers, Norwell, MA

Core T A, Tsang W K, Sherman S J 1993 Fabrication technology for an integrated surface-micromachined sensor. *Solid State Technol.* **26(10)**, 39–47

CoventorWare software, www.coventor.com

Dadap J I, Chu P B, Brener I, Pu C, Lee C D, Bergman K, Bonadeo N, Chau T, Chou M, Doran R, Gibson R, Harel R, Johnson J J, Lee S S, Park S, Peale D R, Rodriguez R, Tong D, Tsai M, Wu C, Zhong W, Goldstein E L, Lin L Y, Walker J A 2003 Modular MEMS-based optical cross-connect with large port-count. *Photon Tech. Lett.* **15(12)**, 1773–5

Dakin T W, Luxa G, Oppermann G, Vigreux J, Wind G, Winkeinkemper H 1974 Breakdown of gases in uniform fields, Paschen curves for nitrogen, air, sulfur hexafluoride. *Electra* **32**, 61–82

Daneman M J, Tien N C, Solgaard O, Pisano A P, Lau K Y, Muller R S 1996 Linear microvibromotor for positioning optical components. *J. Microelectromech. Syst.* **5(3)**, 159–65

Egawa S, Higuchi T 1990 Multi-layered electrostatic film actuator. *Proc. IEEE Micro Electro Mechanical Systems Workshop*, Napa Valley, CA, USA, February 11–14, 1990, pp. 166–71

Fan L S, Tai Y C, Muller R S 1988 IC-processed electrostatic micro motors. *1988 IEEE. Int. Electron. Devices Meeting*, New York, NY, USA, pp. 666–9

Fan L, Wu M C, Choquette K, Crawford M 1997 Self-assembled microfabricated XYZ stages for optical scanning, alignment. *Proc. Int. Conf. Solid-State Sensors and Actuators (Transducers 97)*, Chicago, IL, USA, June 16–19, 1997, pp. 319–22

Feynman R P, Leighton R B, Sands M 1964 *The Feynman Lectures on Physics.* Addison-Wesley Publishing Company, Boston, MA, Vol. II, Chapter 6

Ford J E, Walker J A, Greywall D S, Goossen K W 1998 Micromechanical fiber-optic attenuator with 3 μs response. *J. Lightwave Technol.* **16(9)**, 1663–70

Gere J M, Timoshenko S P 1997 *Mechanics of Materials*, 4th edn. PWS, Boston, MA

Gorthi S, Mohanty A, Chatterjee A 2006 Cantilever beam electrostatic MEMS actuators beyond pull-in. *J. Micromech. Microeng.* **16**, 1800–10

Higuchi T, Watanabe M, Kudoh K 1988 Precise positioner utilizing rapid deformation of piezo electric element. *J. Jpn. Soc. Precision Eng.* **54**, 2107–12

Hirano T, Furuhata T, Gabriel K J, Fujita H 1992 Design, fabrication, operation of sub-micron gap comb-drive microactuators. *J. Microelectromech. Syst.* **1**, 52–9

IntelliSuite software, www.intellisensesoftware.com/

Isamoto K, Kato K, Morosawa A, Chong C, Fujita H, Toshiyoshi H 2004 A 5-V operated MEMS variable optical attenuator by SOI bulk micromachining. *IEEE J. Sel. Top. Quantum Electron.* **10(3)**, 570–8

Isamoto K, Makino T, Morosawa A, Chong C, Fujita H, Toshiyoshi H 2005 Self-assembly technique for MEMS vertical comb electrostatic actuators. *IEICE Electron. Express* **2(9)**, 542–7

Jefimenko O D 1973 Electrostatic motors. *Electrostatics and Its Applications*. John Wiley and Sons, Inc., New York, Chapter 7

Kim C H, Joeng H M, Jeon J U, Kim Y K 2003 Silicon micro XY-stage with a large area shuttle, no-etching holes for SPM-based data storage. *J. Microelectromech. Syst.* **12(4)**, 470–8

Kim J, Christensen D, Lin L 2005 2-D scanning mirror using plastically deformed angular vertical comb drive actuator. *Proc. 13th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers 05)*, Seoul, Korea, June 5–9, 2005, pp. 697–700

Kwon H N, Hwang I-H, Lee J-H 2005 A pulse-operating electrostatic microactuator for bi-stable latching. *J. Micromech. Microeng.* **15**, 1511–16

Kwon H N, Lee J-H, Takahashi K, Toshiyoshi H 2006a Micro XY stages with spider-leg actuator for two-dimensional optical scanning. *Sens. Actuators* **A130-131**, 468–77

Kwon H N, Lee J-H, Takahashi K, Toshiyoshi H 2006b Optical characterization of 9×9 optical cross connect utilizing silicon lens scanners with spider-leg actuators. *Proc. SPIE Photonics West, MOEMS-MEMS 2006, SPIE*, San Jose, CA, USA, Vol. 6114, January 25–26, 2006, pp. 131–8

Kwon S, Lee L P 2002 Stacked two dimensional micro-lens scanner for micro confocal imaging array. *Proc. 15th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2002)*, Las Vegas, NV, USA, January 20–24, 2002, pp. 483–6

Mathcad (Mathsoft home page), www.mathsoft.com/

Mehregany M, Gabriel K J, Trimmer W S N 1988 Integrated fabrication of polysilicon mechanisms. *IEEE Trans. ED* **35(6)**, 719–23

Mehregany M, Bart S F, Tavrow L S, Lang J H, Senturia S D, Schlecht M F 1990a A study of three microfabricated variable-capacitance motors. *Sens. Actuators* **A21**, 173–9

Mehregany M, Nagarkar P, Senturia S D, Lang J H 1990b Operation of microfabricated harmonic and ordinary side-drive motors. *Proc. IEEE Micro Electro Mechanical Systems Workshop*, Napa Valley, CA, USA, February 11–14, 1990, pp. 1–8

Minotti P, Langlet P, Bourbon G, and Masuzawa T 1998 Towards new cylindrical electrostatic micromotors using tubular combinations of arrayed direct drive actuators. *Jpn. J. Appl. Phys.* 37 part 2: L622–5.

Mita M, Arai M, Tensaka S, Kobayashi D, Fujita H 2003 A micromachined impact microactuator driven by electrostatic force. *J. Microelectromech. Syst.* **12**, 37–41

MUMPs: Multi-User MEMS Processes, MEMSCAP, www.memscap.com

Nadal-Guardia R, Dehe A, Aigner R, Castaner L M 2002 Current drive methods to extend the range of travel of electrostatic microactuators beyond the voltage pull-in point. *J. Microelectromech. Syst.* **11(3)**, 255–63

Nathanson H C, Wickstrom R A 1965 A resonant-gate silicon surface transistor with high-Q bandpass properties. *Appl. Phys. Lett.* **7**, 84

Pai M, Tien N C 1997 Current-controlled bidirectional electrothermally actuated vibromotor. *Proc. Transducers 97*, Chicago, IL, USA, June 16–19, 1997, pp. 3P5_7

Paschen F 1889 Ueber die zum Funkenubergang in Luft, Wasserstoff und Kohlensaure bei verschiedenen Drucken erforderlichr Potentialdifferenz. *Ann. Phys.* **27**, 69–96

Patterson P R, Su G-D J, Toshiyoshi H, Wu M C 2000 A MEMS 2-D scanner with bonded single-crystalline honeycomb micromirror. *Solid-State Sensor and Actuator Workshop (Hilton Head 2000)*, Hilton Head Island, SC, USA, June 4–8, 2000, pp. 17–18

Patterson P R, Hah D, Nguyen H, Toshiyoshi H, Chao R-M, Wu M C 2002 A scanning micromirror with angular comb drive actuation. *Proc. 15th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 202)*, Las Vegas, NV, USA, January 20–24, 2002, pp. 544–7

Petersen K E 1982 Silicon as a mechanical material. *Proc. IEEE* **70(5)**, 420–57

Pister K S J, Judy M W, Burgett S R, Fearing R S 1992 Microfabricated hinges. *Sens. Actuators* **A33(3)**, 249–56

Piyawattanametha W, Patterson P R, Yah D, Toshiyoshi H, Wu M C 2005 Surface-, bulk-micromachined two-dimensional scanner driven by angular vertical comb actuators. *J. Microelectromech. Syst.* **14(6)**, 1329–38

Price R, Wood J, Jacobsen S 1989 Modeling considerations for electrostatic forces in electrostatic microactuators. *Sensors and Actuators* **20(1–2)**, 107–14

Roessig T, Arakelian A, Brosnihan T, Clark B, Han D, Judy M, Juneau T, Lemkin M, Sherman S, Swift J 2002 Mirrors with integrated position-sense electronics for optical-switching applications. *Analog Dialogue* **36(4)**, 1–2

Sandia National Laboratories, http://mems.sandia.gov/

Sarajlic E, De Boer M J, Jansen H V, Arnal N, Puech M, Krijnen G, Elwenspoek M 2005 Bulk micromachining technology for fabrication of two-level MEMS in standard silicon substrate. *Proc. 13th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers 05)*, Seoul, Korea, June 5–9, 2005, pp. 1404–5

Saruta K, Fujita H, Toshiyoshi H 2002 Bulk micromachined 2D lens scanners for transparent optical fiber switches. *Conf. Dig. 2002 IEEE/LEOS Int. Conf. Optical MEMS (Optical MEMS 2002)*, Hotel de la Paix, Lugano, Switzerland, August 20–23, 2002, pp. 13–14

Seeger J I, Crary S B 1997 Stabilization of electrostatically-actuated mechanical devices. *Proc. Int. Conf.. Solid State Sensors and Actuators (Transducers 1997)*, Chicago, IL, USA, Vol. 2, pp. 1133–6

Sniegowski J J, Garcia E J 1996 Surface micromachined gear trains driven by an on-chip electrostatic microengine. *IEEE Electron Device Lett.* **17**, 366–8

Tai Y C, Muller R S 1989 IC-processed electrostatic synchronous microrotors. *Sens. Actuators* **20(1/2)**, 41–8

Tai Y C, Muller R S 1990 Friction study of IC processed micromotors. *Sens. Actuators* **A21**, 180–3

Takahashi K, Kwon H N, Mita M, Fujita H, Toshiyoshi H, Suzuki K, Funaki H 2005 Monolithic integration of high voltage driver circuits, MEMS actuators by ASIC-like postprocesses. *Proc. 13th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers 05)*, Seoul, Korea, June 5–9, 2005, pp. 417–20

Takahashi K, Mita M, Fujita H, Toshiyoshi H 2007 Topological layer switch technique for monolithically integrated electrostatic XYZ-stage. *Proc. 20th IEEE Int. Micro Electro Mechanical Systems (MEMS) Conference*, January 21–25, 2007, Kobe, Japan, pp. 651–4

Takahashi K, Mita M, Fujita H, Toshiyoshi H 2006a A high fill-factor comb-driven XY-stage with topological layer switch architecture. *IEICE Electron. Express* **3(9)**, 197–202

Takahashi T, Mita M, Motohara K, Kobayashi N, Kashikawa N, Fujita H, Toshiyoshi H 2006b Electrostatically addressable gatefold micro-shutter arrays for astronomical infrared spectrograph. *Proc.Asia Pacific Conf. Transducers, Micro-Nano Technology*, Marina-Mandarin Hotel, Singapore, June 25–28, 2006

Takeshima N, Gabriel K J, Ozaki M, Takahashi J, Horiguchi H, Fujita H 1991 Electrostatic parallelogram actuators. *Proc. Int. Conf. Solid-State Sensors and Actuators (Transducers 91)*, San Francisco, CA, USA, June 24–27, 1991, pp. 63–6

Tang W C, Nguyen T-C, Howe R T 1989 Laterally driven polysilicon resonant microstructures. *Sens. Actuators* **20**, 25–32

Timoshenko S P, Goodier J N 1970 *Theory of Elasticity*, 3rd edn. McGraw Hill, New York

Toshiyoshi H, Fujita H 1996 Electrostatic micro torsion mirror for an optical switch matrix. *J. Microelectromech. Syst.* **5(4)**, 231–7

Toshiyoshi H, Goto M, Mita M, Fujita H, Kobayashi D, Hashiguchi G, Endo J, Wada Y 1999 Fabrication of micromechanical tunneling probes and actuators on a silicon chip. *Jpn. J. Appl. Phys.* **38(12B, Part 1)**, 7185–9 [Special issue on Microprocesses and Nanotechnology]

Toshiyoshi H, Kobayashi D, Mite M, Hashiguchi G, Fujita H, Endo J, Wada Y 2000 Microelectromechanical digital-to-analog converters of displacement for step motion actuators. *J. Microelectromech. Syst.* **9(2)**, 218–25

Toshiyoshi H, Piyawattanametha W, Chan C-T, Wu M C 2001 Linearization of electrostatically actuated surface micromachined 2-D optical scanner. *J. Miroelectromech. Syst.* **10(2)**, 205–14

Toshiyoshi H, Mite M, Fujita H 2002 A MEMS piggyback actuator for hard-disk drives. *J. Microelectromech. Syst.* **11(6)**, 648–54

Trimmer W S N, Gabriel K J 1987 Design consideration for a practical electrostatic micro-motor. *Sens. Actuators* **11(2)**, 189–206

Tsai J-C, Wu M C 2005 Gimbal-less MEMS two-axis optical scanner array with high fill-factor. *J. Microelectromech. Syst.* **14(6)**, 1323–8

Van Kessel P F, Hornbeck L J, Meier R E, Douglass M R 1998 A MEMS-based projection display. *Proc. IEEE* **86(8)**, 1688–704

Wibbeler J, Pfeifer G, Hietschold M 1998 Parasitic charging of dielectric surfaces in capacitive microelectromechanical systems (MEMS). *Sens. Actuators* **A71(1–2)**, 74–80

Yano M, Yamagishi F, Tsuda T 2005 Optical MEMS for photonic switching – compact, stable optical crossconnect switches for simple, fast, flexible wavelength applications in recent photonic networks. *IEEE J. Sel. Top. Quantum Electron.* **11(2)**, 383–94

Yie R, Conant R A, Pister K S J 1999 Mechanical digital to analog converters. *Proc. 10th Int. Conf. Solid State Sensors and Actuators (Transducers 99)*, Sendai, Japan, June 7–10, 1999, pp. 998–1001

Yoda M, Isamoto K, Chong C, Ito H, Murata A, Kamisuki S, Atobe M, Toshiyoshi H 2005 A MEMS 1-D optical scanner for laser projection display using self-assembled vertical combs, scan-angle magnifying mechanism. *Proc. 13th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers 05)*, Seoul, Korea, June 5–9, 2005, pp. 968–71

Zicker M, Mita M, Ataka M, Fujita H 2002 Design, simulation, fabrication, evaluation of electrostatic vertical comb-drive actuators. *Int. Conf. Elect. Eng. (ICEE-2002)*, Jeju Island, Korea, July 7–11, 2002, pp. 1449–52

Biography

Hiroshi Toshiyoshi received his M.Eng. and Ph.D. degrees in Electrical Engineering from the University of Tokyo, Tokyo, Japan, in 1993 and 1996, respectively. Since April 2002, he has been an Assistant Professor with the Institute of Industrial Science (IIS), the University of Tokyo, Tokyo, Japan. From April 1999 to March 2001, he had been a Visiting Assistant Professor at University of California, Los Angeles for his sabbatical years. He is currently a Co-Director of the first Japan–France international research collaboration LIMMS (Laboratories for Integrated Micro Mechatronic Systems) established between IIS and Centre National de la Recherche Scientifique (CNRS), France. He also conducts a three-year project on Optomechatronics at Kanagawa Academy of Science & Technology (KAST), Kanagawa, Japan since April 2005. His main research interest is optical and RF MEMS.

2.02 Magnetic Actuation

Kwang W. Oh[1] **and Chong H. Ahn**[2], [1]SMALL (nanobioSensors and MicroActuators Learning Lab), Electrical Engineering, University at Buffalo, the State University of New York (SUNY-Buffalo), Buffalo, NY, USA, [2]Center for BioMEMS and NanoBiosystems, Electrical and Computer Engineering and Computer Science, Biomedical Engineering (Secondary), University of Cincinnati, Cincinnati, OH, USA

Published by Elsevier B.V.

Glossary

LOC Lab-on-a-chip
NiFe Permalloy
emf Electromotive force
DNA Deoxyribonucleic acid
RNA Ribonucleic acid
MHD Magnetohydrodynamic
MEMS Microelectromechanical systems
RF Radio frequency
MACS Magnetic cell separation
GMR Giant magnetoresistive
SQUID Superconducting quantum interference devices

2.02.1 Introduction

We have all played with magnets. A pair of magnets by itself makes a wonderful toy. The most familiar magnets may be the small decorative magnets used to fasten notes to the refrigerator door. As you fasten the notes to the refrigerator door with the magnet, you become aware with your fingertips that the space around the magnet has attractive forces. In another familiar type of magnet, a coil is wound around an iron core, the strength

of the external magnetic field being determined by the current in the coil. These two types of magnets, permanent magnets and electromagnets, play a much larger part in our daily lives as well as in microsystems.

At the microscale, magnetic actuation is one of the more promising actuation mechanisms compared to other such mechanisms. Most microactuators actuate mechanical moving parts using electrostatic, piezoelectric, thermal, magnetic, or other actuation mechanisms, as shown in **Figure 1** (Oh and Ahn 2006). For miniaturized structures, electrostatic actuation becomes more attractive due to their fast response time and reliability. However, it is difficult to achieve large forces and deflections because of the extremely high voltages required. Piezoelectric actuation can yield very large forces, but very small deflections even with very high voltages. The drawback of small strokes has been overcome by the hydraulic amplification of the piezoelectric, stacked piezoelectric disks or piezo bimorphs. Thermal actuation can provide large forces via large strokes, but is relatively slow and may not be suitable for many fluids due to heat dissipation. Thermal actuators require a large amount of thermal energy for their operation, and therefore consume substantial electrical energy. High temperature and complicated thermal management are further drawbacks of thermal actuation.

On the other hand, many microactuators employ magnetic actuation, which often involves permanent magnets and electromagnetic coils. Magnetic actuation enables large forces and deflections with low power consumption. The relative advantages and disadvantages of electrostatic actuation, piezoelectric actuation, thermal actuation, and magnetic actuation are well summarized by Liu (2006) in **Table 1**.

Today we can see magnetic actuation being combined with microsystems in an amazing variety of ways. Magnetic actuation can be utilized to miniaturize pumps, valves, mixers, mirrors, and switches. Applications include valving, pumping, and mixing of fluids, as well as the incorporation of mirrors and switches into microsystems. In addition, magnetic forces are used to transport, position, separate, and sort magnetic and nonmagnetic objects such as magnetic particles, magnetically labeled cells, or plugs of ferrofluids. Magnetic actuation is also closely related to magnetic sensing, as magnetic fields are used to produce and sense motion. In particular, magnetic sensors are used to generate an electrical signal to sense motion (Ripka 2001). Indeed, many materials, components, and fabrication methods for magnetic actuation originated in the field of magnetic sensing.

In this chapter, we begin our study of magnetic actuation by discussing the basic theories and

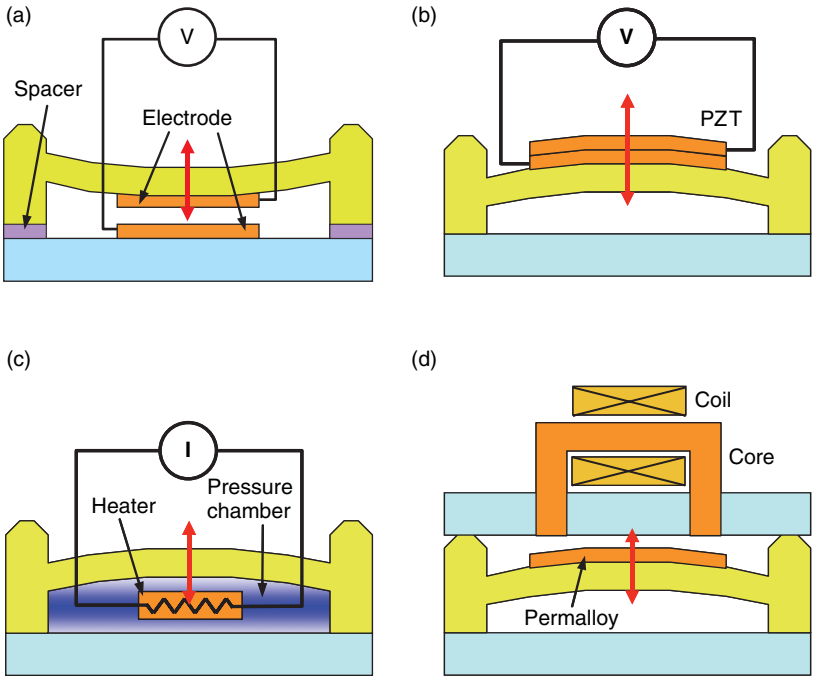


Figure 1 Illustrations of actuation principles of microactuators with mechanical moving parts: (a) electrostatic (b) piezoelectric; (c) thermal; and (d) electromagnetic actuation. The spacer in the electrostatic actuation is working as an insulating layer.

Table 1 Comparison of various actuation methods

	Advantages	*Disadvantages*
Electrostatic actuation	Simplicity of materials; fast actuation response	Trade-off between magnitude of force and displacement; susceptible to pull-in limitation
Piezoelectric actuation	Fast response possible; capable of achieving moderately large displacement	Requires complex material preparation; degraded performance at low frequencies
Thermal actuation	Capable of achieving large displacement (angular or linear); moderately fast actuation response	Relatively large power consumption; sensitivity to environmental temperature changes
Magnetic actuation	Capable of generating large angular displacement; possibility of using very strong magnetic force as bias	Moderately complex processes; difficult to form on-chip, high-efficiency solenoids

Source: Liu (2006).

actuation mechanisms pertaining to microscale magnetic actuation. Then, representative magnetic materials and fabrication processes that are key building blocks in magnetic microsystems are reviewed. Finally, selected examples of magnetic microactuators in the areas of microfluidics, microoptics, microelectronics, and lab-on-a-chip (LOC) devices are discussed.

2.02.2 Principles

2.02.2.1 Scaling Effect

As discussed by Judy (2001), the scaling of any device or system from the macroscale to the microscale dramatically affects the overall design, materials, and fabrication processes. The author reviewed the influences of scaling on material properties, mechanical systems, fluidic systems, chemical and biological systems, thermal systems, electrical and magnetic systems, and optical systems. Especially when scaling down electrical and magnetic systems, there are preferable dimensions suitable for electrostatic or magnetic actuation, yet the dimensions are in the micrometer range.

As summarized in **Table 2**, it is possible to estimate the dimensions where magnetic forces are still dominant. In the case of an electrostatic actuator, the maximum energy density is limited by the maximum electric field that can be applied before electrical breakdown occurs. This maximum electric field is approximately $3\,\mathrm{M\,V\,m^{-1}}$, giving an energy density in vacuum on the order of $4 \times 10^{1}\,\mathrm{J\,m^{-3}}$. In the case of an equivalent magnetostatic actuator, the maximum energy that can be produced is determined essentially by the maximum value of the flux density of the existing materials. Typically, the flux density is on the order of 1 T or $1\,\mathrm{V\,s\,m^{-2}}$, giving a magnetostatic energy of $4 \times 10^{5}\,\mathrm{J\,m^{-3}}$. The ratio of the magnetostatic to the electrostatic energy density at the macroscale is then

$$\frac{U_{\text{magnetostatic\_macro\_max}}}{U_{\text{electrostatic\_macro\_max}}} \approx 10^{4} \qquad [1]$$

which shows why electromagnetic forces dominate in the macroworld and electrostatic forces appear only in very few cases.

Scaling down of the dimension in electrostatic actuation leads to new phenomena. For example, the breakdown voltage is a function of the electrode gap, and a plot of this is known as the Paschen curve

Table 2 Comparison of scaling electrical and magnetic systems

	Electrostatic	*Magnetostatic*
Energy density	$U_{\text{electrostatic}} = \frac{1}{2}\varepsilon E^2$	$U_{\text{magnetostatic}} = \frac{1}{2}\mu^{-1}B^2$
Macroscale	$E_{\max} = \sim 3\,\mathrm{M\,V\,m^{-1}}$ $U_{\max} = \sim 40\,\mathrm{J\,m^{-3}}$	$B_{\max} = \sim 1\,\mathrm{T}$ (or $1\,\mathrm{V\,s\,m^{-2}}$) $U_{\max} = \sim 4\,00\,000\,\mathrm{J\,m^{-3}}$
Microscale	$E_{\max} = F(x)$ (function of electrode gap) $U_{\text{micro\_max}} >> U_{\text{macro\_max}}$ (function of electrode gap)	$B_{\max} =$ const. (dependence on materials) $U_{\text{micro\_max}} = U_{\text{macro\_max}}$ ($10^{6}\,\mathrm{J\,m^{-3}}$ for saturated Fe, $10^{5}\,\mathrm{J\,m^{-3}}$ for saturated Ni)

ε, permittivity; E, electric field; μ, magnetic permeability; B, magnetic flux density; and x, electrode gap
Sources: Judy (2001) and Niachos (2003).

(Von Hippel 1959). Apparently, there are cross-over points from magnetic actuation to electrostatic actuation at the microscale. The points can be estimated by assuming that the maximum electric field is a function of the electrode gap and the maximum magnetic flux density is a constant. They are determined from the above assumption that

$$U_{\text{electrostatic\_micro\_max}} = U_{\text{magnetostatic\_micro\_max}} \quad [2]$$

in which

$$\begin{aligned} U_{\text{electrostatic\_micro\_max}} &= f(x) \\ U_{\text{magnetostatic\_micro\_max}} &= const \end{aligned} \quad [3]$$

This simplified approach gives an order of magnitude of ~2 μm, considering that the most common magnetic materials and structures used for magnetic actuators are iron- and nickel-based alloys (e.g., $10^6\,\text{J m}^{-3}$ for saturated Fe systems and $10^5\,\text{J m}^{-3}$ for saturated Ni systems). In practice, however, the macroscopic assumption is not valid in magnetic microactuators. The size of magnetic domains, i.e., regions of a material with uniform magnetization, is typically on the scale of micrometers in permalloys (NiFe), which are commonly used in magnetic microactuators. Practical issues, such as magnetic domain wall structure, shape, and magnetocrystalline anisotropy as a function of thickness, should be addressed for a more accurate prediction. Still, magnetic actuation is advantageous above a boundary range from 1 to 10 μm gap or deflection (Niarchos 2003).

2.02.2.2 Theory

2.02.2.2.1 Lorentz force

An electromagnetic actuator utilizes the force generated from the magnetic field either through the Lorenz force or through magnetic induction, which we now discuss. Where does the magnetic field come from? Experiments show that it comes from moving electric charges. A charge sets up an electric field irrespective of whether the charge is at rest or is moving. However, a charge sets up a magnetic field only if it is moving. Where are these moving electric charges? In permanent magnets, these are the spinning and circulating electrons in the iron atoms that make up the magnet. In the electromagnets, these are the electrons circulating through the coils of wire that surround these magnets.

In principle, we can set a test charge q in motion in various directions and with various speeds, and can record the force $\mathbf{F}_B$ that acts on the moving charge. If we analyze the results of many such trials, we find that the vector $\mathbf{F}_B$ is related to the velocity vector $\mathbf{v}$ by the following equation:

$$\mathbf{F}_B = q\mathbf{v} \times \mathbf{B} \quad \text{(Lorentz force law for magnetism)} \quad [4]$$

where $\mathbf{B}$ is the magnetic field vector and the SI unit of $\mathbf{B}$ is $\text{N C}^{-1}\,\text{m}^{-1}\,\text{s}$). For convenience, this is called the tesla (T). Recalling that $1\,\text{C s}^{-1} = 1\,\text{A}$, we have

$$1\,\text{tesla} = 1\,\text{T} = 1\,\text{N A}^{-1}\,\text{m}^{-1} \quad [5]$$

An earlier non-SI unit for $\mathbf{B}$, still in common use, is the gauss; the relation is as follows:

$$1\,\text{tesla} = 10^4\,\text{gauss} \quad [6]$$

Note that the earth's magnetic field near the earth's surface is about 10^{-4} T (1 gauss).

If we set up a wire segment of length L making an angle ϕ with a magnetic field (**Figure 2**), the magnetic force is given by the following equation:

$$\mathbf{F}_B = i\mathbf{L} \times \mathbf{B} \quad [7]$$

where $\mathbf{L}$ is a vector that points along the wire segment in the direction of the current. Eqn [7] is equivalent to eqn [4] in that either can be taken as the defining equation for the magnetic field vector $\mathbf{B}$. Equivalently, we can express the magnetic force in terms of the current density $\mathbf{J}$ as

$$\mathbf{F}_B = \int_V \mathbf{J} \times \mathbf{B}\,dV \quad [8]$$

and we can find the resultant force on any given structure of currents by integrating eqn [8] over that structure.

A fundamental fact of electrostatics is that two charges exert forces on each other. Charges produce

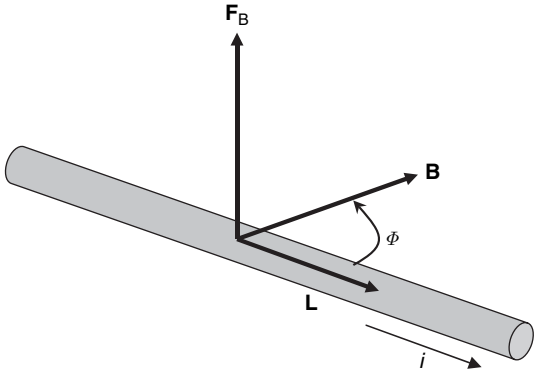


Figure 2 A wire segment of length L makes an angle Φ with a magnetic field.

electric fields and electric fields exert forces on charges. A fundamental fact of magnetics is that two parallel wires carrying currents also exert forces on each other. Currents produce magnetic fields and magnetic fields exert forces on currents. The magnitude of electric field d**E** is given by the following equation:

$$\mathrm{d}\mathbf{E} = \frac{1}{4\pi\varepsilon_0}\frac{\mathrm{d}q}{r^3}\mathbf{r}\,(\text{Coulomb's law}) \qquad [9]$$

where **r** is the distance from the charge element to point P and ε_0 is the permittivity constant. A charge element $\mathrm{d}q$ establishes a differential electric field element d**E** at the point P. By analogy, we can easily guess the magnitude of the magnetic field contribution. The magnetic field contribution d**B** at point P by a given current element is given by the following equation:

$$\mathrm{d}\mathbf{B} = \frac{\mu_0}{4\pi}\frac{i\,\mathrm{d}\mathbf{s}}{r^3} \times \mathbf{r}\,(\text{Biot–Savart law}) \qquad [10]$$

where the vector d**s** is a differential element of length, pointing tangent to the wire in the direction of the current, and μ_0 is the permeability constant. A current element i d**s** establishes a differential magnetic field d**B** at the point P. Eqn [10], called the Coulomb's law for magnetism, is more often called the Biot–Savart law. We at once note a complexity in our analogy of the electrostatic case; the differential charge element $\mathrm{d}q$ is a scalar quantity but the differential current element i d**s** is a vector quantity.

One of the simplest problems in magnetism is that of a long straight wire carrying a current i. The magnitude of the magnetic field $B(r)$ at a perpendicular distance r from such a wire is given by the following equation:

$$B(r) = \frac{\mu_0\, i}{2\pi\, r}(\text{long straight wire}) \qquad [11]$$

Two long parallel wires carrying currents exert forces on each other. **Figure 3** shows two such wires, separated by a distance d and carrying currents i_A and i_B. Wire A produces a magnetic field B_A at all points. The magnitude of B_A at the site of wire B is, from eqn [11], is given as follows:

$$B_\mathrm{A} = \frac{\mu_0\, i_\mathrm{A}}{2\pi\, d} \qquad [12]$$

Wire B, which carries a current i_B, finds itself immersed in an external magnetic field $\mathbf{B}_\mathrm{A}$. A length L of this wire experiences a sideways magnetic force given by eqn [7], whose magnitude is given as follows:

$$F_\mathrm{BA} = i_\mathrm{B}\, L\, B_\mathrm{A} = \frac{\mu_0\, L i_\mathrm{B}\, i_\mathrm{A}}{2\pi\, d}(\text{two long parallel wires}) \qquad [13]$$

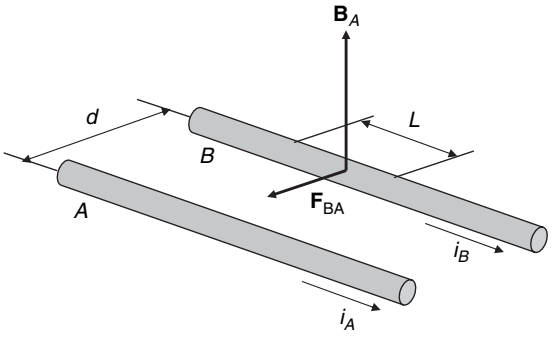


Figure 3 Two parallel wires carrying currents in the same direction attract each other. The magnetic field at wire B set up by the current in wire A is shown. For antiparallel currents, the two wires repel each other.

The rule of vector products says that $\mathbf{F}_\mathrm{BA}$ lies in the plane of the wires and points to the left.

2.02.2.2.2 Magnetic induction

In magnetic actuation, the magnetic field can also be generated by a permanent magnet or a coil. **Figure 4** shows two simple tabletop experiments, which will

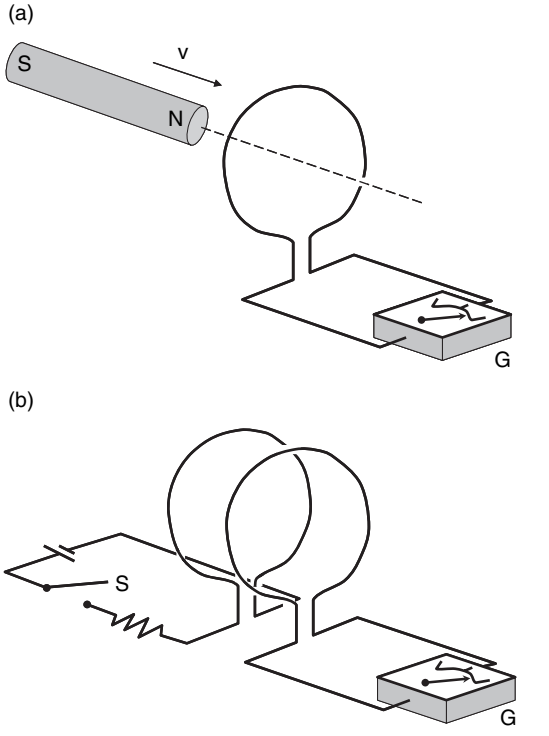


Figure 4 Two simple tabletop experiments guides us to understand Faraday's law of induction. (a) Galvanometer G deflects when the magnet is moving with respect to the coil. (b) Galvanometer G reflects momentarily when switch S is closed or opened. No physical motion of the coils is involved.

guide us to understand the Faraday's law of induction. An induced electromotive force (emf) appears only when something is changing. An induced emf appears in the right-hand loop of **Figure 4** only when the number of magnetic field lines that pass through that loop is changing. Whether you push the magnet toward the loop or pull it away from the loop, you will always experience a resisting force and will thus have to do work. In a static situation, in which no physical objects are moving and the currents are steady, there is no induced emf. This law, with a coil of N turns, in the equation form is as follows:

$$\text{emf} = -N\frac{\mathrm{d}\Phi_\mathrm{B}}{\mathrm{d}t}\,(\text{Faraday's Law}) \qquad [14]$$

where the magnetic flux Φ_B is defined from the following equation:

$$\Phi_\mathrm{B} = \int \mathbf{B}\cdot\mathrm{d}\mathbf{A} \qquad [15]$$

where d**A** is a differential element of surface area, and the integration should be carried out over the entire surface. If the magnetic field has a constant magnitude B and is everywhere at right angles to a plane surface of area A, the absolute value of the flux is $\Phi_\mathrm{B} = BA$. If the rate of change of flux is in webers (1 weber = 1 Wb = 1 T m^2) per second, the induced emf will be in volts.

An inductor is a coil such that if a current i is established in its windings, a magnetic flux Φ links each winding and the inductance L is given by

$$L = N\frac{\Phi}{i}\,(\text{inductance}) \qquad [16]$$

where N is the number of turns and the SI unit of inductance is henry (1 henry = 1 H = 1 T m^2 A^{-1}). By combining eqns [14] and [16], the self-induced emf is given as

$$\text{emf} = -N\frac{\mathrm{d}\Phi}{\mathrm{d}t} = -L\frac{\mathrm{d}i}{\mathrm{d}t} \qquad [17]$$

Thus, in any inductor, such as a coil, a solenoid, or a toroid, a self-induced emf appears whenever the current changes with time.

The two long, rigid, parallel wires (**Figure 3**) carrying current in the same direction attract each other, and energy has to be applied to pull them apart. We can get this stored energy back at any time by letting the wires move back to their original positions. In the case of magnetic actuation, the induction through the magnetic material results in a change in the stored energy U_B, which is given by

$$U_\mathrm{B} = \frac{1}{2}L\,i^2 \qquad [18]$$

In magnetic actuation, the attractive force induced by the current energizing the coil will be

$$F_\mathrm{B} = \frac{\partial U_\mathrm{B}}{\partial x} = \frac{1}{2}i^2\frac{\mathrm{d}L(x)}{\mathrm{d}x} \qquad [19]$$

To achieve a high stored magnetic energy and thus a large attractive force, both the magnetic film and the microcoil need to be designed and fabricated in such a way that they produce the large volume in a given planar area. The reader is referred to a study of an energy-based design criterion for magnetic actuation by one of the authors and colleagues (Nami *et al.* 1996). It suggests that the energy stored in the gap has a maximum value when the magnetic reluctance of the gap is equivalent to the magnetic reluctance of the magnetic core.

2.02.2.2.3 Magnetic torque

In magnetic actuation, a magnetic torque is widely used to rotate cantilever beams in micromirrors and microswitches. As shown in **Figure 5**, when the length l of a soft magnetic cantilever is much larger than its thickness t and width w, the direction along its long axis becomes the preferred direction for magnetization. When such a cantilever is placed in a uniform magnetic field $\mathbf{H}_0$, a magnetic torque $\boldsymbol{\tau}_\text{magnetic}$ exerted on the beam is (Ruan *et al.* 2001a, b)

$$\boldsymbol{\tau}_\text{magnetic} = \mathbf{m}\times\mathbf{B}_0 = \mu_0\mathbf{m}\times\mathbf{H}_0 \qquad [20]$$

where the magnetic induction of the applied uniform permanent magnetic field $\mathbf{B}_0$ and the magnetic moment $\mathbf{m}$ are defined by the following equation:

$$\mathbf{B}_0 = \mu_0\mathbf{H}_0\ (\text{magnetic induction}) \text{ and} \qquad [21]$$

$$\mathbf{m} = \mathbf{M}\,V\,(\text{magnetic moment}) \qquad [22]$$

where μ_0 is the permeability constant, $\mathbf{M}$ is the magnetization in the beam, and V is the volume of the beam.

When the magnetic force is large enough to overcome the mechanical force, the cantilever beam with one fixed end will deflect, as shown in **Figure 5**. According to the elasticity theory, the corresponding mechanical torque $\boldsymbol{\tau}_\text{magnetic}$ needed to cause a deflection Δz at the free end of the cantilever is (Benham *et al.* 1996)

$$\boldsymbol{\tau}_\text{mechanical} = \frac{3EI\Delta z}{l^2} = \frac{Ewt^3\Delta z}{4l^2} \qquad [23]$$

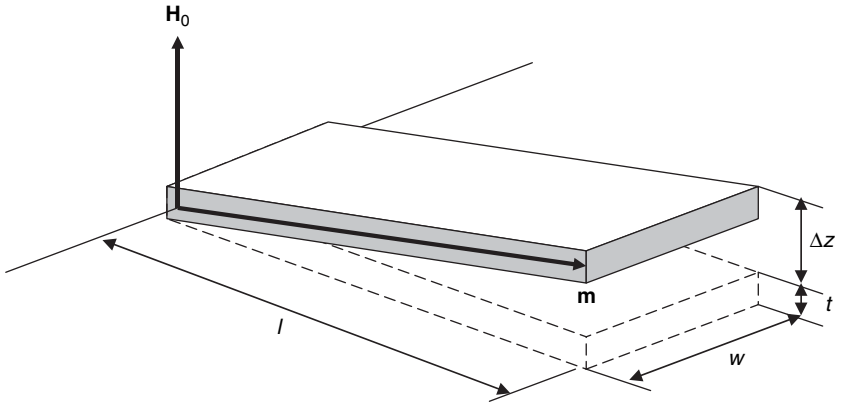


Figure 5 Schematic illustration of a soft magnetic cantilever beam actuator.

where E is Young's modulus of the cantilever beam material and I is the moment of inertia. When $\boldsymbol{\tau}_{\text{magnetic}} > \boldsymbol{\tau}_{\text{mechanical}}$, the cantilever beam will deflect. The larger deflection is possible with a stronger uniform magnetic field $\mathbf{H}_0$ and a stronger magnetization in the beam material.

2.02.2.2.4 Magnetic field gradient

In magnetic actuation, nonuniform magnetic field or magnetic field gradient is often used to transport and manipulate magnetic particles. A magnetic field gradient is required to exert a translation force. As studied in Section 2.02.2.2.3, a uniform field can produce a torque, but no translational action. According to electromagnetic field theory, the magnetic force of attraction acting on the magnetic moment $\mathbf{m}$ from the magnetic field $\mathbf{B}$ can be written as the derivative of the magnetic energy (Gijs 2004, Zborowski *et al.* 1999)

$$\mathbf{F}_{\mathbf{m}} = \nabla(\mathbf{m} \cdot \mathbf{B}) \qquad [24]$$

When we assume that the magnetic moment of the particle is not varying in space ($\nabla \cdot \mathbf{m} = 0$), the magnetic force on the magnetic particle is of the following form:

$$\mathbf{F}_{\mathbf{m}} = (\mathbf{m} \cdot \nabla)\mathbf{B} \qquad [25]$$

$$\mathbf{m} = V \chi_{\text{m}} \mathbf{H} \qquad [26]$$

where V is the volume of the magnetic particle, χ_{m} is the magnetic susceptibility of the particle, and $\mathbf{H}$ is the applied magnetic field intensity. By combining eqns [25] and [26], the magnetic force on the magnetic particle becomes

$$\mathbf{F}_{\mathbf{m}} = V \chi_{\text{m}} (\mathbf{H} \cdot \nabla)\mathbf{B} \qquad [27]$$

Eqn [27] can be expressed in x-, y-, and z-directional forces:

$$F_{\text{m},x} = V \chi_{\text{m}} \left(H_x \frac{\partial}{\partial x} + H_y \frac{\partial}{\partial y} + H_z \frac{\partial}{\partial z} \right) B_x \qquad [28]$$

$$F_{\text{m},y} = V \chi_{\text{m}} \left(H_x \frac{\partial}{\partial x} + H_y \frac{\partial}{\partial y} + H_z \frac{\partial}{\partial z} \right) B_y \qquad [29]$$

$$F_{\text{m},z} = V \chi_{\text{m}} \left(H_x \frac{\partial}{\partial x} + H_y \frac{\partial}{\partial y} + H_z \frac{\partial}{\partial z} \right) B_z \qquad [30]$$

Eqn [28] is preferred, for example, when the magnetic particles are constrained to move in a microchannel in the x-direction. This equation shows that both the magnitude of the applied magnetic field intensity for inducing a large moment and the magnetic field gradient need to be large to have a strong magnetic actuation force.

2.02.2.2.5 Magnetization

Several forms of magnetic behavior have been observed in different materials, including diamagnetism, paramagnetism, and ferromagnetism (Kittel 1996). Ferromagnetism is the normal form of magnetism that most people are familiar with, as exhibited in permanent magnets. However, unlike ferromagnetic materials, diamagnetic and paramagnetic materials do not retain any magnetization in the absence of an externally applied magnetic field. Diamagnetism is the most common magnetic behavior, which is proportional and opposing to the applied magnetic field. Diamagnetic materials are usually considered by nonphysicists as nonmagnetic and include water, DNA, most organic compounds such as oil and plastic, and many metals such as gold and bismuth. Paramagnetic materials present a

magnetization that is proportional to the applied field and adds force to it. Paramagnetism varies inversely with temperature and is characterized by the material's saturation magnetization. Another special case of paramagnetic materials is superparamagnetic particles that are magnetized in a magnetic field. However, once the external field is removed, the particles redisperse and behave like a nonmagnetic material.

The relationship between B and H can be described using the following equation:

$$B = \mu_0 H + M = \mu_0(H + \chi_m H) = \mu_r \mu_0 H \qquad [31]$$

where μ_0 is the magnetic permeability constant, μ_r is the relative permeability of the magnetic material, and M is the internal magnetization. The magnetic susceptibility χ_m is defined as

$$\chi_m = \mu_r - 1 \qquad [32]$$

Diamagnetic materials are repelled by the applied magnetic fields, and therefore they have a negative magnetic susceptibility ($\chi_m < 0$). Paramagnetic materials are attracted toward the applied magnetic fields, and therefore they have a positive magnetic susceptibility ($\chi_m < 0$). Diamagnetic and paramagnetic effects are usually so small that their effects are not observable in our daily lives. In general, the magnetic susceptibility on the order of 10^{-5}–10^{-6} with the negative sign for most diamagnets and on the order of 10^{-3}–10^{-5} with the positive sign for most paramagnets. Therefore, the relative permeability is very close to 1 for diamagnetic and paramagnetic materials.

For ferromagnetic materials such as iron, nickel, cobalt, and some rare earths, the values of the relative permeability are very large ($\chi_m \gg 0$). Ferromagnetic materials are often used in magnetic actuation applications. The linear relationship between B and H (eqn [31]) is valid only within a certain range of H. Ferromagnetic materials can present spontaneous magnetization, and this gives rise to the $B-H$ (or $M-H$) hysteresis loops. Ferromagnetic materials can be characterized by their permeability, Curie temperature (temperature of the phase change to paramagnetic behavior), coercive field (field strength needed to demagnetize the material), and remanent magnetization (magnetization at zero external field).

Hysteresis, the lack of retraceability, is well known in ferromagnetic materials. Magnetization curves for ferromagnetic materials do not retrace themselves as we increase and decrease the applied external magnetic field H. **Figure 6** shows the following operations with a sort of S-shaped loop:

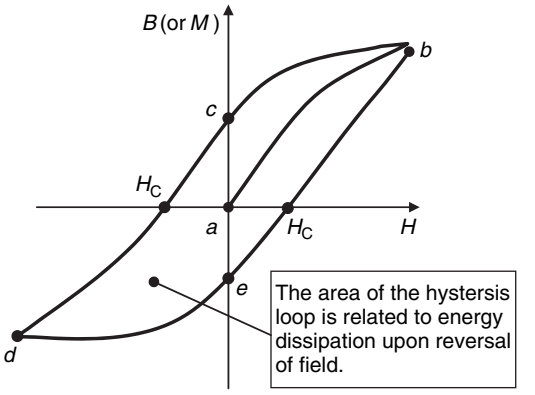


Figure 6 A magnetization curve (ab) for a ferromagnetic specimen and an associated hysterisis loop (bcdeb).

(1) Starting with an unmagnetized material (point *a*), increase the external magnetic field until H has the value corresponding to point *b*.
(2) Reduce the field back to zero (point *c*).
(3) Reverse the field and increase it in magnitude until point *d* is reached.
(4) Reduce the field to zero again (point *e*).
(5) Reverse the field once more until point *b* is reached again.

Note that at points *c* and *e* the material is magnetized, even though there is no field; this is the familiar phenomenon of permanent magnetism, and is called the remanent magnetization. This memory effect of magnetic materials is essential for the magnetic storage of information, as in cassette tapes or computer disks.

The coercivity of a ferromagnetic material is the intensity of the applied magnetic field required to reduce the magnetization of that material to zero after the magnetization of the sample has been driven to saturation. Coercivity is usually measured in oersted (Oe) or ampere per meter units and is denoted as H_C, as shown in **Figure 6**. The area enclosed by the hysterisis curve indicates the amount of magnetic energy stored in the magnetic material. When the coercive field of a ferromagnet is large, the material is said to be a hard or a permanent magnet. A ferromagnet with a low coercive field is said to be a soft magnet. For example, in magnetic actuation, soft magnetic materials such as permalloy (80Ni–20Fe) or orthonol (50Ni–50Fe) are most frequently used for their preferable magnetic properties with high permeability, saturation, and low coercivity.

2.02.2.3 Actuation Mechanisms

As discussed in Section 2.02.3.2, a magnetic field can be used to produce force, torque, and displacement of microstructures or magnetic induction. One can conceptually understand magnetic actuation: materials, initial orientations, and magnetic field uniformities, as illustrated in **Figure 7** (Liu 2006). Two pieces of permanent magnets (pieces A and B) and two soft magnets (pieces C and D) are used as examples. Their initial orientations are different. Pieces A and C are oriented such that their internal magnetization is parallel to the applied magnetic field lines. Pieces B and D are intentionally misaligned.

Under a uniformly applied magnetic field, the soft magnets become polarized so that they are magnetized. When the internal magnetization is parallel to the applied magnetic field, no force or moment is generated (pieces A and C). However, if the internal magnetization is placed at a given angle with respect to the applied magnetic field, the magnetic pieces B and D will experience torques, but not a translational net force. However, a nonuniformly applied magnetic field produces the translational net force. The summary of the force and the torque for each piece under these three situations is provided in **Table 3**.

Note that the direction of the internal magnetization of pieces B and D is aligned along their longitudinal axes, rather than being parallel to the direction of the applied magnetic field. Easy axis refers to the energetically favorable direction of the spontaneous magnetization in a ferromagnetic material. This axis is determined by various factors, including the magnetocrystalline anisotropy and the shape anisotropy. For example, as shown in **Figure 5**, a thin cantilever beam of a ferromagnetic material can exhibit strong magnetization in the plane of the beam rather than vertical to the beam surface, even when the applied induction field $\mathbf{H}_0$ is primarily aligned to the normal direction of the beam surface. This phenomenon is a consequence of energy

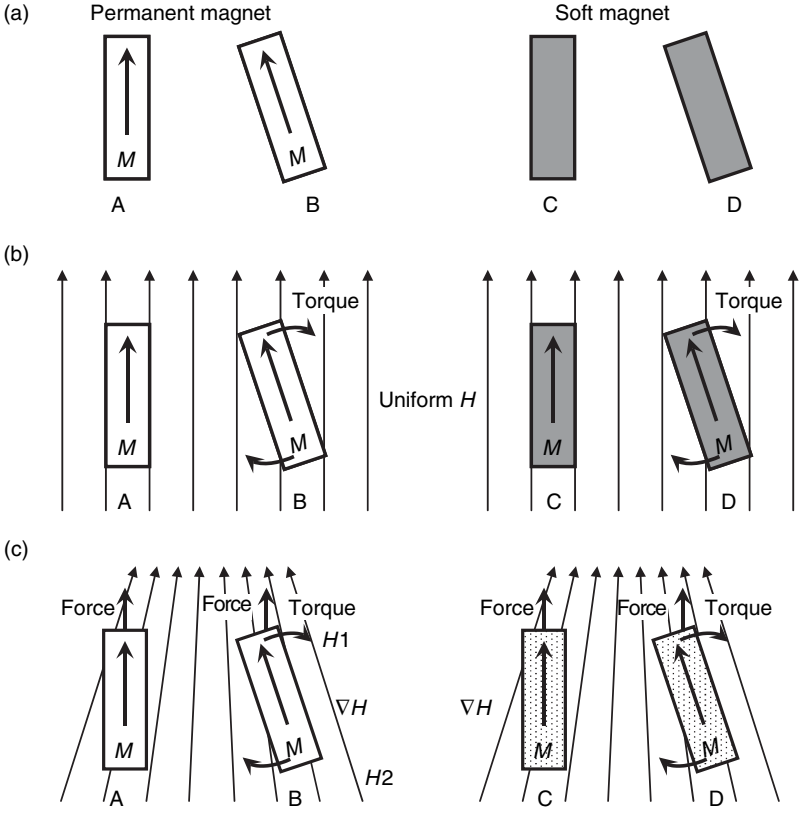


Figure 7 Magnetization states and force states in magnetic field lines. Net magnetic force and torque for each piece are indicated under each biasing case. (a) No applied magnetic field, (b) uniformly applied magnetic field, and (c) nonuniformly applied magnetic field. *H* is the local strength of the biasing magnetic field. ∇H is the difference of the magnetic driving field at two monopoles. (Source: Liu C 2006 *Foundations of MEMS*. Pearson Education, Upper Saddle River, NJ.)

Table 3 Presence of magnetic translational force and rotational torque

	No field (H = 0)				*Uniform field*				*Nonuniform field*			
Pieces	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*
Translation force	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Rotational torque	No	No	No	No	No	Yes	No	Yes	No	Yes	No	Yes

minimization to align magnetic domains. It simply takes less energy to magnetize along the longitudinal direction than that to magnetize to the normal direction of the in-plane direction of the magnetic pieces.

Magnetic actuators shown in **Figure 8** can be categorized according to the types of magnetic sources and the microstructures involved. The source of magnetic field can be an external solenoid, an integrated electromagnetic coil, or a permanent magnet. Multiple types of sources can be used in a hybrid manner. A permanent magnet easily generates a magnetic induction of 0.5–1 T, while the magnetic induction of a simple planar electromagnetic coil with about 10 windings is in the millitesla range (Gijs 2004). An external electromagnet solenoid with increased number of coil windings can generate better magnetic induction than the integrated electromagnetic coil. The microstructures in magnetic actuators, located on a chip, can be a permanent magnet (or hard ferromagnet), a soft ferromagnet (permalloy), or an integrated electromagnetic coil. The movable parts under magnetic source can be in the forms of a membrane, a beam, a spherical ball, magnetic particles, or a ferrofluid plug. The relation among these cases according to the classification scheme is summarized in **Table 4**. For example, in **Figure 8(a)**, a microstructured membrane is directly bonded to a core of external solenoid so that the membrane is movable upward or downward. In **Figure 8(h)**, a beam integrated with microcoils is placed between two permanent magnets, and rotates by Lorentz forces when current passes through the coils.

With a well-designed magnetic structure, magnetic forces can be controlled by changing coil currents in external solenoids or integrated electromagnets. By alternating the currents positively or negatively, the coil can also give different magnetic poles, north or south. In coil-free magnetic actuation, the magnetic forces are generated by permanent magnets, which are usually stationary. Sometimes, the permanent magnets are useful in handling moving balls, particles, or ferrofluids in microfluidic channels to continuously produce the magnetic fields and forces without any external energy. Most magnetic actuators require continuous power to keep the actuators on. This problem can be solved by bistable actuations that require power only in a transient mode between two stable positions. Unique bistable (Bohm *et al.* 2000, Capanu *et al.* 2000, Maekoba *et al.* 2001, Ren *et al.* 1997) and bidirectional (Bhansali *et al.* 2000, Cho and Ahn 2003) actuators can be realized by taking advantages of magnetic pole interactions between the permanent magnet and the electromagnets.

Table 4 Types of magnetic actuators and corresponding case from **Figure 8**

		Movable part				
Source of field	*On-chip microstructure*	*Membrane*	*Beam*	*Ball*	*Particles*	*Ferrofluid*
External solenoid	Permanent magnet					
	Soft ferromagnet	b, c	d			
	Integrated coil					
	Others	a		i, j		
Integrated coil	Permanent magnet		g			
	Soft ferromagnet	e, f				
	Integrated coil					
Permanent magnet	Permanent magnet					
	Soft ferromagnet					
	Integrated coil		h			
	Others				k	l

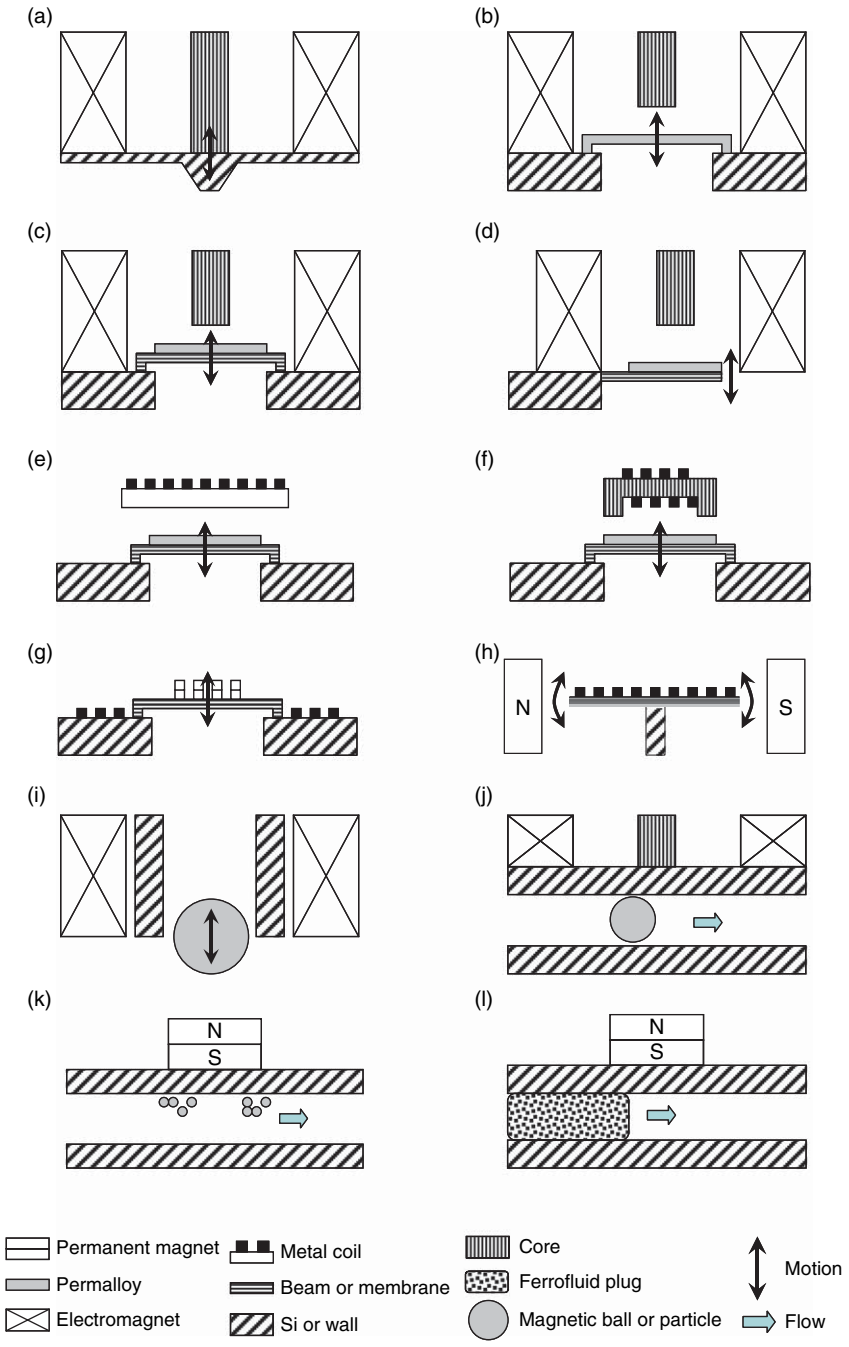


Figure 8 Typical magnetic actuators.

2.02.3 Materials and Fabrication

2.02.3.1 Deposition of Magnetic Materials

Magnetic materials exhibit many unique and useful properties apart from the high permeability of soft magnet materials and the high coercivity of permanent magnet materials. It is well known, for example, that soft magnetic materials enhance the magnetic induction B in the material over what would be achieved in free space by the same applied magnetic field H. In magnetic actuation, it is necessary to fabricate both the permanent magnets and the electromagnets with required properties.

Permanent magnet and electromechanical devices can be found in a wide variety of products (e.g., audio and video players, telecommunications equipment, personal computers, printers, copiers, automobiles, appliances, power tools, data storage devices, and biomedical apparatus) that have an impact on our daily lives (Furlani 2001). A good permanent magnet should produce a high magnetic field with a low mass, and should be stable against influences that would otherwise demagnetize it. In terms of a magnetic moment, miniature magnets fare much better than coils as their size is reduced. The desirable properties of such magnets are typically stated in terms of the remanence and coercivity of the magnet materials. **Table 5** contains some data about materials used as permanent magnets (Myers 1997). Both the coercivity and remanence are quoted in tesla, the basic unit for magnetic field *B*. Besides coercivity and remanence, a quality factor for permanent magnets is the quantity $(BH_0)_{max}$. A high value for this quantity implies that the required magnetic flux can be obtained with a lesser volume of the material, making the device lighter and more compact. The most common and the most powerful micromagnets currently used in magnetic actuators are individually micromachined from bulk Nd–Fe–B or Sm–Co magnets. However, this method is incompatible with full integration or batch fabrication (Yan *et al.* 2001).

Besides the form of bulk permanent magnets, the form of film permanent magnets is very useful in magnetic actuators (Chin 2000). Since biasing of magnetoresistive transducers by permanent magnets films was demonstrated for high-resolution recording head (Bajorek and Thompson 1975), various applications based on permanent magnet films have been developed. Some films are thin films up to a few hundred nanometers, while others are thick layers of a few micrometers, even to tenths of a millimeter. These films can be deposited by metallurgical processes (e.g., sintering, pressure bonding, injection molding, casting, extruding, and calendaring), vacuum processes (e.g., evaporation, sputtering, molecular beam epitaxy, chemical vapor deposition), and electroplating processes (Myung *et al.* 2003). Electroplating (Cho *et al.* 2000, Liakopoulos *et al.* 1996) and screen printing of bonded powders (Dutoit 1999, Lagorce *et al.* 1999) are very well adapted to microfabrication but the resulting magnetic properties are relatively poor compared to bulk permanent magnets. Other techniques such as sputtering (Jeong *et al.* 2000, Lileev *et al.* 2002), pulsed laser deposition (Hannemann *et al.* 2002), or low-pressure plasma spraying (Rieger *et al.* 2000) give excellent magnetic properties but either the thickness of the deposited layer is too thin or the process is too difficult to adapt to microtechnology and batch fabrication due to high deposition temperature, chemical pollution, slow deposition rate, small deposition surface. Various permanent magnet films for microelectromehanical systems (MEMS) applications have been reviewed by several authors (Chin 2000, Cugat *et al.* 2003, Gibbs *et al.* 2004, Myung *et al.* 2003, Niarchos 2003) and summarized in **Table 6** (Chin 2000).

Micromachined thick CoNiMnP permanent magnet arrays on a silicon substrate were fabricated by electroplating (Liakopoulos *et al.* 1996). The magnet arrays had 1500 micromagnets of 40 μm × 40 μm × 50 μm dimension in a cubic shape. They had high vertical coercivity of $0.6 - 1.0 \times 10^5$ A m^{-1}, a remanence of 0.2–0.3 T, a saturation magnetic induction of 1.2–1.3 T, and a maximum energy density of 14 kJ m^{-3}. Electroplated thick CoNiMnP-based permanent magnet arrays with controlled direction of magnetization and improved magnetic properties were reported by applying external magnetic fields during electroplating (Cho *et al.* 2000). The optimized processing conditions with external magnetic fields applied during electroplating have improved the coercivity and the retentivity of the magnets by more than 200% and 350%, respectively,

Table 5 The coercivity, remanence, and a quality factor $(BH_0)_{max}$ for permanent magnets

Material	*Coercivity (T)*	*Remanence (T)*	*$(B H_0)_{max}$ (kJ m^{-3})*
Alnico IV	0.07	0.6	10.3
$BaFe_{12}O_{19}$	0.36	0.36	25
Alcomax I	0.05	1.2	27.8
MnBi	0.37	0.48	44
Alnico V	0.07	1.35	55
$Ce(CuCo)_5$	0.45	0.7	92
$SmCo_5$	1.0	0.83	160
Sm_2Co_{17}	0.6	1.15	215

Source: Myers (1997).

Table 6 The comparison among candidate permanent magnet films

Films	B_r *(kG)*	H_C *(kOe)*	$(BH)_m$ *(MGOe)*	*Corrosion resistance*	*Ease of magnetization*	*Electroplating*
$SmCo_5$	8	23	18	Poor	Yes	No
$Sm_2(Co, Fe, Zr)_{17}$	9–11	7	21–28	Poor	No	No
$Nd_2Fe_{14}B$	7–12	8–15	20–40	Poor	Yes	No
$Nd(Fe,Ti)_{12}N_9$	12–15	8.7	46.3	Fair	Yes	No
PtCo/Ag		1–17		Excellent	No	Yes
Pt/Fe	10	10	40	Good	No	Yes
CoNiMnP	2–3	0.8–1.3	1.8	Good	Yes	Yes
CoCrTaX	7–9	2–4		Good	Yes	Yes
FeCrCo	8–12	0.7		Excellent	Yes	Yes
MnAl	4–5	3		Good	Yes	No
Resin + RPM	8–10	8.5	8–19	Fair	Fair	No
Resin + Ferr.	3	4	2–4	Excellent	Yes	No
Ba/Sr-ferrite	2–3	3–4	3–5	Excellent	Yes	No
γ-Fe_2O_3	3	2–4		Excellent	Yes	Maybe

Note: B_r, residual magnetic induction; H_C, coercivity; $(BH)_m$, energy density; 1 KG = 0.1 T; 1Oe = $1000/(4\pi)$ A m^{-1} = 79.6 A m^{-1}; 1 kOe = $7.96\times10^4\times$ A m^{-1}; 1 MGOe = 7.96 kJ m^{-3}.
Source: Chin (2000).

compared with those processed without external magnetic fields. As shown in **Figure 8(g)**, using integrated spiral coils around the electroplated magnets enables microactuators, microsensors, and other MEMS devices.

As discussed, electroplating is very well adapted to microfabrication for depositing ferromagnetic materials, as shown in **Figure 9**. The process used in electroplating is called electrodeposition, and it is analogous to a galvanic cell acting in reverse. The sample to be coated is placed into a container containing a solution of one or more metal salts. The sample is connected to an electrical circuit forming the cathode (negative) of the circuit, while an electrode typically of the same metal source to be plated forms the anode (positive). When an electric current is passed through the circuit, metal ions in the solution take up excess electrons from the sample, resulting in the deposition

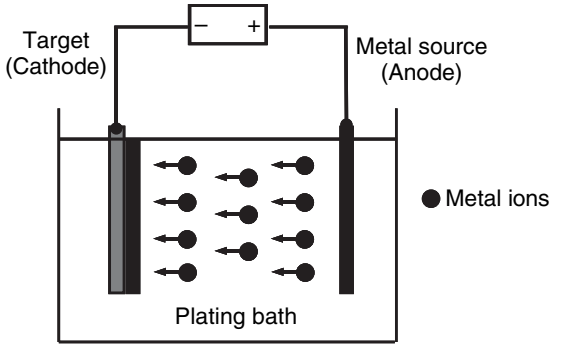


Figure 9 Typical electroplating bath. A target is connected to the cathode and a metal source is connected to the anode.

of a layer of metal on the sample. The electroplating process is often desirable over other vacuum-based thin-film deposition methods because it is relatively easy and cheap to reach appreciable thickness from a few micrometers up to hundreds of micrometers. The electroplating rate can be controlled by the applied current density; the greater the current density, the faster the electroplating.

Electroplating is a very useful process for producing metal structures (e.g., Au, Ag, Cu, Ni, Pt, hard and soft magnetic materials) above the target substrate in predefined patterns followed by photolithography, as shown in **Figure 10**. Since the wafer is not conductive, the surface of a wafer is first coated with a thin-film conductive metal layer called the seed layer. Common seed layer materials are copper, aluminum, or gold. Thin metal layers of Cr or Ti are often used to enhance adhesion between the seed layer metal and the substrate. A thick mold, typically using thick photoresists (PRs) (e.g., AZ4620, SU-8, Futurrex NR-9000), is deposited and patterned on top of the seed layer. The wafer is immersed in an electroplating solution, and magnetic materials grow in the open windows where the seed layer is exposed to the electroplating bath. Depending on the current density and the duration of electroplating, the thickness of the electroplated magnetic materials can be controlled. After removing the thick PR mold, the electroplated magnetic materials remain on the surface of the wafer, as shown in **Figure 11**.

Soft magnetic materials such as permalloy, with about 20% iron and 80% nickel content, or orthonol,

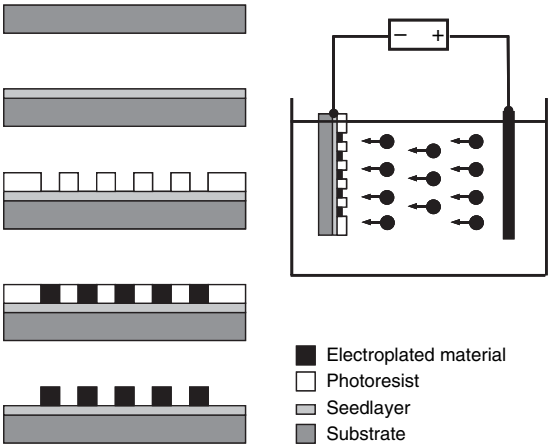


Figure 10 A process for the electroplating of magnetic materials and its typical electroplating bath.

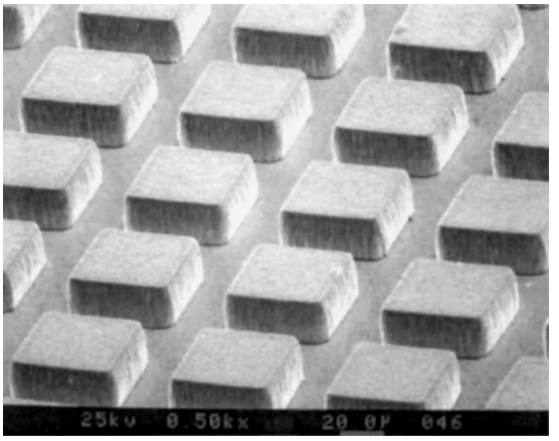


Figure 11 After removing the thick photoresist (PR) mold, the electroplated magnetic materials remain on the surface of the wafer. (Source: Cho H J, Bhansali S, Ahn C H 2000 Electroplated thick permanent magnet arrays with controlled direction of magnetization for MEMS application. *J. Appl. Phys.* **87**, 6340–2.)

with about 50% iron and 50% nickel content, are most frequently used for their preferable properties and availability of electroplating methods and baths. Permalloy is a commonly used soft magnetic material in MEMS because of its relatively high magnetic saturation ($B_S = \sim 1$ T), low coercivity (low hysteresis loss), near-zero magnetostriction (i.e., magnetic properties not affected by film stress), and significant anisotropic magnetoresistance (Myung *et al.* 2003, Niarchos 2003). Permalloy has been used for many years for the fabrication of magnetic heads in the recording industry and for magnetic microactuators. **Table 7** gives typical plating bath compositions for soft and hard magnetic materials, including permalloy, orthonol, and CoNiMnP (Cho *et al.* 2000, Liakopoulos *et al.* 1996, Park and Allen 1998).

2.02.3.2 Fabrication of Electromagnetic Coils

At the macroscale, conventional electromagnets usually have a solenoid shape, which can be manufactured by simply winding many turns of a conducting coil around a ferromagnetic core. However, miniaturization of integrated electromagnetic coils in magnetic microactuators is a very challenging task due to limited thickness and cross section. Although a hybrid approach can be found in the assembly of wound coils (Fischer and Guckel 1998, Guckel *et al.* 1991, 1992), for true integration, the planar fabrication methods are favorable. Three types of integrated electromagnetic coils, spiral type, solenoid type, and meander type, are possible, as illustrated in **Figures 12** and **13** (Ahn and Allen 1998).

The easiest form of integrated electromagnetic coils is a single-layer spiral-type planar coil with an air core. However, this coil cannot generate a strong magnetic flux due to the lack of a magnetic core and the lateral spreading of coils away from the center. It is difficult to guide the magnetic flux to the required point in this structure without using a magnetic core. By adding a magnetic core made of a high permeability soft magnetic material, a spiral-type planar coil with the core can perform better than the coil without the core. Furthermore, an arrangement of multilayer, spiral-type coils with a core to increase the number of turns can improve its performance.

To fabricate solenoid-type electromagnetic coils on a planar substrate, quasi 3D micromachining techniques are suggested (Ahn and Allen 1998). This solenoid- or bar-type planar coil with a closed magnetic circuit can be considered as one of the favorable structures as a magnetic flux generator in microscale, since the wires wrapped around a closed magnetic core result in low leakage flux, and the generated flux can be flexibly guided to the required points.

Brief fabrication steps for a solenoid-type electromagnetic coil are shown in **Figure 14**, which has been applied for the magnetic actuation of microvalves (Sadler *et al.* 1999). In this integrated coil, the process involves first etching and electroplating magnetic vias or through-holes, and then building the electromagnetic coil on top of the through-hole

Table 7 Electroplating bath compositions for soft and hard magnetic materials

Magnetic material	*Electroplating solution*	*Amount*
Permalloy: Ni(80%)–Fe(20%)	$NiSO_4 \bullet 6H_2O$ (g l^{-1})	200
	$FeSO_4 \bullet 7H_2O$ (g l^{-1})	8
	$NiCl_2 \bullet 6H_2O$ (g l^{-1})	5
	H_3BO_3 (g l^{-1})	25
	Saccharin (g l^{-1})	3
	PH	2.5–3.0
	Temperature (°C)	25–30
	Current density (mA cm^{-2})	13
Orthonol: Ni(50%)–Fe(50%)	$NiSO_4 \bullet 6H_2O$ (g l^{-1})	168
	$FeSO_4 \bullet 7H_2O$ (g l^{-1})	81
	$NiCl_2 \bullet 6H_2O$ (g l^{-1})	135
	H_3BO_3 (g l^{-1})	50
	Saccharin (g l^{-1})	3
	pH	3.5–4.0
	Temperature (°C)	55–60
	Current density (mA cm^{-2})	30
Permant magnet: CoNiMnP	$CoCl \bullet 6H_2O$ (g l^{-1})	24
	$NaCl \bullet 6H_2O$ (g l^{-1})	24
	$MnSO_4 \bullet H_2O$ (g l^{-1})	3.4
	NaCl (g l^{-1})	23.4
	$B(OH)_3$ (g l^{-1})	24
	$NaH_2PO_2 \bullet xH_2O$ (g l^{-1})	4.4
	Sodium lauryl sulfate (g l^{-1})	0.2
	Schchain (g l^{-1})	1.0
	Current density (mA cm^{-2})	10–20

Sources: Cho *et al.* (2000), Liakopulos *et al.* (1996), and Park and Allen (1998).

wafer using a multilayer thick PR process. Through-holes are patterned onto the backside of a 250-μm-thick silicon wafer using standard photolithography and are then etched in a 45% KOH solution. After etching is completed, the empty through-holes are electroplated with NiFe permalloy using a backside Cr/Cu/Cr metal layer as an electroplating seed layer. The through-holes are plated until they reach the top of the wafer, thus forming the magnetic vias, which allow magnetic flux to pass through the silicon wafer. After the magnetic vias are complete, the solenoid-type coil is ready to be fabricated on the topside of the wafer. Fabrication of coil on top of the through-hole wafer is based on a thick PR process. AZ-4000 series PR is used to produce a 25-μm-thick electroplating mold. Copper lines are then electroplated on top of the mold from an underlying Ti/Cu seed using standard electroplating techniques. After removing the PR and the seed layer, a new layer of PR is spun, via openings are patterned, and the PR is hard cured at 220°C to form a permanent and planarizing dielectric layer. Another seed layer is then deposited from which first vias and then a magnetic core are patterned and electroplated. Conductor vias are electroplated copper with a cross section of 50 μm × 50 μm and a thickness of 25 μm. The magnetic core is formed from electroplated NiFe permalloy and is also 25 μm thick. Then photoresist and seed layers are removed, and a layer of photoresist is spun and hard cured to form another permanent and planarizing dielectric layer. Finally, top conductor lines are patterned and grown from another seed layer using exactly the same method as for the bottom conductors. Again, the conductor thickness is 25 μm.

An integrated meander-type coil can be realized, where meander refers to the wrapping approach, by interchanging the roles of the conductor wire and the magnetic core in the solenoid-type coil. In the meander-type electromagnetic coil, a magnetic core is wrapped around a conductor line, i.e., a multilevel magnetic core is wrapped around a planar conductor. This structure has the advantage that a relatively short, planar conductor is used, thus reducing the total conductor resistance. Using this coil structure, a planar magnetic micromotor and a micropump with a fully integrated stator and wrapped coils were realized (Ahn and Allen 1995, Ahn *et al.* 1993).

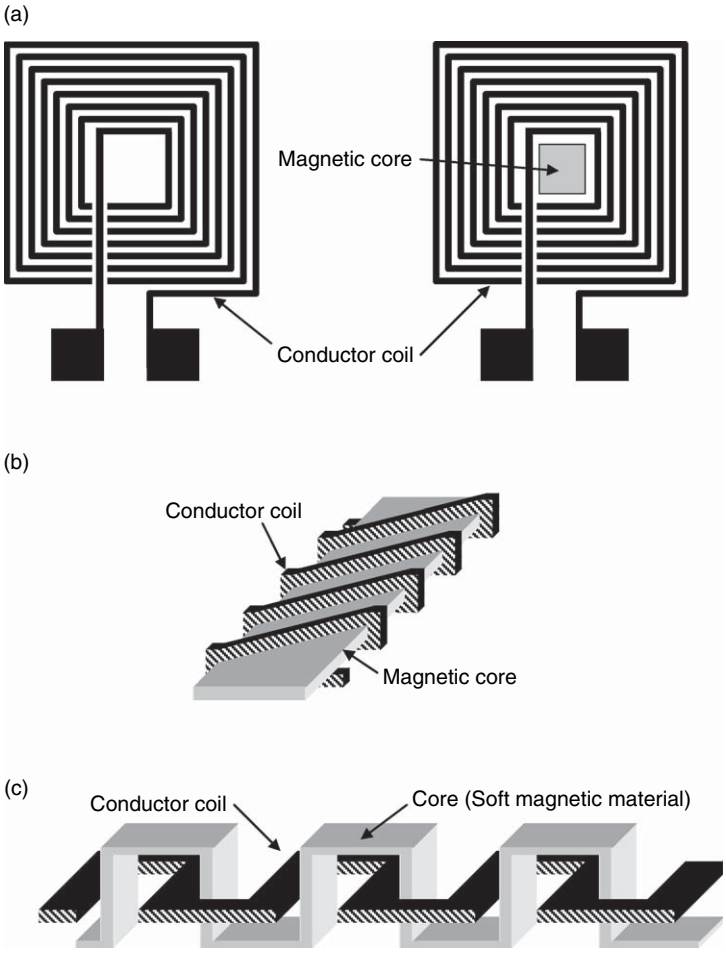


Figure 12 Schematic diagrams of three types of integrated electromagnetic coils: (a) spiral type without a core and with a core, (b) solenoid type, and (c) meander type.

2.02.3.3 Magnetic Particles

In addition to the thin films or mechanical structures of hard and soft magnetic materials, mobile magnetic particles, ranging in size from a few nanometers to many micrometers, offer tremendous applications in microfluidic-based magnetic actuation (Gijs 2004, Pamme 2006). Due to the presence of embedded magnetic entities, they can be magnetically manipulated by magnetic fields, in parallel with normal microfluidic or biological processes. In magnetic actuation, nonuniform magnetic field or magnetic field gradient is frequently used to transport and manipulate these magnetic particles.

The majority of magnetic particles are superparamagnetic. Superparamagnetic particles are magnetized in a magnetic field since the particles have a core of small iron oxide crystals encased by a polymer shell. Iron oxides, such as magnetite (Fe_3O_4) or maghemite (γ-Fe_2O3) are more stable against oxidation, and are preferentially used as core material instead of iron. Biomolecules such as DNA strands, antigens, or antibodies can easily be attached to their surface due to their engineered polymer shell surface for selectively binding the biomolecules of interest. However, since they have no magnetic memory, once the external field is removed, the particles redisperse and behave like a nonmagnetic material. Once suspended magnetic particles are selectively removed from a solution using a magnetic field, they remain suspended even after the removal of the field. Therefore, it is very easy to manipulate the magnetic particles in microfluidic-based magnetic actuation. Another advantage of using small particles is a large surface-to-volume ratio for chemical binding.

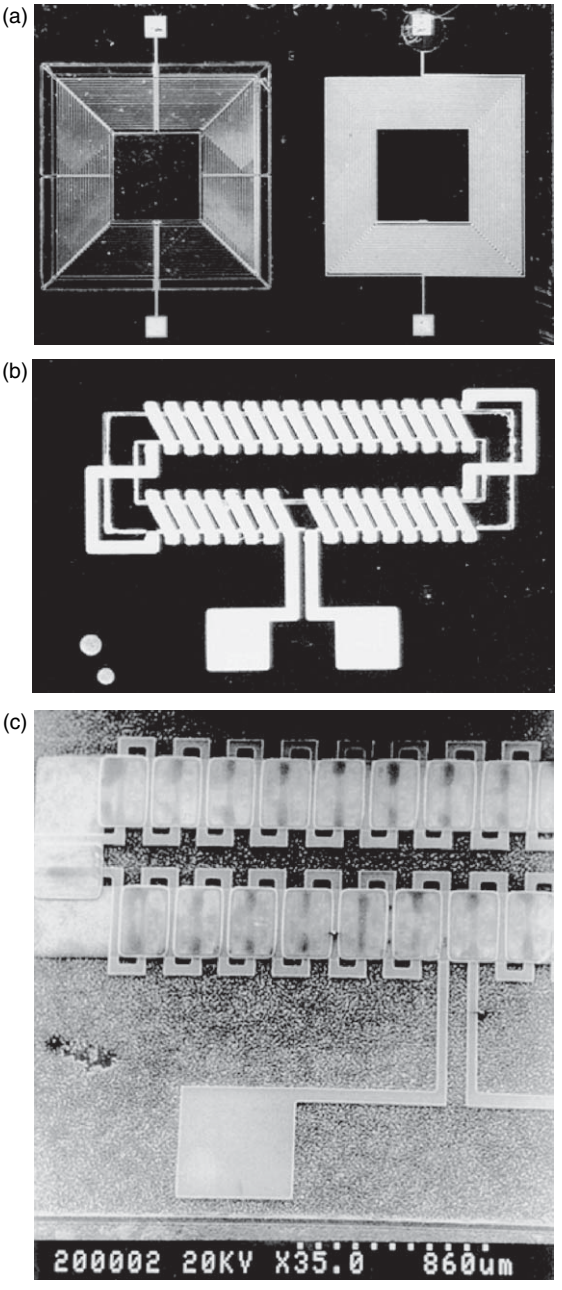


Figure 13 Fabricated examples of three types of integrated electromagnetic coils: (a) spiral type without a core and with a core, (b) solenoid type, and (c) meander type. (Source: Ahn C H, Allen M G 1998 Micromachined planar inductors on silicon wafers for MEMS applications. *IEEE Trans. Ind. Electron.* **45**, 866–76.)

Magnetic fluids or ferrofluids are another class of superparamagnetic material that can be used in microfluidic devices. Developed by NASA in the 1960s, ferrofluids are used today in many applications such as loudspeakers, CD-ROMs, computers,

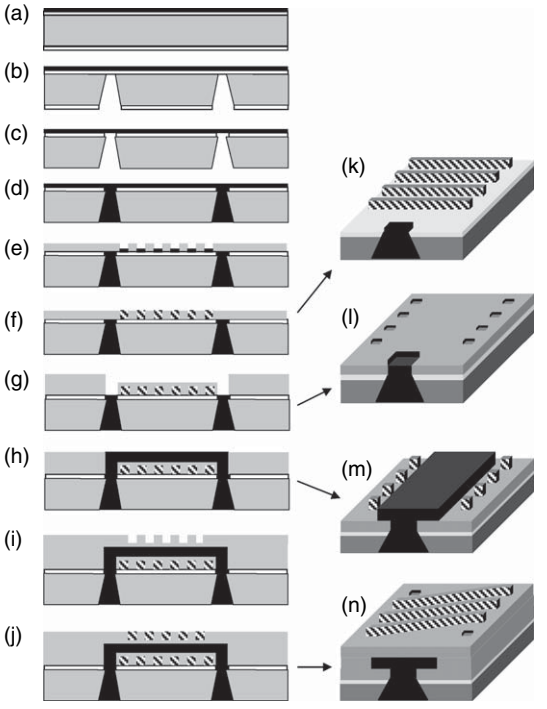


Figure 14 Bar-type magnetic actuator fabrication steps: (a) silicon wafer with Cr/Cu seed layer; (b) through-hole etch in KOH; (c) oxide removal; (d) NiFe through-hole plating; (e) thick photoresist (PR) patterning for copper conductor lines; (f, k) electroplating of thick copper conductor lines; (g, l) via opening and PR hard curing; (h, m) electroplating of vias and magnetic core; (i) thick PR patterning for copper conductor lines; and (j, n) via opening on second PR layer, hard curing, and top conductor line electroplating.

and semiconductor fabrication (Rosensweig 1985). The ferrofluid is a stable colloidal suspension of subdomain magnetic particles in a liquid carrier such as water or an organic solvent. The particles, which have an average size of about 10 nm, are coated with a stabilizing dispersing surfactant, which prevents particle agglomeration even when a strong magnetic field gradient is applied to the ferrofluid. Therefore, in microfluidic channels the ferrofluid is a good plug that moves easily by external magnetic forces.

2.02.4 Applications

2.02.4.1 Microfludics: Valves, Pumps, and Mixers

In microfluidic systems, magnetic actuators such as valves, pumps, and mixers have been realized for many years. The research on microfluidics, including

microvalves (Hesketh *et al.* 2004, Oh and Ahn 2006), micropumps (Laser and Santiago 2004, Woias 2005), micromixers (Nguyen and Wu 2005), and world-to-chip microfluidic interfaces (Fredrickson and Fan 2004) has been thoroughly reviewed. In this section, selected examples of magnetically driven microvalves, micropumps, and micromixers are covered.

2.02.4.1.1 Microvalves

Solenoid-shaped external electromagnets are routinely used in magnetic actuation, and can be simply attached to the microvalve structures. In 1979, a miniaturized electromagnetic microvalve was demonstrated using a solenoid plunger, which was physically connected to a silicon micromachined membrane (Terry *et al.* 1979). The microvalve was the first active micromachined valve, and a component of an integrated gas chromatography system, as shown in **Figure 15** and also indicated in **Figure 8(a)**. A microvalve with a thin permalloy membrane as a cap inside a small tube and a solenoid coil mounted outside was built by Yanagisawa *et al.* (1993). Various types of ball-type microvalves making use of solenoid-shaped electromagnets were reported (Fu *et al.* 2003, Krusemark *et al.* 1998, Oh *et al.* 2001). Although some pinch valves are commercially available (from Bio-Chem Valve, Boonton, NJ), a miniaturized pinch-type valve using a stand-alone conventional solenoid compliant with a surface-mountable scheme on an integrated microfluidic biochemical detection system for magnetic bead-based immunoassay has been reported (Choi *et al.* 2002, Oh *et al.* 2005a, b).

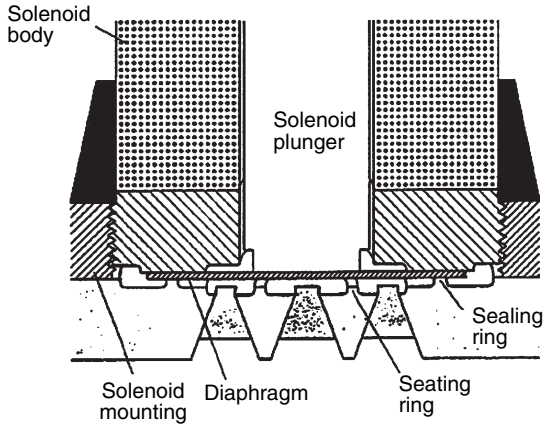


Figure 15 The microvalve with a solenoid plunger as a component of an integrated gas chromatography system. (Source: Terry S C, Jerman J H, Angell J B 1979 A gas chromatographic air analyzer fabricated on a silicon wafer. *IEEE Trans. Electron. Devices* **26**, 1880–6.)

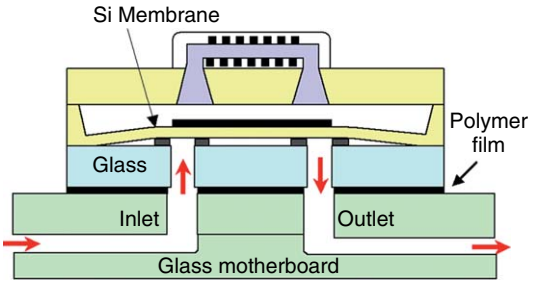


Figure 16 The normally closed magnetic microvalve with an integrated electromagnet. (Source: Oh K W, Han A, Bhansali S, Ahn C H 2002 A low-temperature bonding technique using spin-on fluorocarbon polymers to assemble microsystems. *J. Micromech. Microeng.* **12**, 187–91.)

Active microvalves with a fully integrated solenoid-type electromagnet (see **Figure 14**), a deflectable silicon membrane with a permalloy thin film, and a stationary inlet/outlet valve seat were developed (Cho *et al.* 2001, Choi *et al.* 2001a, b, Oh *et al.* 2002, Sadler *et al.* 1999). A schematic representation of the microvalve mounted on a glass motherboard is shown in **Figure 16**. The electromagnet, valve components, and the glass motherboard were fabricated individually and then bonded together using a low-temperature polymer bonding technique to fabricate the microvalve (Oh *et al.* 2002). The electromagnet acted as a flux generator and produced sufficient forces to pull the silicon membrane by magnetically coupling with the NiFe permalloy electroplated on the silicon membrane. The achievable flow rate increased quickly after an applied current of about 250 mA for N_2 gas.

Movable membranes hybrid integrated with permanent magnets to increase magnetic forces with low power consumption are possible in two ways: integrated with coils (in this case, the permanent magnet is fixed) (Bosh *et al.* 1993, Goettsche *et al.* 2005, Meckes *et al.* 1999) or mounted with permanent magnets (the permanent magnet is free to move by actuating external coils) (Bae *et al.* 2002, 2003, Jerman 1994). Most microvalves require continuous power to keep the valves on or off. This problem can be solved by bistable actuations that require power only in a transient mode between two stable positions. Unique bistable microvales can be realized by taking advantages of magnetic pole interactions between the permanent magnet and the electromagnets (Bohm *et al.* 2000, Capanu *et al.* 2000, Maekoba *et al.* 2001,

Ren and Gerhard 1997) or by combining electromagnetic and electrostatic actuation (Bosh *et al.* 1993).

In addition, a novel phase change bistable microvalve using a paraffin-based ferrofluid plug was reported, as shown in **Figure 17** (Oh *et al.* 2005a, b). The material (called Ferro-Wax) was formed by mixing ferrofluids and paraffin waxes in a volume ratio of 1:2. The Ferro-Wax plug was essentially leak-proof because of the phase change nature of the material; once the plug is solidified, it forms a solid seal. A top glass substrate with holes (φ 500 μm) and microchannels (a width of 300 μm and a depth of 150 μm) was anodically bonded with a wet etched silicon wafer using a Ti/Pt heater/sensor. For injection of the paraffin-based ferrofluid plug, the chip (20 mm × 20 mm × 0.8 mm) was heated to 80°C on a hot plate and the plug of 0.2 μl was loaded into the stem hole. The meltable magnetic plug changed the phase from solid to liquid by an on-chip heating and moved remotely in a channel by magnetic actuation without pneumatic air/vacuum systems. The bistable microvalve performed excellently up to 50 psi. The use of only ferrofluids as microvalves (Hartshorne *et al.* 2004) and micropumps (Hatch *et al.* 2001, Yamahata *et al.* 2005) has also been suggested.

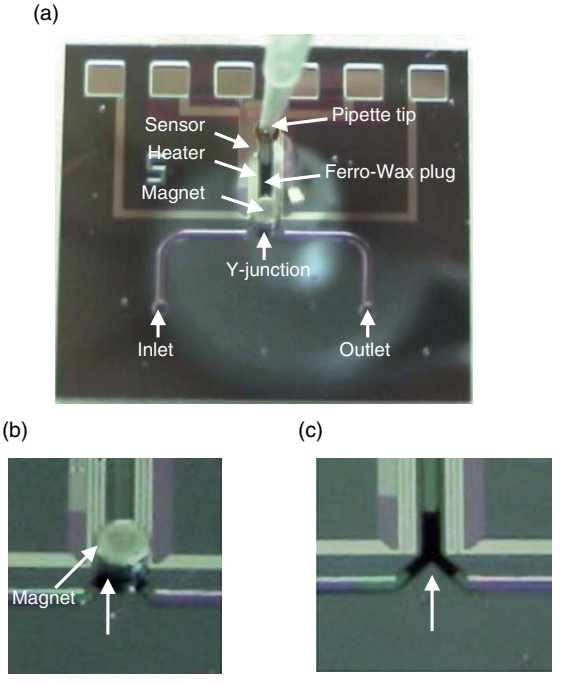


Figure 17 (a) The chip was heated to 80°C on the hot plate and meltable magnetic plug of 0.2 μl was loaded into the vent hole, (b) the molten plug was stopped at the Y-junction in the presence of the permanent magnet, and (c) the plug was solidified after cooling to 60°C.

2.02.4.1.2 Micropumps

A planar magnetic micromotor and a micropump with a fully integrated stator and wrapped coils were realized (Ahn and Allen 1995, Ahn *et al.* 1993). A meander-type integrated inductive component was used in the motor for flux generation (see **Figure 12(c)**). As shown in **Figure 18**, the stators were made of the meander-type integrated electromagnets, whereas the rotor was made of a permalloy material. The pump actuator was based on the variable reluctance magnetic micromotor with fully integrated stator and coils, which had 12 stator poles in three phases and 10 rotor poles. When phase I coils were actuated, the rotor poles located closest to the excited stator poles were attracted to the stator pole. The actuated phase coils were turned off, and the next phase II coils were actuated for continuous rotation. In this design, the wound poles of all phases were arranged in pairs of opposite polarity to achieve adjacent pole paths of short lengths. The stator and coils arranged in one or more sets and phases were actuated in sequence to rotate the rotor, therefore a rotary magnetic micropump can be demonstrated.

When three microvalves are serially connected and controlled sequentially, they can work as a peristaltic micropump. Zhang and Ahn (1996) built a peristaltic micropump using fully integrated three solenoid-type electromagnets on a silicon wafer. Pan *et al.* (2005) presented a polydimethylsiloxane

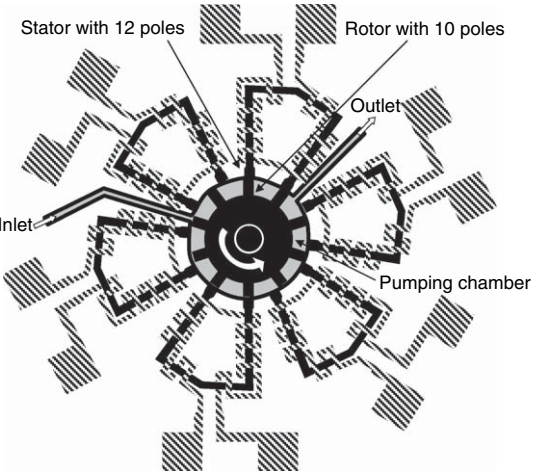


Figure 18 A planar magnetic micromotor and a micropump with a fully integrated stator and wrapped coils.

(PDMS) peristaltic micropump encapsulated with three small permanent magnets in three small chambers. A small DC motor with three permanent magnets stagger-mounted on its shaft was used to pull down and actuate the membrane-mounted magnets to generate a peristaltic waveform. Oh *et al.* (2005a, b) developed a pinch-type peristaltic pump by connecting three pinch-type valves in series, showing self-priming and bidirectional pumping capabilities.

Magnetic micropumps can be easily built by incorporating two check valves in an inlet and an outlet with a pumping chamber where the magnetic actuator produces a stroke volume. McDonald *et al.* (2004) used a PDMS membrane, driven by magnetic force of two miniature permanent magnets, with two one-way microball valves in the inlet and the outlet. A small DC motor with two permanent magnetic disks (NdFeB) mounted on its shaft is used to actuate the membrane-mounted magnet. Santra *et al.* (2002) fabricated a magnetic micropump in a modular fashion using a thin valve membrane sandwiched between two molded bodies and a magnetically oriented permanent magnet composite pumping membrane on top. Bohm *et al.* (1999) developed an electromagnetic actuator consisting of a permanent magnet placed in a coil and secondly a disk. By sandwiching a thin plastic membrane between two molded parts, planar passive membrane valves with low opening pressure resulted.

Magnetic pumping using ferrofluids has also been implemented without using mechanical moving parts in the microstructure. A micropump consisting of two serial check valves converted the periodic motion of a ferrofluidic plug into a pulsed quasi-continuous flow (Yamahata *et al.* 2005). The ferrofluid was actuated by the mechanical motion of an external NdFeB permanent magnet. A planar rotary micropump using ferrofluids in a circular microchannel with an inlet and an outlet was designed (Hatch *et al.* 2001). Two permanent magnets were used, one as a valve actuator and another as a plug actuator: a fixed magnet to create a stationary ferrofluid plug between the inlet and the outlet and a rotational magnet to create a moving plug. By separating the plug into one stationary and one moving plug, the fluid was primed into the inlet and pushed out of the outlet. However, in this pumping method there is a risk of cross-contamination of the fluid with the ferrofluid plugs due to the direct contact. Using an indirect contact method between the fluid and the ferrofluid plugs, a peristaltic rotary pump was miniaturized (Kim *et al.* 2006). Rotating rollers were replaced with rotating ferrofluid plugs that gently squeeze a silicon rubber-based microchannel and that push the fluid from the inlet to the outlet.

Several magnetohydrodynamic (MHD) micropumps have been reported where current-carrying ions in aqueous solutions are subjected to a magnetic field to impart a Lorentz force (eqn [7]) on the liquid and to induce flow. The flow direction depends on the direction of the magnetic field and the direction of the applied electrical field, as shown in **Figure 2**. Because the Lorentz force acts on the bulk fluid and creates a pressure gradient, MHD pumps generate a parabolic velocity profile, similar to pressure-driven flow. Jang and Lee (2000) reported a MHD micropump with a 40-mm-long pumping channel with hydraulic diameter on the order of 1 mm. To avoid electrolysis associated with DC operation, Lemoff and Lee (2000) used a miniaturized electromagnetic coil operating (along with the electric field) at 1 kHz. Several papers have discussed microscale applications of MHD effects (Bau *et al.* 2001, Huang *et al.* 2000, Leventis and Gao 2001, Zhong *et al.* 2002).

2.02.4.1.3 Micromixers

Due to the laminar flow effect at the microscale, mixing multiple streams of fluids with short mixing length for LOC applications has been a difficult task. Various methods for mixing fluids in microscale channels have been demonstrated (Nguyen and Wu 2005). If one can stir fluids in the microchannels, rapid mixing will be produced at the stirrer site, which can be easily done by magnetic actuation. When a rotating external magnetic field is presented, a magnetic microstirrer will experience a torque through interaction between its internal magnetization and the external field. A simple magnetic stir bar structure was integrated directly into a channel (Lu *et al.* 2002, Ryu *et al.* 2004). The permalloy stir bar of 400 μm width can be fitted into a microchamber of 420 μm diameter, as shown in **Figure 19**. The stir bar was activated using the rotating magnet in a commercial hotplate/stirrer. Small magnetic nickel bars controlled by three external electromagnets were also suggested for mixing in microchannels (Barbic *et al.* 2001).

A chain of magnetic microparticles bonded together with linker molecules has been used to mix fluids in a microchannel (Biswal and Gast 2004). A rotating magnetic field generated by two pairs of electromagnets, mixed fluids quickly upon rotation of the chains. A dye and two buffer streams all

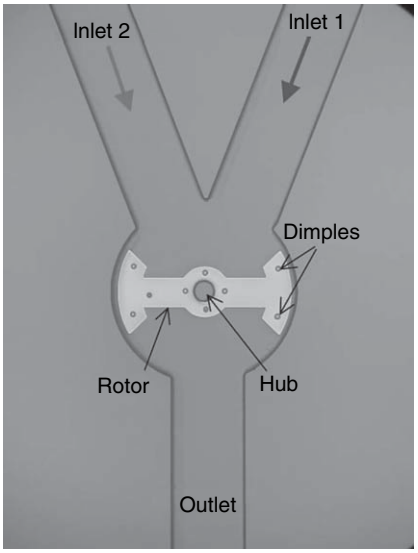


Figure 19 Magnetic mixing with a permalloy rotor controlled by a conventional benchtop stirrer plate. (Source: Ryu K S, Shaikh K, Goluch E, Fan Z F, Liu C 2004 Micro magnetic stir-bar mixer integrated with parylene microfluidic channels. *Lab Chip* **4**, 608–13.)

containing particle chains were merged and mixed. Also, an acid and a base were mixed in the presence of a fluorescent pH indicator. Magnetic particles moving at random were also suggested for mixing of two fluid streams (Rida and Gijs 2004). A plug of magnetic particles was formed in a microchannel by means of an electromagnet, and the particles were moved over short distances due to the ever-changing magnetic field, which was modified by an arbitrary waveform generator. Chaotic mixing was reported in a serpentine microchannel with copper wire electromagnets, embedded transversely to the direction of flow to attract nearby magnetic particles (Suzuki *et al.* 2004). The combination of time-varying magnetic fields and the serpentine geometry resulted in efficient mixing.

MHD flow, described above as a pumping method, can also be utilized for mixing fluids by integrating several MHD pumps into a microfluidic device and by utilizing time-varying flow patterns in order to achieve mixing. Bau *et al.* (2003) reported a circulatory fluid motion to mix fluids in a microchamber with several electrodes along the walls and additional electrodes at the bottom with a large Nd magnet. Time and space variations in the electric field led to chaotic motion and thus mixing, as visualized using a dye. West *et al.* (2003) used a circulating fluid in a circular channel with a MHD micropump for fluid circulation. Two solutions were injected into the channel and the MHD parabolic flow profile led to a large increase in interfacial surface area and thus to mixing.

2.02.4.2 Microoptics: Mirrors

A digital micromirror device has been fabricated with great success by Texas Instruments using electrostatic actuation for large-scale projection (http://www.dlp.com). An alternative good candidate to replace the electrostatic actuation in micromirrors is the magnetic actuation due to its large angular deflection. In addition, arc failure, which is a problem in the electrostatic actuation, can be avoided due to its current-driven actuation. However, the magnetic micromirrors may not be suitable for high-frequency modulation of more than a few kilohertz when compared to the electrostatic one.

One successful example of magnetic micromirrors was an electromagnetic optical scanner for horizontal scanning of a commercial confocal laser scanning microscope by Olympus (Miyajima *et al.* 2003). Magnetic actuation, suitable for moving large structure, was selected because of its millimeter-sized mirror, and single-crystal silicon hinge was used for realizing high-speed scanning (~4 kHz) with sufficient scan angle (2.1–16°). In order to maintain a mirror flatness for high-quality optical requirement, the whole wafer thickness (300 μm) was used as the mirror, resulting in a large moment of inertia. Lorenz force between a planar spiral-type coil on the mirror body and a large magnetic field produced by a bulk permanent magnet was used to actuate the mirror, as shown in **Figure 20** (see also **Figure 8(h)**). A pair of permanent magnets were magnetized in the x-direction, causing the magnetic field mainly in the x-direction, and parts of the actuation coil adjacent to the magnets had currents in the $\pm y$-direction (parallel to the axis of rotation). Therefore, Lorentz force was generated in the $\pm z$-direction, causing torsional actuation. It has successfully satisfied all the specifications including not only the fundamentals such as resonant frequency and scan angle but also the ones for the commercial product such as scanning stability and reliability.

An alternative approach using a hard magnetic CoNiMnP thick film as a permanent magnet component and an external solenoid electromagnet as a variable field source was reported by Cho and Ahn (2002, 2003). **Figure 21** shows a schematic view of a serpentine torsion beam-type scanner with bidirectional scanning capability and **Figure 22** shows its

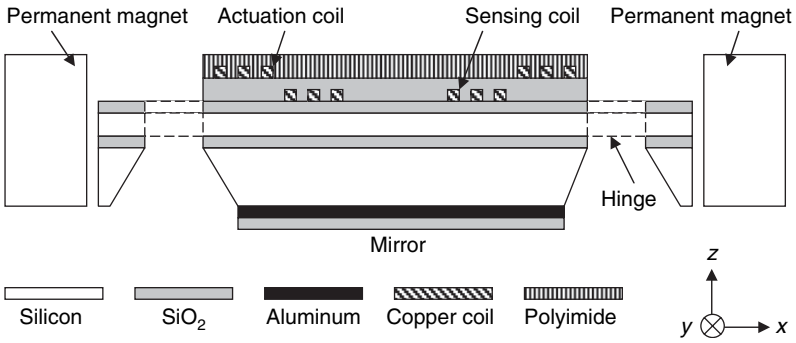


Figure 20 A cross-sectional view of an electromagnetic optical scanner for horizontal scanning of a commercial confocal laser scanning microscope by Olympus.

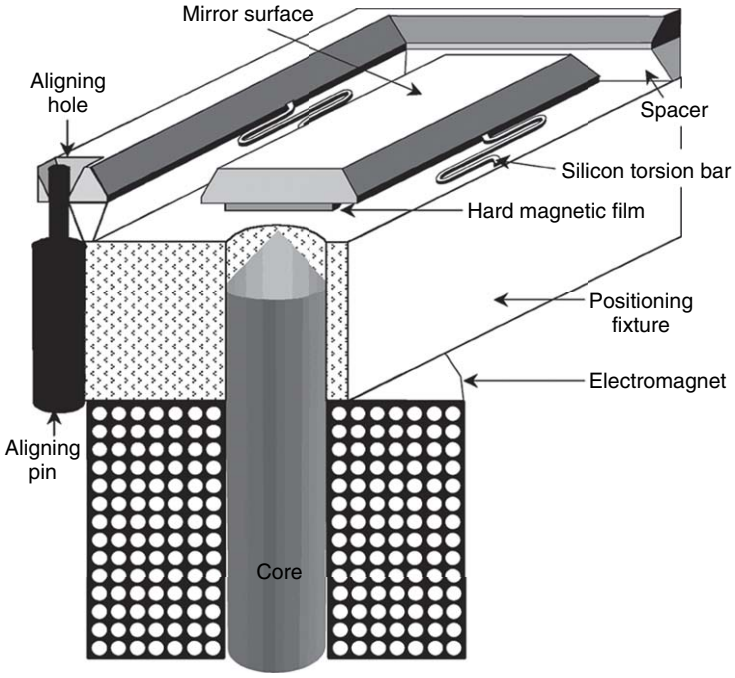


Figure 21 Schematic view of the bidirectional magnetic scanning mirror. (Source: Cho H J, Ahn C H 2003 Magnetically-driven bi-directional optical microscanner. *J. Micromech. Microeng.* **13**, 383–9.)

bidirectional scanning motion. The scanner can be operated bidirectionally with a large deflection of $\pm 70\,\mu m$ with actuation currents of $\pm 60\,mA$ under the condition of a static operation. Under dynamic operation, the prototype scanner showed stable bidirectional scanning performance at the operating frequency of 30 Hz, corresponding to 60 Hz in the regular unidirectional scanner. The average power consumption was 63 mW for AC operation, which is considerably lower when compared to the conventional scanner. At a scanning frequency of 30 Hz, the product of the scan angle and the mirror size $\theta \times D$ (deg $\times$ mm) was 5.0, which is large enough for practical applications. This unique function, moving a large mirror bidirectionally with a large deflection, cannot be easily obtained by using any other actuation mechanism.

Different types of magnetically driven optical micromirrors, scanners, and optical switches have also been studied: a deformable mirror for application to adaptive optics in astrophysics (Divoux *et al.* 1998), an out-of-plane micromirror array for hologram

Figure 22 Bidirectional scanning motion of the magnetic scanning mirror: (*top*) pull; (*middle*) rest; and (*bottom*) push. (Source: Cho H J, Ahn C H 2003 Magnetically-driven bi-directional optical microscanner. *J. Micromech. Microeng.* **13**, 383–9.)

memory devices (Jang and Kim 2003), mirror arrays for 3D optical switching application (Bernstein *et al.* 2004), a magnetoresistive 2D optical scanner (Bourouina *et al.* 2002, Debray *et al.* 2004), a galvanometric scanner (Ferreira and Moehlecke 1999), an electromagnetic 2×2 optical switch (Ji *et al.* 2004), and a double-focus zoom lens by a magnetic mirror for a thin camera (Kinoshita *et al.* 2006).

In addition, a magnetic fluid-based optical switch was designed and characterized (Horng *et al.* 2004). The optical switch was formed by sealing a magnetic fluid between two glass prisms. When light was incident on one side of one of the prisms, the reflected light from the magnetic fluid film came out from the same prism, whereas the transmitted light through the film was emitted from the other prism. It was found that the intensity ratio of the reflected light to the transmitted light can be manipulated by varying the external magnetic field strength.

2.02.4.3 Microelectronics: Switches

Microelectromechanical systems (MEMS) switches or relays have been miniaturized for applications in the areas of automotive control, power electronics, and communications. Most of the MEMS switches (or relays) reported to date have used electrostatic actuation (Gretillat *et al.* 1999, Kruglick and Pister 1999, Receveur *et al.* 2005, Schiele *et al.* 1998). To overcome the disadvantages of high actuation voltage and low mechanical stability in electrostatic actuation, a fully integrated magnetic switch was suggested by Taylor *et al.* (1998). The device with a single-layer coil to actuate a movable upper magnetically responsive platform can make and break 1.2 A of current through the relay contacts when the relay was electromagnetically switched. However, the actuation required continuous power during a given switching period.

For on–off switching operations, bistable latching is preferred, which can eliminate or reduce power consumption, heat generation, and thermal distortion (Cho *et al.* 2005, Cugat *et al.* 2006, Ruan *et al.* 2001a, b). Ruan *et al.* (2001a, b) reported a magnetic switch with the bistable latching operation. The structure and the principle of the bistable magnetic switch is illustrated in **Figure 23** (see **Figure 5**). A cantilever beam

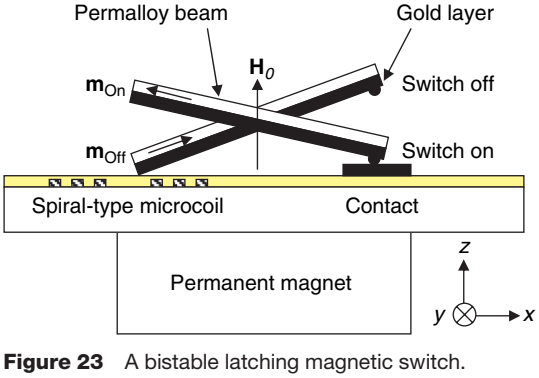


Figure 23 A bistable latching magnetic switch.

consisting of permalloy film on the top and a layer of gold on the bottom was elevated above the substrate by a torsion support. When the beam was placed on a uniform magnetic field $\mathbf{H}_0$ in the z-direction by a permanent magnet, a magnetic torque $\boldsymbol{\tau}_{\text{magnetic}} = \mu_0 \mathbf{m} \times \mathbf{H}_0$ (see eqn [20]) was exerted in either the clockwise or the counterclockwise direction, depending on the initial orientation of the beam with respect to the magnetic field. The bottom permanent magnet was used to hold the beam in one state until the next switching took place. A spiral-type planar microcoil beneath the beam generated switching magnetic field lines pointing in both the x- and z-directions. The magnetic field in the x-directional component was mainly responsible for the switching. When a switching field was a sufficient magnetic field greater than a critical field to toggle the orientation of the $\mathbf{m}$, the induced magnetic field caused a magnetization toggling between $\mathbf{m}_{\text{on}}$ and $\mathbf{m}_{\text{off}}$, resulting in rotation of the beam. By reversing the current in the coil, the original position was restored.

By combining electromagnetic and electrostatic actuation with a latching mechanism, two types of radio frequency (RF) MEMS switches (series type and shunt type) have been demonstrated by Cho *et al.* (2005). The switch used electromagnetic force only during switching transition periods and for the rest of the holding period it used electrostatic force to maintain low power consumption. When a spiral-type coil was excited counterclockwise, a membrane with the coil was actuated downward by Lorentz force in an external magnetic field by a permanent magnet and the contact metal made an electrical connection to the underlying coplanar waveguide signal line. Then, electrostatic force between the coils and the bottom electrode was activated to hold the membrane for maintaining its on state. For switching its state back to the off state, a reverse current was applied and then the contact metal was detached from the bottom signal line by Lorentz force as well as by the mechanical restoring force.

2.02.4.4 LOC: Magnetic Manipulation and Separation

Magnetic actuation has also been used in various LOC devices for bioassay processes such as sample preparation, sample injection, sample manipulation, reaction, separation, and detection. A magnetic field gradient was used to either remotely manipulate or selectively separate biological materials attached to magnetic particles, as well as cells. Magnetic actuation is generally not affected by surface charges, pH, ionic concentrations, or temperature; therefore the use of magnetic particles in biomedical devices gives an additional degree of freedom. **Figure 24** shows several concepts to manipulate magnetic particles as well as magnetically labeled cells in microfluidic devices (Pamme 2006). Magnetic particles can be transported by time-varying electromagnetic fields along a track of meandering gold wires (Deng *et al.* 2001), along a sawtooth gold wire (Wirix and de Boeck 2004), and along a gold wire mesh (Lee *et al.* 2004). In magnetic transport, manipulation of magnetic particles is a difficult task because the magnetic susceptibility of the magnetic particles is rather weak due to small magnetic core volumes and the demagnetization effects of the particles. On the other hand, magnetic separation is different from magnetic transport in the sense that the magnetic particles are captured and therefore separated by magnetic actuation, but are transported using a liquid flow (Gijs 2004).

Magnetic separation of cells (magnetic-activated cell separation, MACS) labeled with magnetic beads in tubes or capillaries has become a common method for the separation of antigens, antibodies, DNA, and even cells (Dunnill and Lilly 1974, Kondo *et al.* 1994, Koneracka *et al.* 2002, Kourilov and Steinitz 2002, Mosbach and Andersson 1977, Rashkovetsky *et al.* 1997). In microscales, H-shaped channel networks can be used for continuous flow magnetic separation as illustrated in **Figure 25**. Kim and Park (2005) reported a microfluidic immunoassay utilizing binding of superparamagnetic nanoparticles to beads and deflection of these beads in a magnetic field as the signal for measuring the presence of analyte. The superparamagnetic 50-nm nanoparticles and fluorescent 1-μm polystyrene beads were immobilized with specific antibodies. When target analytes reacted with the polystyrene beads and superparamagnetic nanoparticles simultaneously, the superparamagnetic nanoparticles were attached to the microbeads by the antigen–antibody complex. With a high-gradient magnetic field, only the microbeads conjugated with superparamagnetic nanoparticles by analytes moved to the magnetic fields and were consequently separated through a corresponding outlet channel.

Pamme and Manz (2004) demonstrated on-chip free-flow magnetophoresis capable of separating magnetic from nonmagnetic particles as well as separating magnetic particles of different sizes from

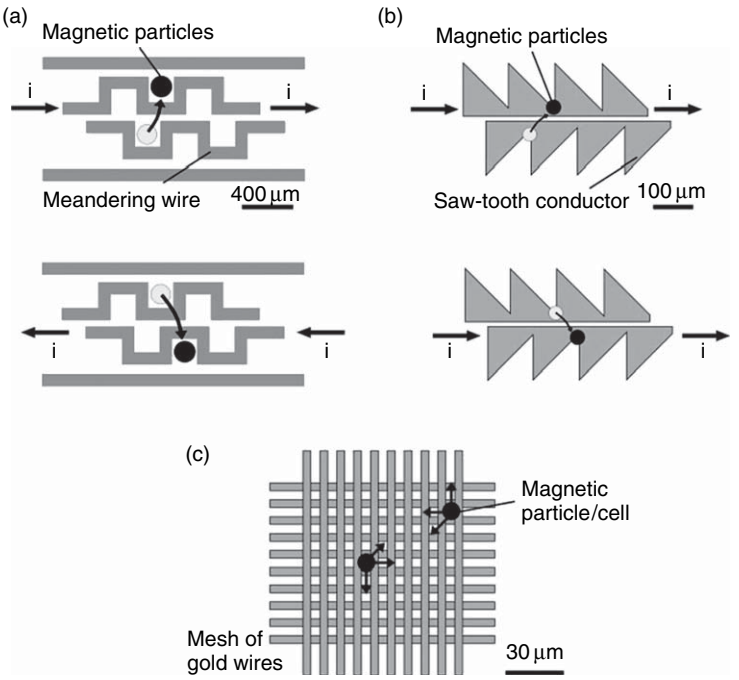


Figure 24 Magnetic particles can be transported by time-varying electromagnetic fields: (a) along a track of meandering gold wires; (b) along a sawtooth gold wire; and (c) along a gold wire mesh. (Source: Pamme N 2006 Magnetism and microfluidics. *Lab Chip* **6**, 24–38.)

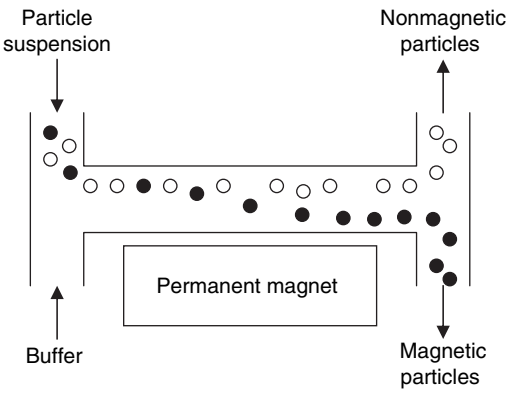


Figure 25 Principle of H-shaped separators for continuous flow separation.

each other (see eqn [27]). In continuous flow, magnetic particles were deflected from the direction of laminar flow by a perpendicular magnetic field by means of an assembly of NdFeB magnets, depending on their magnetic susceptibility and size and on the flow rate. The 4.5-μm particles with larger susceptibility were deflected further from the direction of laminar flow than the 2.0-μm magnetic particles. Nonmagnetic 6-μm polystyrene beads, however, were not deflected at all.

Instead of using permanent magnets to generate nonuniform magnetic field or magnetic field gradient for trapping magnetic particles in flow, Choi, Ahn, and coworkers reported several microfluidic devices with integrated electromagnets. Ahn and coworkers fabricated 3D electromagnets with poles adjacent to a microchannel (Ahn *et al.* 1996), spiral-type electromagnets (Choi *et al.* 2001a), and serpentine-type electromagnets (Choi *et al.* 2000, 2001b, 2002). Depending on the electromagnet design, currents of 500–30 mA were required to capture magnetic particles at the electromagnet poles. **Figure 26** shows the concept of magnetic particle-based bioseparation using the serpentine-type electromagnet for the case of antigen–antibody sandwich immunoassay. Antibody-coated magnetic particles are introduced on the electromagnet (**Figure 26(a)**) and are separated by applying magnetic fields. While holding the antibody-coated magnetic particles (**Figure 26(b)**), antigens are injected into the channel (**Figure 26(c)**). Only target antigens are immobilized and, thus, separated onto the magnetic particle surface due to antigen–antibody interaction (**Figure 26(d)**). Other antigens are washed out with the flow. In a subsequent step, labeled secondary antibodies are introduced and incubated with the immobilized

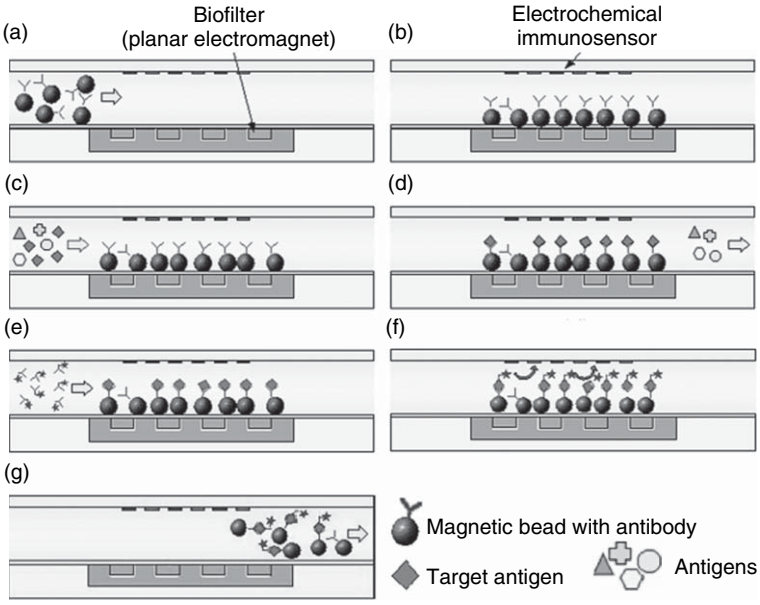


Figure 26 Conceptual illustration of biosampling and immunoassay procedure using magnetic particle approach: (a) injection of magnetic particles labeled with antibody; (b) separation and holding of particles; (c) flowing sample; (d) immobilization of target antigen; (e) flowing labeled antibody; (f) electrochemical detection after adding enzyme substrate; and (g) washing out magnetic particles and ready for another immunoassay. (Source: Choi J-W, Oh K W, Thomas J H, Heineman W R, Halsall H B, Nevin J H, Helmicki A J, Henderson H T, Ahn C H 2002 An integrated microfluidic biochemical detection system for protein analysis with magnetic bead-based sampling capabilities. *Lab Chip* **2**, 27–30.)

antigens after which the chamber is rinsed to remove all unbound secondary antibodies (**Figure 26(e)**). The substrate solution, which will react with the enzyme, is injected into the channel and the electrochemical detection is performed (**Figure 26(f)**). Finally, the magnetic particles are released and the bioseparator is ready for the next sample (**Figure 26(g)**).

At the microscale, magnetic microparticles coated with biomolecules were often used as solid supports for antigens/antibodies (Choi *et al.* 2002, Kim and Park 2005), DNA (Fan *et al.* 1999, Liu *et al.* 2004), RNA (Jiang and Harrison 2000), and even cells (Furdui *et al.* 2003). However, in the cases of red blood cells (Han and Frazier 2006) and magnetostatic bacteria (Lee *et al.* 2004) the additional magnetic particles coated with specific biomolecules were not necessary due to their natural magnetophoretic properties. Furlani (2006) used analytical analysis to study magnetophoretic particle transport and capture a microsystem consisting of an array of integrated soft magnetic elements embedded in a nonmagnetic substrate beneath a microfluidic channel. In addition to magnetophoretic particle transport, manipulation, and separation, the magnetic properties of the particles can also be used for detection using sensitive magnetic sensors such as giant magnetoresistive (GMR) sensors, spin-valve sensors, superconducting quantum interference devices (SQUID), and miniaturized hall sensors (Pamme 2006).

2.02.5 Summary

Magnetic actuation is one of the most widely used actuation principles found in our daily lives. In this chapter, we studied magnetic actuation and discussed several related topics including basic magnetic theory, scaling effect, the Lorentz force, magnetic induction, magnetic force and torque, magnetic field gradient, magnetization, and magnetic actuation mechanisms pertaining to microscale magnetic actuators. We also discussed representative magnetic materials such as permanent magnets, hard and soft magnetic materials, and magnetic particles. Fabrication processes such as electroplating of ferromagnetic materials and electromagnetic coils that are key building blocks in the magnetic microsystems were also covered. Finally, selected examples of magnetic microactuators in the areas of

valves, pumps, mixers, mirrors, switches, as well as manipulation of magnetic particles and magnetic cell separation were reviewed. In addition to magnetic actuation, the magnetic microsystems can also be advantageous to magnetic sensing. With advancements in these magnetic-based microsystems, magnetic actuation will be widely applicable to many microsystems.

References

Ahn C H, Kim Y J, Allen M G 1993 A planar variable reluctance magnetic micromotor with fully integrated stator and wrapped coils. *J. Microelectomech. Syst.* **2**, 165–73

Ahn C H, Allen M G 1995 *Proc. MEMS '95, 8th IEEE International Workshop, Micro Electro Mechanical Systems*, Amsterdam, The Netherlands, pp. 408–12

Ahn C H, Allen M G 1998 Micromachined planar inductors on silicon wafers for MEMS applications. *IEEE Trans. Ind. Electron.* **45**, 866–76

Ahn C H, Allen M G, Trimmer W, Jun Y N, Erramilli S 1996 A fully integrated micromachined magnetic particle separator. *J. Microelectromech. Syst.* **5**, 151–8

Bae B, Kim N, Kee H, Kim S-H, Lee Y, Lee S, Park K 2002 Feasibility test of an electromagnetically driven valve actuator for glaucoma treatment. *J. Microelectromech. Syst.* **11**, 344–54

Bae B, Kee H, Kim S, Lee Y, Sim T, Kim Y, Park K 2003 In vitro experiment of the pressure regulating valve for a glaucoma implant. *J. Micromech. Microeng.* **13**, 613–9

Bajorek C H, Thompson D A 1975 Permanent magnet films for biasing of magnetoresistive transducers. *IEEE Trans. Mag.* **11**, 1209–11

Barbic M, Mock J J, Gray A P, Schultz S 2001 Electromagnetic micromotor for microfluidics applications. *Appl. Phys. Lett.* **79**, 1399–401

Bau H H, Zhong J H, Yi M Q 2001 A minute magneto hydro dynamic (MHD) mixer. *Sens. Actuators B* **79**, 207–15

Bau H H, Zhu J Z, Qian S Z, Xiang Y 2003 A magneto-hydrodynamically controlled fluidic network. *Sens. Actuators B* **88**, 205–16

Benham P P, Crawford R J, Armstrong C G 1996 *Mechanics of Engineering Materials*. Longman, Essex

Bernstein J J, Taylor W P, Brazzle J D, Corcoran C J, Kirkos G, Odhner J E, Pareek A, Waelti M, Zai M 2004 Electromagnetically actuated mirror arrays for use in 3-D optical switching applications. *J. Microelectromech. Syst.* **13**, 526–35

Bhansali S, Zhang A L, Zmood R B, Jones P E, Sood D K 2000 Prototype feedback-controlled bidirectional actuation system for MEMS applications. *J. Microelectromech. Syst.* **9**, 245–51

Biswal S L, Gast A P 2004 Micromixing with linked chains of paramagnetic particles. *Anal. Chem.* **76**, 6448–55

Bohm S, Olthuis W, Bergveld P 1999 A plastic micropump constructed with conventional techniques and materials. *Sens. Actuators* **77**, 223–8

Bohm S, Burger G J, Korthorst M T, Roseboom F 2000 A micromachined silicon valve driven by a miniature bi-stable electro-magnetic actuator. *Sens. Actuators* **80**, 77–83

Bosh D, Heimhofer B, Muck G, Seidel H, Thumser U, Welser W 1993 A silicon microvalve with combined electromagnetic/electrostatic actuation. *Sens. Actuators A* **37/38**, 684–92

Bourouina T, Lebrasseur E, Reyne G, Debray A, Fujita H, Ludwig A, Quandt E, Muro H, Oki T, Asaoka A 2002 Integration of two degree-of-freedom magnetostrictive actuation and piezoresistive detection: Application to a two-dimensional optical scanner. *J. Microelectromech. Syst.* **11**, 355–61

Capanu M, Boyd J G, Hesketh P J 2000 Design, fabrication, and testing of a bistable electromagnetically actuated microvalve. *J. Microelectromech. Syst.* **9**, 181–9

Chin T S 2000 Permanent magnet films for applications in microelectromechanical systems. *J. Magn. Magn. Mater.* **209**, 75–9

Cho H J, Ahn C H 2002 A bidirectional magnetic microactuator using electroplated permanent magnet arrays. *J. Microelectromech. Syst.* **11**, 78–84

Cho H J, Ahn C H 2003 Magnetically-driven bi-directional optical microscanner. *J. Micromech. Microeng.* **13**, 383–9

Cho H J, Bhansali S, Ahn C H 2000 Electroplated thick permanent magnet arrays with controlled direction of magnetization for MEMS application. *J. Appl. Phys.* **87**, 6340–2

Cho H J, Oh K W, Ahn C H, Boolchand P, Nam T-C 2001 Stress analysis of silicon membranes with electroplated permalloy films using Raman scattering. *IEEE Trans. Magn.* **37**, 2749–51

Cho I, Song J T, Baek S H, Yoon E 2005 A low-voltage, low-power RF MEMS series, shunt switches actuated by combination of electromagnetic, electrostatic forces. *IEEE Trans. Microw. Theor. Tech.* **53**, 2450–7

Choi J W, Ahn C H, Bhansali S, Henderson H T 2000 A new magnetic bead-based, filterless bio-separator with planar electromagnet surfaces for integrated bio-detection systems. *Sens. Actuators B* **68**, 34–9

Choi J W, Liakopoulos T M, Ahn C H 2001a An on-chip magnetic bead separator using spiral electromagnets with semi-encapsulated permalloy. *Biosens. Bioelectron.* **16**, 409–16

Choi J-W, Oh K W, Han A, Wijayawardhana C A, Lannes C, Bhansali S, Schlueter K T, Heineman W R, Halsall H B, Nevin J H, Helmicki A J, Henderson H T, Ahn C H 2001b Development and characterization of microfluidic devices and systems for magnetic bead-based biochemical detection. *Biomed. Microdevices* **3**, 191–200

Choi J-W, Oh K W, Thomas J H, Heineman W R, Halsall H B, Nevin J H, Helmicki A J, Henderson H T, Ahn C H 2002 An integrated microfluidic biochemical detection system for protein analysis with magnetic bead-based sampling capabilities. *Lab Chip* **2**, 27–30

Cugat O, Delamare J, Reyne G 2003 Magnetic micro-actuators and systems (MAGMAS). *IEEE Trans. Magn.* **39**, 3607–12

Cugat O, Reyne G, Delamare J, Rostaing H 2006 Novel magnetic micro-actuators and systems (MAGMAS) using permanent magnets. *Sens. Actuators A* **129**, 265–9

Debray A, Ludwig A, Bourouina T, Asaoka A, Tiercelin N, Reyne G, Oki T, Quandt E, Muro H, Fujita H 2004 Application of a multilayered magnetostrictive film to a micromachined 2-D optical scanner. *J. Microelectromech. Syst.* **13**, 264–71

Deng T, Whitesides G M, Radhakrishnan M, Zabow G, Prentiss M 2001 Manipulation of magnetic microbeads in suspension using micromagnetic systems fabricated with soft lithography. *Appl. Phys. Lett.* **78**, 1775–7

Divoux C, Cugat O, Reyne G 1998 Deformable mirror using magnetic membranes: Application to adaptive optics in astrophysics. *IEEE Trans. Magn.* **34**, 3564–7

Dunnill P, Lilly M D 1974 Purification of enzymes using magnetic bio-affinity materials. *Biotechnol. Bioeng.* **16**, 987–90

Dutoit B 1999 High performance micromachined Sm_2Co_{17} polymer bonded magnets. *Sens. Actuators* **77**, 178–82

Fan Z H, Mangru S, Granzow R, Heaney P, Ho W, Dong Q P, Kumar R 1999 Dynamic DNA hybridization on a chip using paramagnetic beads. *Anal. Chem.* **71**, 4851–9

Ferreira L O S, Moehlecke S 1999 A silicon micromechanical galvanometric scanner. *Sens. Actuators A* **73**, 252–60

Fischer K, Guckel H 1998 Long throw linear magnetic actuators stackable to one millimeter of structural height. *Microsyst. Technol.* **4**, 180–3

Fredrickson C K, Fan Z H 2004 Macro-to-micro interfaces for microfluidic devices. *Lab Chip* **4**, 526–33

Fu C, Rummler Z, Chomburg W 2003 Magnetically driven micro ball valves fabricated by multilayer adhesive film bonding. *J. Micromech. Microeng.* **13**, S96–102

Furdui V I, Kariuki J K, Harrison D J 2003 Microfabricated electrolysis pump system for isolating rare cells in blood. *J. Micromech. Microeng.* **13**, S164–70

Furlani E P 2001 *Permanent Magnet and Electromechanical Devices: Materials, Analysis and Applications*. Academic Press, New York

Furlani E P 2006 Analysis of particle transport in a magnetophoretic microsystem. *J. Appl. Phys.* **99**, 024912

Gibbs M R J, Hill E W, Wright P J 2004 Magnetic materials for MEMS applications. *J. Phys. D. Appl. Phys.* **37**, R237–44

Gijs M A M 2004 Magnetic bead handling on-chip: New opportunities for analytical applications. *Microfluid Nanofluid* **1**, 22–40

Goettsche T, Kohnle J, Willmann M, Ernst H, Spieth S, Tischler R, Messner S, Zengerle R, Sandmaier H 2005 Novel approaches to particle tolerant valves for use in drug delivery systems. *Sens.Actuators A* **118**, 70–7

Gretillat M A, Gretillat F, de Rooij N F 1999 Micromechanical relay with electrostatic actuation and metallic contacts. *J. Micromech. Microeng.* **9**, 324–31

Guckel H, Skrobis K J, Christenson T R, Klein J, Han S, Choi B, Lovell E G, Chapman T W 1991 Fabrication and testing of the planar magnetic micromotor. *J. Micromech. Microeng.* **1**, 135–8

Guckel H, Christenson T, Skrobis K 1992 Metal micromechanisms via deep X-ray lithography, electroplating and assembly. *J. Micromech. Microeng.* **2**, 225–8

Han K H, Frazier A B 2006 Paramagnetic capture mode magnetophoretic microseparator for high efficiency blood cell separations. *Lab Chip* **6**, 265–73

Hannemann U, Neu V, Fähler S, Holzapfel B, Schultz L 2002 Hard magnetic Nd–Fe–B and Sm–Co films: Influence of the Cr buffer microstructure on the magnetic properties. *IEEE Trans. Magn.* **38**, 2949–51

Hartshorne H, Backhouse C J, Lee W E 2004 Ferrofluid-based microchip pump and valve. *Sens. Actuators B* **99**, 592–600

Hatch A, Kamholz A E, Holman G, Yager P, Boringer K F 2001 A ferrofluidic magnetic micropump. *J. Microelectromech. Syst.* **10**, 215–21

Hesketh P J, Bintoro J S, Luharuka R 2004 Microvalve for fuel cells and miniature gas chromatographic system. *Sensor. Update* **13**, 233–302

Horng H E, Chen C S, Fang K L, Yang S Y, Chieh J J, Hong Chin Y, Yang H C 2004 Tunable optical switch using magnetic fluids. *Appl. Phys. Lett.* **85**, 5592–4

Huang L, Wang W, Murphy M C, Lian K, Ling Z G 2000 LIGA fabrication and test of a DC type magnetohydrodynamic (MHD) micropump. *Microsyst. Technol.* **6**, 235–40

Jang J S, Lee S S 2000 Theoretical and experimental study of MHD (magnetohydrodynamic) micropump. *Sens. Actuators A* **80**, 84–9

Jang Y H, Kim Y K 2003 Design, fabrication and characterization of an electromagnetically actuated addressable out-of-plane micromirror array for vertical optical source applications. *J. Micromech. Microeng.* **13**, 853–63

Jeong S, Hsu Y N, Laughlin D E, McHenry M E 2000 Magnetic properties of nanostructured CoPt and FePt thin films. *IEEE Trans. Magn.* **36**, 2336–8

Jerman H 1994 Electrically activated, normally closed diaphragm valves. *J. Micromech. Microeng.* **4**, 210–6

Ji C H, Yee Y, Choi J, Kim S H, Bu J U 2004 Electromagnetic 2×2 MEMS optical switch. *IEEE J. Sel. Top. Quant. Electron.* **10**, 545–50

Jiang G F, Harrison D J 2000 mRNA isolation in a microfluidic device for eventual integration of cDNA library construction. *Analyst* **125**, 2176–9

Judy J W 2001 Microelectromechanical system (MEMS): Fabrication, design and applications. *Smart Mater. Struct.* **10**, 115–34

Kim E-G, Oh J-G, Choi B 2006 A study on the development of a continuous peristaltic micropump using magnetic fluids. *Sens. Actuators A* **128**, 43–51

Kim K S, Park J K 2005 Magnetic force-based multiplexed immunoassay using superparamagnetic nanoparticles in microfluidic channel. *Lab Chip* **5**, 657–64

Kinoshita H, Hoshino K, Matsumoto K, Shimoyama I 2006 A thin camera with a zoom function using reflective optics. *Sens. Actuators A* **128**, 191–6

Kittel C 1996 *Introduction to Solid State Physics*. Wiley, New York

Kondo A, Kamura H, Higashitani K 1994 Development and application of thermosensitive magnetic immunomicrospheres for antibody purification. *Appl. Microbiol. Biotechnol.* **41**, 99–105

Koneracka M, Kopcansky P, Timko M, Ramchand C N 2002 Direct binding procedure of proteins and enzymes to fine magnetic particles. *J. Magn. Magn. Mater.* **252**, 409–11

Kourilov V, Steinitz M 2002 Magnetic-bead enzyme-linked immunosorbent assay verifies adsorption of ligand and epitope accessibility. *Anal. Biochem.* **311**, 166–70

Kruglick E J J, Pister K S J 1999 Bistable MEMS relays and contact characterization. *J. Microelectromech. Syst.* **8**, 264–71

Krusemark O, Feustel A, Muller J 1998 Micro ball valve for fluidic micropumps and gases. *Proceedings of Micro Total Analysis Systems '98 Workshop*, Banff, Canada, pp. 399–402

Lagorce L K, Brand O, Allen M G 1999 Magnetic microactuators based on polymer magnet. *IEEE J. Microelectromech. Syst.* **8**, 2–9

Laser D J, Santiago J G 2004 A review of micropumps. *J. Micromech. Microeng.* **14**, R35–64

Lee H, Purdon A M, Chu V, Westervelt R M 2004 Controlled assembly of magnetic nanoparticles from magnetotactic bacteria using microelectromagnets arrays. *Nano Lett.* **4**, 995–8

Lemoff A V, Lee A P 2000 An AC magnetohydrodynamic micropump. *Sens. Actuators B* **63**, 178–85

Leventis N, Gao X R 2001 Magnetohydrodynamic electrochemistry in the field of Nd–Fe–B magnets. Theory; experiment; and application in self-powered flow delivery systems. *Anal. Chem.* **73**, 3981–92

Liakopoulos T M, Zhang W, Ahn C H 1996 Micromachined thick permanent magnet arrays on silicon wafers. *IEEE Trans. Magn.* **32**, 5154–6

Lileev A S, Parilov A A, Blatov V G 2002 Properties of hard magnetic Nd–Fe–B films vs different sputtering conditions. *J. Magn. Magn. Mater.* **242–245**, 1300–3

Liu C 2006 *Foundations of MEMS*. Pearson Education, Upper Saddle River, NJ

Liu R H, Yang J N, Lenigk R, Bonanno J, Grodzinski P 2004 Self contained, fully integrated biochip for sample preparation, polymerase chain reaction amplification, and DNA microarray detection. *Anal. Chem.* **76**, 1824–31

Lu L H, Ryu K S, Liu C 2002 A magnetic microstirrer, array for microfluidic mixing. *J. Microelectromech. Syst.* **11**, 462–9

Maekoba H, Helin P, Reyne G, Bourouina T, Fujita H 2001 Self-aligned mirror and V-grooves applied to an optical switch: Modeling and optimization of bi-stable operation by electromechanical actuation. *Sens. Actuators A* **87**, 172–8

McDonald S, Pan T, Ziaie B 2004 *Proc. 26th Annu. Int. Conf. IEEE EMBS*, San Francisco, CA, USA, pp. 2650–3

Meckes A, Behrens J, Kayser O, Benecke W, Becker T, Muller G 1999 Microfluidic system for the integration and cyclic operation of gas sensors. *Sens. Actuators A* **76**, 478–83

Miyajima H, Asaoka N, Isokawa T, Ogata M, Aoki Y, Imai M, Katashiro F O M, Matsumoto K 2003 A MEMS electromagnetic optical scanner for a commercial confocal laser scanning microscope. *J. Microelectromech. Syst.* **12**, 243–51

Mosbach K, Andersson L 1977 Magnetic ferrofluids for preparation of magnetic polymers and their application in affinity chromatography. *Nature* **270**, 259–61

Myers H P 1997 *Introductory Solid State Physics*. 2nd edn. Taylor & Francis, London

Myung N V, Park D Y, Yoo B Y, Sumodjo P T A 2003 Development of electroplated magnetic materials for MEMS. *J. Magn. Magn. Mater.* **265**, 189–98

Nami Z, Ahn C H, Allen M G 1996 An energy-based design criterion for magnetic microactuators. *J. Micromech. Microeng.* **6**, 337–44

Nguyen N-T, Wu Z 2005 Micromixers – A review. *J. Micromech. Microeng.* **15**, R1–6

Niarchos D 2003 Magnetic MEMS: key issues and some applications. *Sens. Actuators A* **109**, 166–73

Oh K W, Ahn C H 2006 A review of microvalves. *J. Micromech. Microeng.* **16**, R13–39

Oh K W, Rong R, Ahn C H 2001 In-line micro ball valve through polymer tubing. *Proceedings of Micro-TAS 2001*, Monterrey, CA, USA, pp. 407–8

Oh K W, Han A, Bhansali S, Ahn C H 2002 A low-temperature bonding technique using spin-on fluorocarbon polymers to assemble microsystems. *J. Micromech. Microeng.* **12**, 187–91

Oh K W, Rong R, Ahn C H 2005a Miniaturization of pinch-type valves and pumps for practical micro total analysis system integration. *J. Micromech. Microeng* **15**, 2449–55

Oh K W, Namkoong K, Chinsung P 2005b *Proceedings of Micro-TAS 2005*, Boston, MA, USA, pp. 554–6

Pamme N 2006 Magnetism and microfluidics. *Lab Chip* **6**, 24–38

Pamme N, Manz A 2004 On-chip free-flow magnetophoresis: Continuous flow separation of magnetic particles and agglomerates. *Anal. Chem.* **76**, 7250–6

Pan T, McDonald S J, Kai E M, Ziaie B 2005 A magnetically driven PDMS micropump with ball check-valves. *J. Micromech. Microeng.* **15**, 1021–6

Park J Y, Allen M G 1998 Development of magnetic materials and processing techniques applicable to integrated micromagnetic devices. *J. Micromech. Microeng.* **8**, 307–16

Rashkovetsky L G, Lyubarskaya Y V, Foret F, Hughes D E, Karger B L 1997 Automated microanalysis using magnetic beads with commercial capillary electrophoretic instrumentation. *J. Chromatogr. A* **781**, 197–204

Receveur R A M, Marxer C R, Woering R, Larik V C M H, de Rooij N F 2005 Laterally moving bistable MEMS DC switch for biomedical applications. *J. Microelectromech. Syst.* **14**, 1089–98

Ren H, Gerhard E 1997 Design and fabrication of a current-pulse-excited bistable magnetic microactuator. *Sens. Actuators A* **58**, 259–64

Rida A, Gijs M A M 2004 Manipulation of self-assembled structures of magnetic beads for microfluidic mixing and assaying. *Anal. Chem.* **76**, 6239–46

Rieger G, Wecker J, Rodewald W, Sattler W, Bach F-W, Duda T, Unterberg W 2000 Nd–Fe–B permanent magnet (thick films) produced by a vacuum-plasma-spraying process. *J. Appl. Phys.* **87**, 5329–31

Ripka P 2001 *Magnetic Sensors and Magnetometers*. Artech House, Norwood, MA

Rosensweig R E 1985 *Ferrohydrodynamics*. Cambridge University Press, Cambridge

Ruan M, Shen J, Wheeler C B 2001a Latching microelectromagnetic relays. *Sens. Actuators A* **91**, 346–50

Ruan M, Shen J, Wheeler C B 2001b Latching micromagnetic relays. *J. Microelectromech. Syst.* **10**, 511–17

Ryu K S, Shaikh K, Goluch E, Fan Z F, Liu C 2004 Micro magnetic stir-bar mixer integrated with parylene microfluidic channels. *Lab Chip* **4**, 608–13

Sadler D J, Oh K W, Ahn C H, Bhansali S, Henderson H T 1999 *10th Int. Conf. Solid-State Sensors and Actuators, Transducers '99*, Sendai, Japan, pp. 1812–15

Santra S, Holloway P, Batich C D 2002 Fabrication and testing of a magnetically actuated micropump. *Sens. Actuators B* **87**, 358–64

Schiele I, Huber J, Hillerich B, Kozlowski F 1998 Surface micromachined electrostatic microrelay. *Sens. Actuators A* **66**, 345–54

Suzuki H, Ho C M, Kasagi N 2004 A chaotic mixer for magnetic bead-based micro cell sorter. *J. Microelectromech. Syst.* **13**, 779–90

Taylor W P, Brand O, Allen M G 1998 Fully integrated magnetically actuated micromachined relays. *J. Microelectromech. Syst.* **7**, 181–91

Terry S C, Jerman J H, Angell J B 1979 A gas chromatographic air analyzer fabricated on a silicon wafer. *IEEE Trans. Electron. Devices* **26**, 1880–6

Von Hippel A R 1959 *Molecular Science and Molecular Engineering*. The Technology Press of MIT and Wiley, New York, NY

West J, Gleeson J P, Alderman J, Collins J K, Berney H 2003 Structuring laminar flows using annular magnetohydrodynamic actuation. *Sens. Actuators B* **96**, 190–9

Wirix-Speetjens R, de Boeck J 2004 On-chip magnetic particle transport by alternating magnetic field gradients. *IEEE Trans. Magn.* **40**, 1944–6

Woias P 2005 Micropumps – Past, progress and future prospects. *Sens. Actuators B* **105**, 28–38

Yamahata C, Chastellain M, Parashar V K, Petri A, Hofmann H, Gijs M A M 2005 Plastic micropump with ferrofluidic actuation. *J. Microelectromech. Syst.* **14**, 96–102

Yan C, Zhao X, Ding G, Zhang C, Cai B 2001 Axial flux electromagnetic micromotor. *J. Micromech. Microeng.* **11**, 113–7

Yanagisawa K, Kuwano H, Tapo A 1993 An electromagnetically driven microvalve. *Proceedings of Transducers '93*, Yokohama, Japan, pp. 102–5

Zborowski M, Sun L P, Moore L R, Williams P S, Chalmers J J 1999 Continuous cell separation using novel magnetic quadrupole flow sorter. *J. Magn. Magn. Mater.* **194**, 224–30

Zhang W, Ahn C H 1996 A bi-directional magnetic micropump on a silicon wafer. *Proc. Solid-State Sensor and Actuator Workshop*, Hilton Head, SC, USA, pp. 94–7

Zhong J H, Yi M Q, Bau H H 2002 Magneto hydrodynamic (MHD) pump fabricated with ceramic tapes. *Sens. Actuators A,* 59–66

Biographies

Kwang W. Oh joined the Department of Electrical Engineering at the University at Buffalo (UB), the State University of New York, Buffalo (SUNY-Buffalo), NY, in March 2006. After receiving his Ph.D. degree from the University of Cincinnati in 2001, he worked as a technical staff member in the biolab at Samsung Advanced Institute of Technology, Korea. At Samsung, he was responsible for microfluidics in developing lab-on-a-chip (LOC) platforms for clinical diagnostic applications. His main research activity in SMALL (nanobioSensors and MicroActuators Learning Laboratory) at UB is Small (nano & micro) Science & Technology, including sample-to-answer nano/microfluidic system; nonobiosensors and microactuators; world-to-chip interfacing and packaging; and single cell manipulation, sorting, and detection. He has an excellent record of publications with more than 42 journal and conference proceeding papers, including one topical review paper on microvalves and 24 patents in the fields of MEMS, BioMEMS, and microfluidics.

Chong H. Ahn received his Ph.D. degree in Electrical and Computer Engineering from the Georgia Institute of Technology, Atlanta, GA, in 1993. From 1993 to 1994, he was a postdoctoral associate at the Georgia Institute of Technology and at IBM T.J. Watson Research Center, Yorktown Heights, NY. In 1994, he joined the Department of Electrical and Computer Engineering and Computer Science at the University of Cincinnati, Cincinnati, OH, where he currently holds the rank of professor in the Departments of Electrical Engineering and Biomedical Engineering. He is also the director of the microsystems and Bio-MEMS lab at the University of Cincinnati. He has published approximately 170 journal and conference proceeding papers related to the MEMS, BioMEMS, and microfluidics.

2.03 Thermal Actuation

Long Que, Micro- and Nanostructure Technologies Center, GE Global Research, One Research Circle, Niskayuna, NY, USA

Glossary

MTI Microfluidic Transdermal Interface

OCT Optical Coherence Tomography

2.03.1 Introduction

The microactuator is one of the key components to form a real comprehensive microsystem. Other key components are microsensors, control circuits, and communication interface (Najafi *et al.* 1994). Over the past two decades, extensive efforts have been devoted to the microactuator development. Different actuation mechanisms have been described and summarized in detail (Que and Gianchandani 2005). Those most widely explored so far include electrostatic (Tang *et al.* 1989), thermal (Que *et al.* 2001a, b), piezoelectric (Schiller and Polla 1993), shape memory alloy (SMA) (Ma *et al.* 2000), electromagnetic (Guckel *et al.* 1992), solid–liquid phase change, thermopneumatic (Zdeblick *et al.* 1994), and microbubble and microsteam (Lin *et al.* 1991, Sniegowski 1993).

Output force F, displacement u, volume v, and response time t are four critical parameters to quantify the performance of microactuators. Another useful parameter, work output per unit volume (W), is defined as

$$W = \frac{\vec{F} \cdot \vec{u}}{v} \qquad [1]$$

Typically, as the displacement of an actuator increases, the available force at that displacement decreases. Usually, the displacement of a microactuator refers to the peak displacement with no loading force under a fixed actuation voltage, current, or

power, whereas the output force refers to the peak force with no displacement at the same actuation condition. The offered output force at each intermediate displacement can be obtained based on those two peak values. Several performance parameters for major actuators are summarized in **Table 1**.

The thermal actuator can provide both high force (up to several millinewtons) and large displacement (up to several hundred micrometers), but its speed spans from several hertz to several tens of kilohertz depending on the structural material, design, and target displacement, which is limited by the cooling procedure during each actuation cycle. However, its actuation voltage can be kept <15 V, which is compatible with standard integrated chip (IC) testing. More will be discussed in the following sections.

2.03.2 Thermal Actuation

2.03.2.1 Basic Theory

The thermal expansion or contraction property of a material is the basis for thermal actuation. Therefore, the materials can be gas, liquid, or solid as far as they are expandable or shrinkable when heated or cooled down, respectively. The heating of the structural material can be generated by either passing current through an embedded resistor (Joule heating) or using optical excitation (optothermal effect) (Oliver *et al.* 2003). In contrast, the cooling of the structural material can be achieved by immersing the actuator into a low-temperature environment, such as low-temperature liquid or water.

Even though the materials for thermal actuation cover a very broad range and the configurations of each specific actuator are different, a basic heat transfer model can be applicable to all of them (Que 2000). This model can help to analyze the performance of thermal actuators, including their displacement, volume expansion, output force, response speed, power consumption, and even their reliability. Therefore, it provides an excellent guidance and tool for the thermal actuator design. In order to describe this model and understand its basic principles and approaches, a specific actuator design is illustrated in the following.

In this case, a simple bent-beam electrothermal actuator is used as an example to analyze the heat transfer process (Lott *et al.* 2002, Que and Gianchandani 2005). The basic schematic diagram of a bent-beam actuator is shown in **Figure 1**, which includes all the design parameters essential for the simulation. The device consists of two connected shallow-angled beams suspended above the substrate while their ends are anchored to the substrate. The device is typically operated in air at temperature T_{air}. When the current I is applied, the temperature of the beams rises due to Joule heating. There are three different heat dissipation mechanisms: (1) the heat directly conducts through the beam itself, to the anchor, and then to the substrate; (2) the heat is dissipated by conductive and convective mechanisms through the air; and (3) the heat is dissipated into the environment in the form of infrared radiation.

Table 1 Major actuators and their typical performance

Actuator type	*Displacement (μm)*	*Output force (μN)*	*Actuation voltage (V)*	*Speed (Hz)*	*W (J m^{-3})*	*Comments*
Electrostatic	0.1–30	0.1 to 1×10^3	~50 to 120	3×10^3	7.0×10^2 to 1.8×10^5	Parallel plate Comb drive Force array
Thermal	10–100	10 to 1×10^4	<20	10^2–10^3	4.6×10^5	Bimorph Pseudo-bimorph Bent-beam actuator Thermal responsive polymer
Piezoelectric	~10	10 to 1×10^6	20 to 10^3	10×10^3	1.2×10^5 1.8×10^2	PZT-based device ZnO-based device
SMA	10–570	$(10200) \times 10^3$	1–3	20	2.5×10^7 to 6.0×10^6	Ni–Ti-based device
Magnetic	Up to 10^3	Up to 10^5	<10 mA (current)	10^2–10^4	1.6×10^3 to 4.0×10^5	Magnetostatic Electromagnetic Magnetostrictive

Source: Que and Gianchandani (2005).

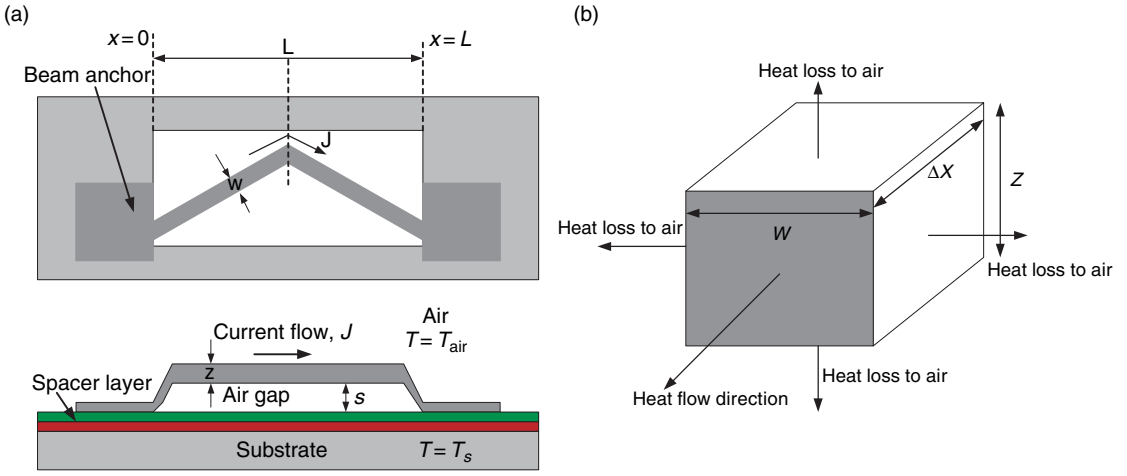


Figure 1 (a) Top view and cross-sectional view of the bent-beam actuators and (b) the differential element of the bent beam.

In order to set up a finite difference model, the device should be divided into discrete volume elements. The differential element of the bent beam of width w, thickness z, and length Δx is given in Figure 1(b). The heat balance equation for the element is

$$Q_{\text{total}} = P_{\text{g}} + Q_{\text{conduction}} + Q_{\text{convection}} + Q_{\text{radiation}} \quad [2]$$

where Q_{total} is the net rate of the change of the internal energy of the element, P_{g} the power generated inside the element by the Joule heating, $Q_{\text{conduction}}$ the conduction heat loss in the element, $Q_{\text{convection}}$ the convection heat loss, and $Q_{\text{radiation}}$ the radiation heat loss.

Specifically, Q_{total} is expressed as

$$Q_{\text{total}} = \rho_{\text{m}} cwz\Delta x\left(\frac{\partial u}{\partial t}\right) \quad [3]$$

where c and ρ_{m} are the specific heat and density of the beam material, respectively, and u is the average temperature. The power generated by Joule heating is given by

$$\begin{aligned} P_{\text{g}} &= \mathcal{J}^2 \rho(u) wt\Delta x \\ &= \mathcal{J}^2 \rho_0 wz\Delta x(1 + \text{TCR} \times (u - T_0)) \end{aligned} \quad [4]$$

where $\mathcal{J}$ is the current density in the element, ρ_0 its resistivity at T_0 (room temperature), and TCR the temperature coefficient of the resistivity of the material.

The conductive loss component consists of two parts. One is the conduction through the beam itself, e.g., through adjacent differential elements to the substrate to dissipate in the anchor region, and the other is the conduction of heat through the air to the substrate. It is given by

$$\begin{aligned} Q_{\text{conduction}} = &- k_{\text{b}} wz\left(\left.\frac{\partial u}{\partial x}\right|_x - \left.\frac{\partial u}{\partial x}\right|_{x+\Delta x}\right) \\ &- k_{\text{air}} \eta w\Delta x \frac{u - T_{\text{s}}}{s} \end{aligned} \quad [5]$$

where k_{b} and k_{air} are the thermal conductivity of the beam material and the air, respectively, and η is the excess flux coefficient. η accounts for the fringing heat flux of each differential element and s is the air gap between the beam and the substrate.

The convection loss component is expressed as

$$Q_{\text{convection}} = 2b(V_{\text{air}}, P)(w + z)\Delta x(u - T_{\text{air}}) \quad [6]$$

where $b(V_{\text{air}}, P)$ is the convective heat transfer coefficient from the beam to the air as a function of the air flow velocity V_{air} and the absolute pressure P.

Assume that the upper and side faces of the beam radiate to the surrounding air, and the bottom face radiates to the substrate. The radiation loss component is given by

$$\begin{aligned} Q_{\text{radiation}} = &\sigma_{\text{b}}(w + 2z) \\ &\Delta x(u^4 - T_{\text{air}}^4) + \sigma_{\text{b}} w\Delta x(u^4 - T_{\text{s}}^4) \end{aligned} \quad [7]$$

where σ_b is the Stefan–Boltzmann constant for the structural material. Based on eqns [6] and [7] and boundary condition

$$u(x)|_{x=0} = u(x)|_{x=L} = T_{\text{s}} \quad [8]$$

the steady-state solution of this system can be obtained. For most cases, the heat loss by infrared radiation is negligible since the average actuation temperature of

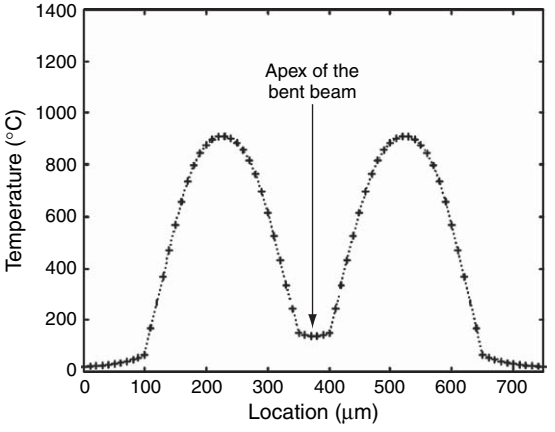


Figure 2 A descriptive calculated temperature profile showing the temperature drop around the apex of the bent-beam actuator. (Source: Lott C, Mclain T, Harb J, Howell L 2002 Modeling the thermal behaviour of a surface-micromachined linear-displacement thermomechanical microactuator. *Sensors and Actuators A* **101**, 239–50.)

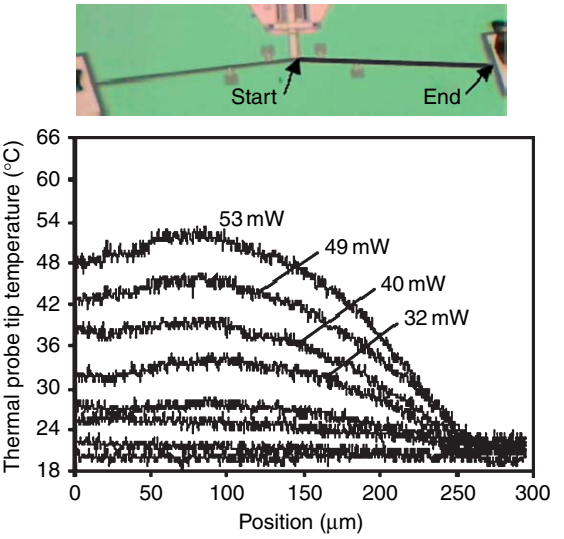


Figure 3 Range on the bent beam of the temperature measurement (*upper*) and measured temperature profile (*lower*). (Source: Chu *et al*. 2006.)

the bent-beam actuator does not exceed 600°C. If the actuator is driven in a vacuum system, the convection loss is also negligible due to very little free molecule air conduction. The average temperature of the beam is obtained by integrating the temperature profiles along the beam length and dividing by the beam length. It is numerically calculated by

$$T_{\text{average}} = \frac{\int_0^L u(x)\mathrm{d}x}{L} \qquad [9]$$

Once the temperature variation and profile in the structure are obtained, the output force and displacement can be estimated.

For a qualitative description, a calculated temperature profile of the bent-beam actuator based on eqn [9] is given in **Figure 2** as an example (Lott *et al.* 2002). The apex region of the bent beam has a relatively lower temperature, while the peak temperature actually occurs roughly at 2/3 of the distance from the anchor to the apex, which results from the significant larger heat loss through the connected beam at the apex (**Figure 2**). The descriptive temperature profile along the beam has also been measured by thermal scanning microscopy (**Figure 3**) (Chu *et al.* 2006). The length of each beam is 300 μm and the width is 7.4 μm. It shows a trend similar to the modeling results. Particularly, the larger the applied actuation power, the more similar the trend of the temperature profile to the modeling results, indicating the validation of the analytical model. It can be anticipated that if there is no additional beam connected at the apex, the peak temperature should be around this region. The thermal properties of the widely used structural materials in microactuators are summarized in **Table 2**.

2.03.2.2 Major Fundamental Thermal Actuators

2.03.2.2.1 Bimorph and multimorph

If two or more layers of dissimilar materials are grown adjacently to each other, actuation can be generated when heated since they have different coefficients of thermal expansion (CTE). These kinds of actuators are called bimorphs or multimorphs, depending on the number of the layers of the materials. A simple mathematical model can be used to describe their performance.

A bimorph or a multimorph typically consists of a cantilever with two or more layers, respectively. One layer of heating resistor is sandwiched between them (**Figure 4**). Assume that each layer is homogeneous in the X- and Y-directions. The length L of each layer is the same. When the current is passed through the heating resistor layer, a thermal stress σ_{T} will be generated between different layers due to the temperature difference ΔT_i. It is given by

$$\sigma_{\mathrm{T}} = E_i \alpha_i \Delta T_i \qquad [10]$$

where E_i and α_i denote the Young's modulus and the CTE of the ith layer, respectively. The thermal stress causes the cantilever to bend. The bending amount and direction of each cantilever depends on the CTE

Table 2 Thermal properties of some solid and polymer materials

Material	*CTE (ppm K^{-1})*	*C_p (cal $g^{-1}\,°C^{-1}$)*	*E (GPa)*	*T_M (°C)*	*κ ($cm^2\,s^{-1}$)*	*K ($W\,cm^{-1}\,°C^{-1}$)*	*ρ ($g\,cm^{-3}$)*
Si	3.0–4.2	0.168	130	1412	0.9	1.5	2.33
Polysilicon	NA	0.168	165	NA	0.185	0.34	2.32
SiO_2	0.50	0.239	57–92	1713	0.006	0.014	2.19
Si_xN_y	4.0	0.120	150	1900	0.013	~0.032	~3.0
Al	25	0.215	70	660	NA	2.37	2.7
Ni	13	0.106	206	1453	NA	0.899	8.91
Au	14.2	0.031	80	1064	NA	3.15	19.3
Cu	16.6	0.092	128	1083	NA	3.98	8.96
GaAs	6.86	0.084	85.5	1238	0.44	0.46	5.32
SU8	52	NA	4–5	NA	NA	NA	NA
		T_m (K)				Volume change (%)	
PolyNIPAAm		NA				Up to 150%	
Paraffin		~325 (with 50 carbon atoms)				10–30%	

Sources: Mutlu *et al*. (2003), Que and Gianchandani (2005), Selvaganapathy *et al*. (2003). C_p, specific heat; CTE, coefficient of thermal expansion; *E*, Young's modulus; *K*, thermal conductivity; *k*, coefficient of thermal diffusivity; ρ , density; T_m, melting temperature.

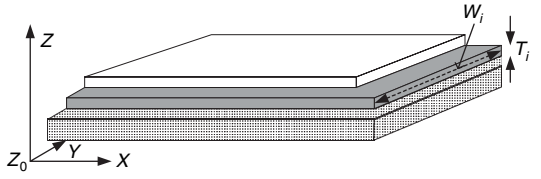


Figure 4 Schematic of bimorph with one layer sandwiched in the middle as the heating resistor.

difference, temperature difference, dimensional parameters, and mechanical properties of adjacent contacting materials.

For a bimorph actuator, the displacement of the cantilever tip at a small deflection angle is given as follows (Zhang *et al*. 1999):

$$\delta=\frac{3W_1W_2E_1E_2t_1t_2(t_1+t_2)(\alpha_2-\alpha_1)\Delta T\times L^2}{(E_1W_1t_1^2)^2+(E_2W_2t_2^2)^2+2W_1W_2E_1E_2t_1t_2(2t_1^2+3t_1t_2+2t_2^2)} \quad [11]$$

and the output force is given as follows:

$$F=\frac{3EI}{L^3}\times\frac{3\times\Delta\alpha\times\Delta T\left(1+\frac{t_1}{t_2}\right)L^2}{t_2\left\{\left(\frac{W_1E_1}{W_2E_2}\right)\left(\frac{t_1}{t_2}\right)^3+\frac{W_2}{W_1}\frac{E_2}{E_1}\left(\frac{t_2}{t_1}\right)+2\left[2\left(\frac{t_1}{t_2}\right)^2+3\left(\frac{t_1}{t_2}\right)+2\right]\right\}} \quad [12]$$

where t_i and W_i are the thickness and the width of the *i*th layer, respectively, $\Delta\alpha=\alpha_2-\alpha_1$, $\Delta T=T_2-T_1$, and *EI* is the equivalent stiffness of the bimorph cantilever, which is given by

$$EI=\frac{W_2^3t_1E_1E_2}{12(E_1t_1+E_2t_2)}K \quad [13]$$

$$K=\left[4+6\frac{t_1}{t_2}+4\left(\frac{t_1}{t_2}\right)^2+\frac{E_1}{E_2}\left(\frac{t_1}{t_2}\right)^3+\frac{E_2}{E_1}\frac{t_2}{t_1}\right] \quad [14]$$

These equations can be used as a guide to design bimorph actuators to fulfill the required performance, including desired displacement and output force. Typical displacement can be up to several hundred micrometers while the speed of the actuators as high as 1 kHz was obtained.

For a multimorph actuator, similar equations can be obtained to describe its performance and can be used to design a specific multimorph to fulfill a specific requirement.

One of the major issues for bimorph or multimorph is the bending of the actuator along the perpendicular direction of the layered structure once it is released from the substrate, resulting in an unpredictable initial position. The bending is caused by residual stress in each layer. The residual stress consists of two components: thermal stress and intrinsic stress. While the thermal stress due to CTE mismatch is the basis for thermal actuation, the intrinsic stress affects the repeatability and the reliability. The intrinsic stress results from the defects, grain size, reorganization of the structural material, etc., during fabrication process flow. Research found that the intrinsic stress can be mitigated by ion beam machining and rapid thermal annealing (RTA) (Bifano *et al*. 2002, Huang *et al*. 2006), therefore eliminating the initial bending of the released bimorph actuator.

Experiments found that bimorphs showed very nice reliability with less than ±2% hysteretic error

(Xie *et al.* 2003) and that they can be operated up to several-million-cycle actuation without observable degradation.

2.03.2.2.2 Pseudo-bimorph actuator

While bimorph and multimorph typically offer displacement and force perpendicular to the substrate, the other type of thermal actuator called the pseudo-bimorph actuator can offer displacement and force parallel to the substrate. It was first reported by Henry Guckel in the early 1990s (Guckel *et al.* 1992). This actuator is made of just one single layer of the structural material and the thermal expansion difference, resulting from different Joule heatings due to different resistances, is because of the two different shapes in the structure.

It has two types of designs. In both the cases, it has two arms connected in a U-shape. One design has two connected arms of different widths but the same lengths (**Figure 5(a)**). The other design, in contrast, has two arms of same width but different lengths (**Figure 5(b)**) (Pan and Hsu 1997). The actuator shown in **Figure 5(a)** consists of two beams of different widths. When the current is passed through the beams, the narrower beam becomes hotter than the wider one due to its larger resistance. Thus, it expands more than the wider beam. Since the two arms are joined at the free end, an arc deflection of the tip of the actuator toward the cold wider beam occurs. A deflection of several tens of micrometers and an output force of several tens of micronewtons can be generated for a 200- to 300-μm-long polysilicon actuator with power consumption of about 40 mW. The operating frequency can reach 2.5 kHz.

Assuming that the temperature difference between the hot and the cold beam is T_{HC}, the deflection of the tip is given by the following equation (Guckel *et al.* 1992):

$$\Delta X = \frac{\alpha \Delta T_{HC} L^2}{g\left(0.7707 + 0.3812\,\frac{t^2}{g^2}\right)} \qquad [15]$$

where α is CTE, g the gap between the hot and the cold beam, L the half-length of the hot and cold beams, and t the thickness of the structural material. For this structure, the deflection of the tip is independent of the Young's modulus E of the material, which is confirmed by measured data and a finite element model (Guckel *et al.* 1992).

In the design shown in **Figure 5(b)**, when the current is passed through the beams, the heat generated in the conductive long and short beams causes the temperature to change. Because of the difference between the temperature increment ΔT_L in the long beam and the temperature increment ΔT_S in the short beam ($\Delta T_L > \Delta T_S$), and because of the different beam lengths ($L_L > L_S$), the long beam has more elongation than the short beam. Therefore, the long beam curls toward the short beam.

Out-of-plane actuation to the substrate can be realized by arranging and modifying the pseudo-bimorphs properly. Two designs are schematically shown in **Figure 6** (Cowan and Bright 1997).

The vertical displacement is achieved by the thermal mismatch between the hot and the cold arm when current is applied through the beams. Experiments and modeling both found that the

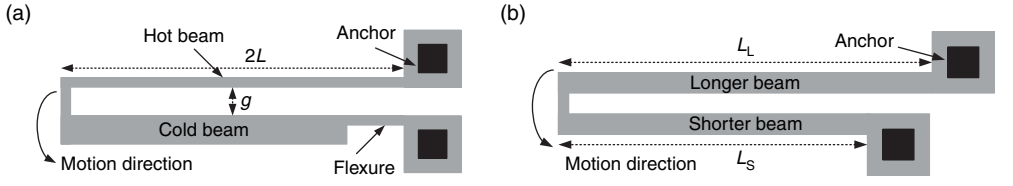


Figure 5 Schematic of (a) design I and (b) design II of pseudo-bimorph actuator.

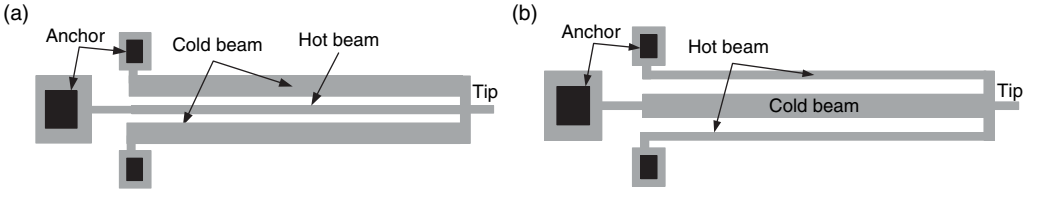


Figure 6 Two designs of a vertical electrothermal actuator: (a) hot beam in the middle and (b) cold beam in the middle.

device with the hot beam in the middle (**Figure 6(a)**) is a better design, because it has a consistent temperature difference between the hot and the cold beam and thus can maintain the repeatable up-and-down movement.

Reliability testing on loaded pseudo-bimorphs fabricated using the SUMMiT™ process developed has been performed. Some devices can be operated up to 980 million cycles at 2 kHz and still reach their full deflection at the end of the operation (Comtois *et al.* 1997).

2.03.2.2.3 Bent-beam actuator

Using just one single structural material, a class of widely used electrothermal actuators, bent-beam actuators (Cragun and Howell 1999, Que 2000, Que *et al.* 1999, 2001a, b, 2003), has been developed by Gianchandani's group in 1998 and Howell's group in 1999. It offers large in-plane, rectilinear displacement and high output force, different from the pseudo-bimorphs providing arc motion described above.

The sketch of the basic and one-level cascaded bent-beam electrothermal actuator is shown in **Figure 7**. It consists of a V-shaped beam anchored at both ends. When an electric current is passed through the beam, thermal expansion caused by Joule heating pushes the apex outward. The displacement of the apex is a function of the beam dimensions and slope, given the same actuation voltage. The force generated at the apex is linearly related to the displacement. It can easily offer peak force up to the range of several millinewtons, about 200× the forces in electrostatic comb drives of similar dimensions. Furthermore, the output force and displacement can easily be scaled by modifying the dimensions and slopes of the beams. Larger forces can be generated by placing these actuators in parallel, whereas larger displacements up to several hundred micrometers can be obtained by cascading the actuators.

The lifetime of a bent-beam actuator can extend beyond several tens of millions of cycles, depending on design and actuation temperature. The speed of the actuator has been evaluated experimentally and can be estimated by the electrothermal model previously described as well (Lott *et al.* 2002). Besides the thermal mass, the different heat transfer paths in air and vacuum also affect the temperature response time, thus the speed of the bent-beam actuator. By using the mathematical model, it has been shown that the time for the cooling cycle is much shorter in air than in vacuum, since the heat loss and conduction is much faster in the air. The theoretical predictions match reasonably well with experimental results. Typical responding time is in the order of milliseconds and can be reduced into the submilliseconds range if the thermal mass is cut or the displacement is reduced as mentioned earlier. Furthermore, as discussed in Section 2.03.2.5, the response can be enhanced and power consumption can be reduced dramatically if pulsed actuation or burst-pulsed actuation scheme is implemented. More details can be found in Que *et al.* (2004) and Shimamura *et al.* (2006).

2.03.2.2.4 Thermal responsive polymer actuator

The thermal actuators described so far can only offer either in-plane or out-of-plane actuation. In other words, they can only offer 1-directional or 2-directional actuation. Polymer-based thermal actuators are a group of actuators very attractive in microfluidics and bioapplications in that they can offer all-directional actuation since their volumes can be dynamically changed in all directions.

Polymers used for thermal actuation are usually cheap, biocompatible, and compatible with standard micromachining technology, thus good for disposable and for bioapplications. Furthermore, the most important feature of polymer-based actuation is that it can be integrated with the microfluidic lab-on-chip

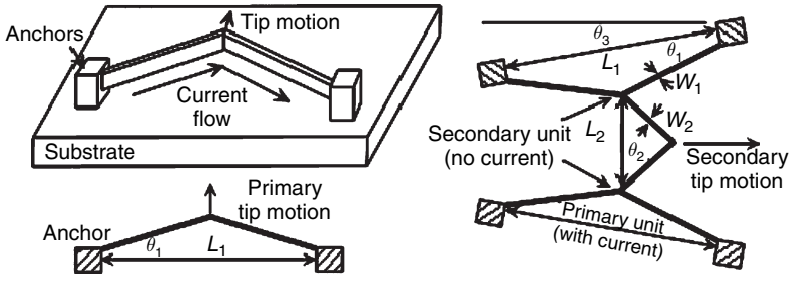


Figure 7 A basic (*left*) and one-level cascaded (*right*) bent-beam actuator.

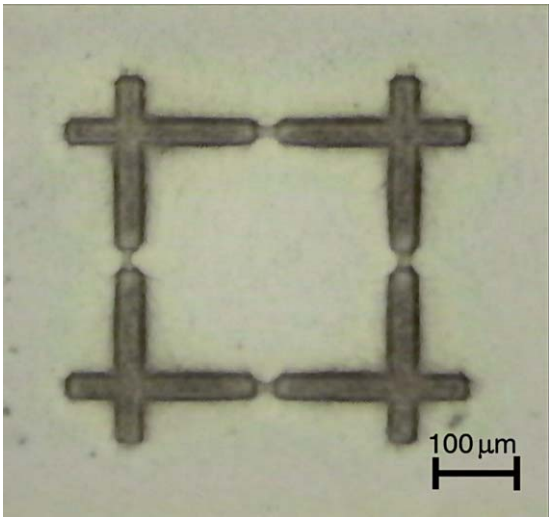


Figure 8 Top view of polymer inside water at 40°C. (Source: Mutlu S, Yu C, Svec F, Mastrangelo C, Frechet J, Gianchandani Y B 2003 A thermally responsive polymer microvalve without mechanical parts photo-patterned in a parylene channel. *Int. Conf. Solid State Sensors, Actuators and Microsystems (Transducers '03)*, Boston, MA, USA, pp. 802–5.)

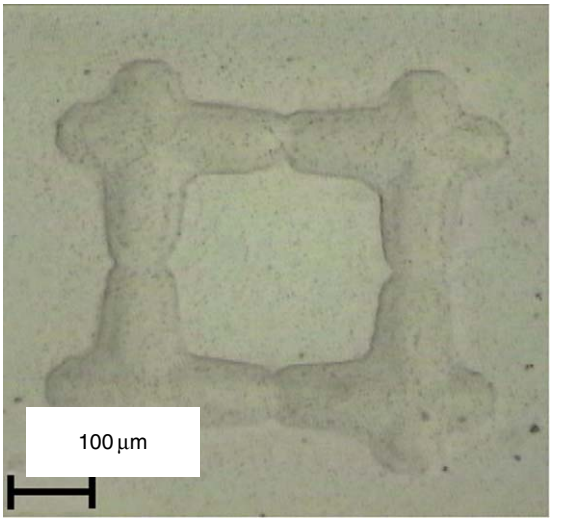


Figure 9 Same pattern inside water at 15°C. (Source: Mutlu S, Yu C, Svec F, Mastrangelo C, Frechet J, Gianchandani Y B 2003 A thermally responsive polymer microvalve without mechanical parts photo-patterned in a parylene channel. *Int. Conf. Solid State Sensors, Actuators and Microsystems (Transducers '03)*, Boston, MA, USA, pp. 802–5.)

systems much more easily than other actuation mechanisms, such as electrostatic, magnetic, piezoelectric, and bimorph actuation.

The thermally responsive polymers and microspheres can be either expandable or shrinkable with increased heating or cooling, respectively. Paraffin (Selvaganapathy *et al.* 2003), SU8 (Chronis and Lee 2005), poly(dimethylsiloxane) (PDMS) and microspore mixture (Samel *et al.* 2003) exhibit volume expansion if heated, whereas the volume of poly(*N*-isopropyl acrylamide-*co*-allyl methacrylate) (polyNIPAAm) (Mutlu *et al.* 2003) is shrinkable with increased temperature. Their major properties are summarized in **Table 2**. The volume change of the UV-patterned features of polyNIPAAm is given in **Figures 8** and **9**. PolyNIPAAm is one of the best-known temperature-sensitive polymers with a photoresist-like film characteristic. It shrinks rapidly with reversible phase transition from extended hydrated chains (swollen state) to collapsed hydrophobic coils (shrunk state) from the lower critical solution temperature (LCST) of 32°C to the temperature above LCST. Experiments found that its volume has a 2.5-fold increase as the temperature decreases from 28 °C to 10 °C. The build-up pressure, therefore, increases from ~16 to ~20 psi. The response time is about 7 s by repeatedly measuring the polymer swelling and shrinking (Multu *et al.* 2003). All the experiments for proof-of-principle were performed in water, the temperature of which can be tuned relatively easily.

2.03.2.3 Power Management and Solution

Thermal actuators are based on the heating or cooling of the structural materials; therefore, one important issue is about the power consumption and management. For any microsystem, the power consumption should ideally be minimal. This not only saves the operation power but also significantly extends its lifetime and reliability.

Several efforts have been explored to address this problem for silicon-based devices (Que *et al.* 2002, 2004). One efficient way is to reduce the thermal mass. The power P required for a certain temperature change is given in the following equation:

$$P = \frac{V^2}{R} = c \times M \times \Delta T \qquad [16]$$

where V is the actuation voltage, R the resistance of the bent beam, c the specific heat, M the thermal mass, and ΔT the temperature change.

The power consumption is proportional to the thermal mass if the temperature change is fixed. Therefore, for the same material, reduction of

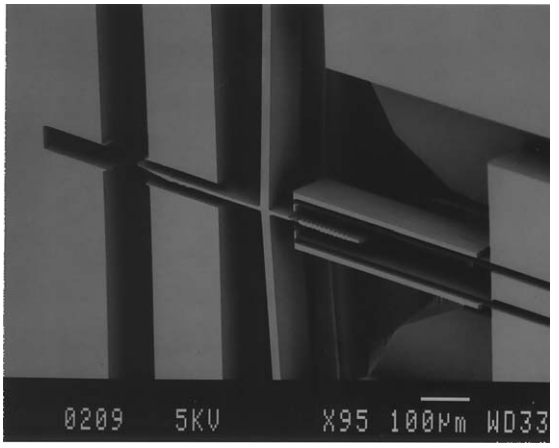


Figure 10 Scanning electron micrograph (SEM) of the digital latching micromechanics. (Source: Que *et al*. 2003.)

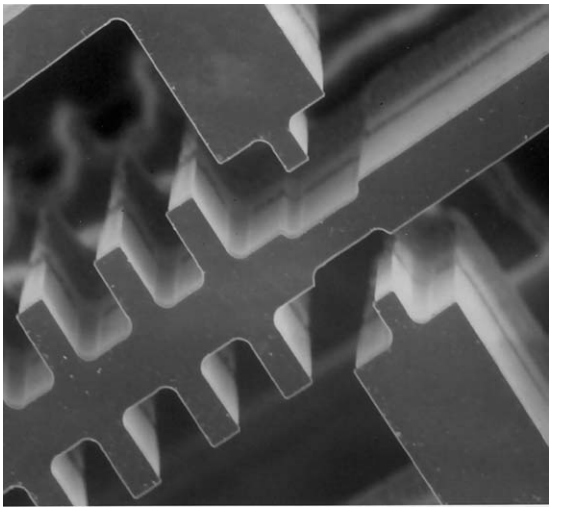


Figure 11 Close-up view of the fabricated digital latching micromechanics. (Source: Que *et al*. 2003.)

thermal mass is a way to cut the power consumption. However, the disadvantages are the potential complexity of the fabrication process and the constant power consumption still needed to maintain the position of the actuator.

The latching micromechanics (Que *et al.* 2002, 2004), in contrast, is the other efficient approach to abate the power consumption. This mechanism works for solid-state material-based actuators, such as silicon or metal actuators. Digital and analog latching are two representatives for this purpose. For both versions, the basic idea is to be able to turn off the power once the actuator has reached the target position. Thanks to the latching mechanism, the actuator, therefore, can retain its position. Under this situation, the power is only required during the actuation period, typically 1 ms or less, resulting in significant reduction in power consumption. One example for the digital latching is shown in **Figures 10** and **11** (Que *et al.* 2002). To describe this functionality briefly, the tooth-shaped structures are micromachined in the center beam, which is driven by a bent-beam actuator. The complementary tooth-shaped structures are micromachined into the pseudo-bimorph. These two tooth-shaped structures are used to engage each other to achieve digital latching. The two thermally driven bimorphs first are open, offering a path for the movement of the center beam. Once the center beam moves to the target position, the bimorphs close and the tooth-shaped structures engage to each other; thus, the center beam is locked mechanically. At this time, the power can be removed from the actuator while the center beam still maintains the position. The other digital latching mechanism, i.e., bistable latching, and analog latching mechanism will be described in Section 2.03.3.

2.03.2.4 Thermal Isolation Strategies

Thermal isolation is very important for the thermal actuator in a microsystem. The heat flow can propagate from the actuator to other components due to temperature gradient if no proper thermal isolation is implemented, resulting in the degradation of the performance of the system or even damage in the samples to be manipulated or under test. In order to isolate the heat flowing from the actuator to the other section in the microsystem, several approaches have been proposed. The basic principle for all approaches is the same, whereas different fabrication processes and materials are utilized.

One approach is to design a trench in the region that mechanically connects the actuator and the other components. The trench is formed by deep reactive ion etching (RIE) and later is filled by materials with a low thermal conduction coefficient, such as silicon dioxide or silicon nitride (Que *et al.* 2001a, b). The fabricated structure is shown in **Figure 12**. The other approach deposits a silicon nitride layer, which is used to connect two separated polysilicon beams in **Figure 13**; thus, the heat flow can be minimized from the actuator to the radio frequency (RF) switch (Wang *et al.* 2002). For both cases, the basic idea is to use some materials with low thermal

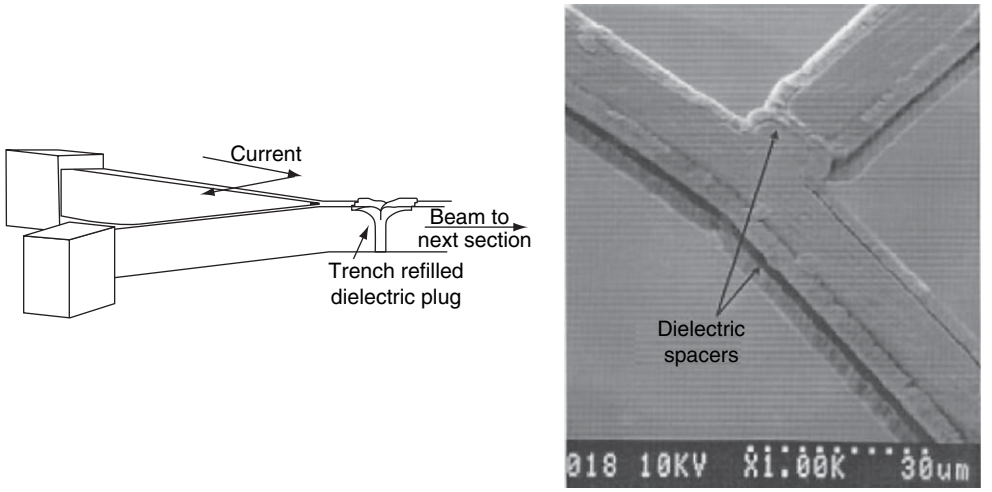


Figure 12 Trench refilled with thermal isolation material silicon dioxide.

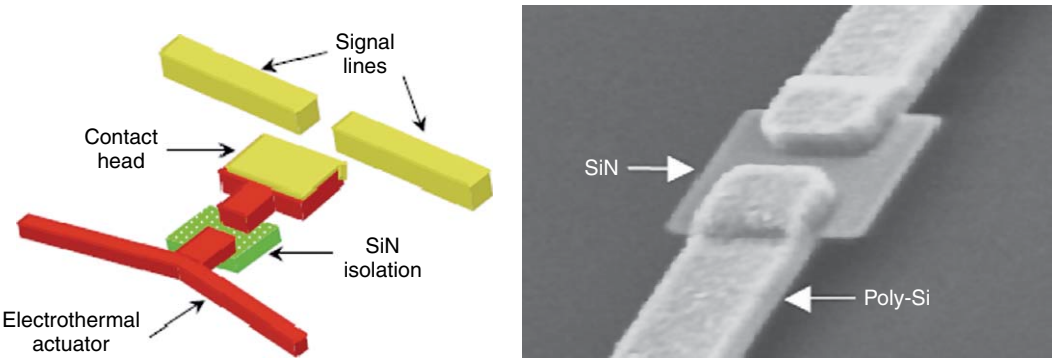


Figure 13 Silicon nitride (SiN) layer connecting polysilicon layer for thermal isolation. (Source: Wang Y, Li Z, McCormik D, Tien N C 2002 Low-voltage lateral-contact microrelays for RF applications. *15th IEEE Int. Conf. Micro-electro-mechanical Systems*, Las Vegas, NV, USA, pp. 645–8.)

conduction to reduce the heat flow, thereby maintaining reasonably low temperature of other parts in the microsystem.

For microfluidics and bioapplications, thermal isolation or mitigation is extremely critical. A scheme using polyimide as an isolation layer for a microvalve is proposed. Its basic idea is similar to the two aforementioned approaches. In contrast, polyimide is used as the thermal isolation material in this case. The sketch of the silicon bimorph and scanning electron micrograph (SEM) of the fabricated device are shown in **Figure 14**. The measurement shows the temperature distribution of the actuator (112°C) and at the thermally isolated region (33°C), indicating excellent thermal isolation between them. Using this approach, a thermally isolated microvalve has been demonstrated (Kawada *et al.* 2003).

2.03.2.5 Response Speed

The heat transfer process directly affects the speed of the thermal actuators, which depends on how fast the resistor can be heated to a target temperature in the heating cycle and how fast the heat can be dissipated in the cooling cycle. The heating process is at the rate of energy transfer between the electron gas and the lattice vibration, and thus quite fast. However, the cooling procedure is relatively slower. Generally, the speed of thermal actuators is limited by this cycle. Therefore, its speed can be potentially increased by integrating heat sink components or cutting the thermal mass.

For both cycles, the effective thermal mass m_{thermal} of the actuator is the major factor in response speed (Lott *et al.* 2002). A bent-beam actuator is used as an example to estimate the speed response of the thermal actuator. Eqns [1]–[8] can be rewritten as the

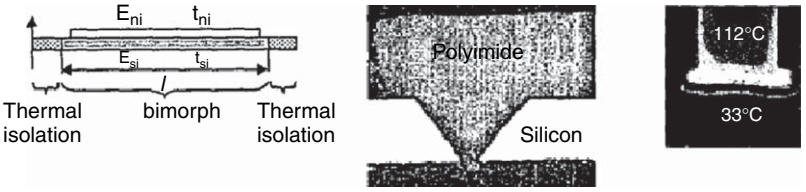


Figure 14 Sketch of the bimorph with polyimide as isolation layer (*left*); scanning electron micrograph (SEM) of the device showing the thermal isolation region (*middle*); and temperature distribution (*right*). (Source: Kawada H, Yoshida H, Kamakura M, Yoshida K, Saitou M, Kawahito K, Tomonari S 2003 Thermally driven microactuator containing thermal isolation structure with polyimide and its application to microvalve. *12th Int. Conf. Solid State Sensors, Actuators and Microsystems (Transducers '03)*, Boston, MA, USA, pp. 1935–8.)

following explicit finite difference model expression. The temperature $T_i(j+1)$ at time $j+1$, therefore, can be obtained using temperatures $T_i(j)$, $T_{i-1}(j)$, $T_{i+1}(j)$ at time j, the current input $\mathcal{J}(j)$, and boundary conditions, where i refers the ith differential element of the bent beam.

$$T_i(j+1) = \frac{\Delta t}{\rho c \Delta V_i}\left[\mathcal{J}^2 \rho_r \Delta V_i + \frac{SA_u}{G_u}(T_{sub} - T_i(j)) + \varepsilon\sigma A_s(T_{sur}^4 - T_i^4(j)) + \frac{k_p A_x}{\Delta x}(T_{i-1}(j) - T_i(j)) + \frac{k_p A_x}{\Delta x}(T_{i+1}(j) - T_i(j))\right] \quad [17]$$

where Δt is the time increment, ρ and c the density and specific heat of the structural material, ΔV_i the element volume, ρ_r the material resistivity, S the shape factor, A_u surface area of the element facing the substrate, G_u the effective thermal conductivity of the element, T_{sub} and T_{sur} the temperature of the substrate and the surroundings, A_s and A_x the area of the exposed surface of the element and cross-sectional area of the element, Δx the length of the element in the direction of heat transfer, k_p the thermal conductivity of the structural material, ε and σ the emissivity of the structural material and the Stefan–Boltzman constant.

The transient temperature response can be calculated using this formula (**Figure 15**). The thermal mass (ρ, c) affects the response.

Pulsed actuation has been explored to improve the transient response and power consumption both theoretically and experimentally by different groups (Que *et al.* 2004, Shimamura *et al.* 2006). It turns out that the response time can be improved while the power consumption is significantly reduced. Furthermore, burst pulse (a series of short pulses) actuation can further improve the efficiency of the actuators. Typically, the response time can be around 0.2 ms compared with several milliseconds for constant current actuation while the required energy is reduced from 1000 to 200 μJ.

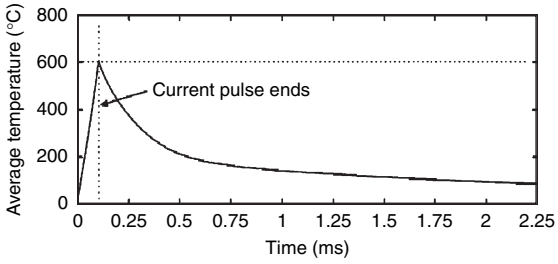


Figure 15 Calculated temperature response of the bent-beam actuator. (Source: Lott C, Mclain T, Harb J, Howell L 2002 Modeling the thermal behaviour of a surface-micromachined linear-displacement thermomechanical microactuator. *Sensors and Actuators A* **101**, 239–50.)

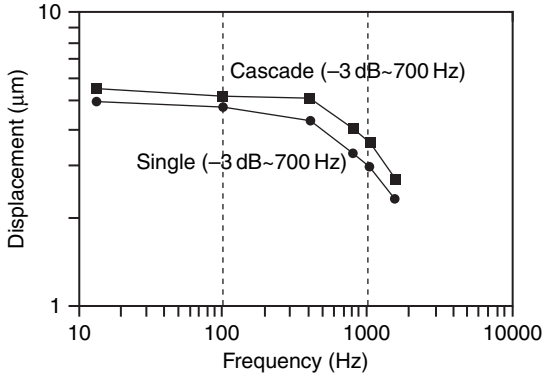


Figure 16 A typical measured response of the bent-beam actuator. (Source: Que *et al*. 2001b.)

One typical measurement result in **Figure 16** shows roughly 1.5 ms response time under constant current actuation (Que *et al.* 2002).

An additional factor on the response speed is the targeted displacement. The smaller the targeted displacement, the lower the heating power required reaching the target temperature, thus the faster the cooling cycle. Experiments found that for pseudo-bimorph, it can be operated as fast as 13 kHz if the displacement is about 2 μm (Comtois *et al.* 1997). It is also true for bent-beam actuators and other thermal actuators.

2.03.2.6 Materials

Theoretically, all materials expandable or shrinkable with increased temperature can be potentially used as the structural material for the thermal actuators. Thus, the material can be gas, liquid, or solid. Widely used materials include doped silicon or polysilicon, electroplated metals, i.e., nickel or copper, and polymers such as SU8, PDMS composite, and Parylene. Recently, some advanced nanoscale materials, such as carbon nanotubes (CNTs), have also been exploited as structural material for optothermal actuation (Lu and Panchapakesan 2005).

While silicon and some metals form the foundation mostly for mechanical microelectromechanical systems (MEMS) applications, the availability of polymers for thermal actuation paves the way for bioMEMS and microfluidics in biological and medical applications. Since these materials are biocompatible and most can be micromachined using a standard microfabrication process, they have been successfully used to construct some major microfluidic components as well as various integrated microfluidic systems. For bioMEMS devices and implantable devices, low voltage is essential for physiologic compatibility; therefore, the thermal actuation has been utilized.

2.03.2.7 Repeatability and Reliability

Repeatability and reliability (lifetime) are two important factors evaluating the performance of the microactuators. The repeatability is that the actuator can always offer the same outputs at the same driving condition at different times. More specifically, the importance of the repeatability is to ensure that the same displacement, volume expansion, or force can be achieved for targeted functionalities; thus, the output of the integrated microsystem can be predictable and dependable, which is the key for maintaining the performance of the integrated microsystem. The hysteresis, for example, of the displacement and driving voltage is definitely not good, and therefore, it should be eliminated or mitigated for robust actuator design.

While repeatability offers the spatial resolution of the actuators, the reliability gives the temporal dependence. The reliability (lifetime) is how long the actuator can maintain the same performance without degradation under continuous actuation. Lifetime is determined by several mechanisms: stress distribution in the actuator, structural material properties, microdefects, microcracks and their propagation in the whole device, etc.

The repeatability and reliability test of bent-beam actuators is used as an illustrative example. The repeatability of bent-beam actuators has been evaluated by measuring the displacements while the driving voltages increase and then decrease in a specific range (**Figure 17**) (Que and Gianchandani 2005). These devices are fabricated using silicon-on-insulator (SOI) wafers; the actuators are micromachined in the device layer, which is single-crystal silicon. Measurements find that as long as the driving voltages are below the upper limit of the voltage, the actuator shows very good repeatability (less than $\pm 1\%$ hysteretic error). Note that at the upper limit of the actuation voltage, too much heating would be generated. Therefore, significant thermal stress and self-annealing would be generated in the device, resulting in inelastic deformation of the actuator. On the contrary, in terms of reliability, the displacement of the actuators has been measured under repeated square-wave-voltage actuation, which shows no degradation after several million actuation cycles for cascaded devices if the actuation temperature is far below the melting temperature at 1412°C (**Figure 18**). In contrast, for the boron-doped (p^{++} Si)

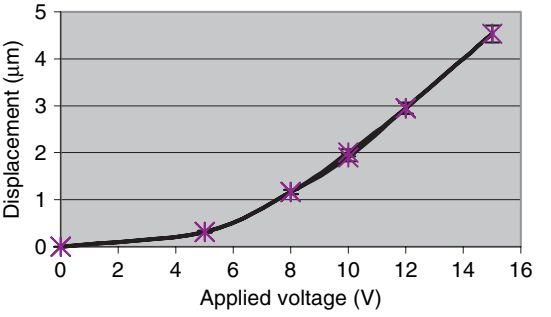


Figure 17 Repeatability test of bent-beam actuator.

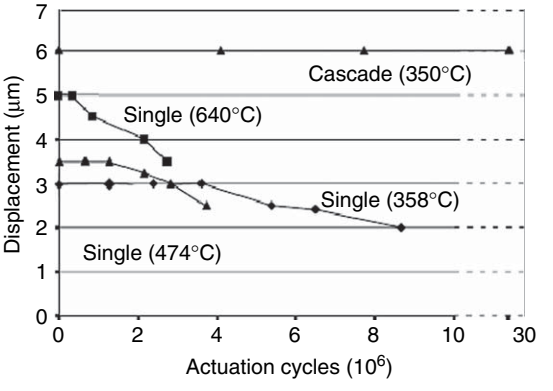


Figure 18 Lifetime test of bent-beam actuator.

devices, microcracks inside the devices and their propagation degraded their performance, SEM and atomic force microscopic (AFM) studies have confirmed this kind of fatigue (Chu *et al.* 2006). For polysilicon devices, thorough measurements have also been performed. It turns out that the micrograins of polysilicon anneal and grow during thermal actuation (**Figure 19**). Changes in the grain structure can affect many of the mechanical, electrical, and thermal properties of the material, and thereby the performance of the actuators.

Other factors, such as plastic deformation, dopant diffusion, and redistribution in the actuator, can also effect its lifetime (Chu *et al.* 2006).

For other thermal actuators, their repeatability and reliability can also be evaluated similarly. Note that for polymer-based devices, their chemical compositions and structures also play an essential part in their performance. Therefore, the synthesis and preparation of the polymer before the micromachining could be very critical.

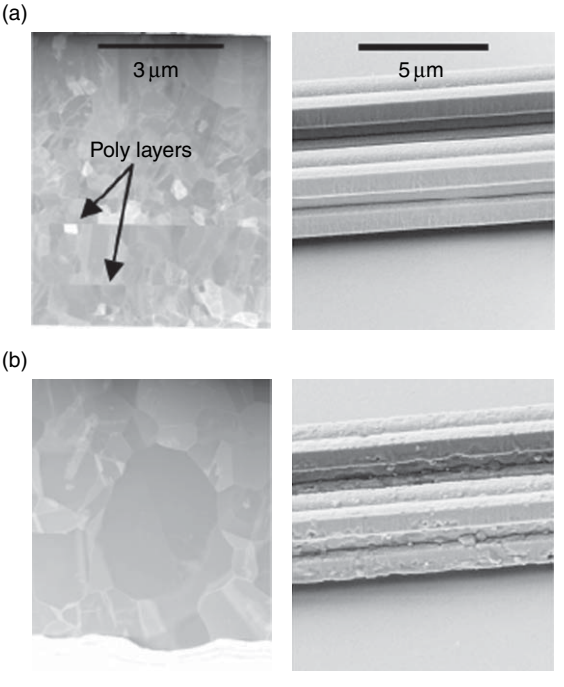


Figure 19 Transmission electron microscopic (TEM) image and scanning electron micrograph (SEM) of bent beams subjected to different levels of actuation power: (a) an unheated section of bent beam and (b) heated after 30 million cycles at 43 mW. (Source: Chu *et al*. 2006.)

2.03.3 Applications

2.03.3.1 Microoptics and Biophotonics

In the field of microoptics, microbiophotonics, and bioimaging, more and more optical elements are required to be integrated in order to form a compact, cost-effective, and high-performance microsystem (Que and Gianchandani 2005). Microoptical scanning functionality plays a very important role and could be essential for some medical diagnostic instrumentations, usually enabled by the micromachined mirrors. For optical communication, dynamic optical power management and configuration of the network also requires reconfigurable optical components, such as variable optical attenuators (VOAs) or optical switches. In addition, for some other specific applications such as dynamic optical pulse shaping for ultrafast optics, one approach is to dynamically change the refractive index or the pitch of the fiber Bragg gratings (FBGs). Therefore, periodic high-force array (typically in the range of millinewtons) offered by arrayed microactuators on FBGs is needed to modify the refractive index of the fiber core. In this case, thermal actuation, which offers a reasonably fast speed (up to kilohertz range), is a good option. In the following, some applications of thermal actuation in biophotonics, microoptics and optical communication will be illustrated and discussed.

Microscanner-enabled OCT. The microscanner has been developed for various applications, from project display for high-definition TV, to maskless lithography for microarrays for genomics (Singh-Gasson *et al.* 1999), to optical coherence tomography (OCT) for biomedical imaging. The actuation mechanism can be electrostatic, thermal, piezoelectric, magnetic, SMA, etc. Thermal actuation-based microscanner typically utilizes the bimorph actuator. One type of MEMS scanning mirror for endoscopic OCT (EOCT) is shown in **Figure 20**, fabricated using complementary metal oxide semiconductor-MEMS (CMOS-MEMS) process (Xie *et al.* 2003).

OCT is an emerging imaging technique that offers cross-sectional images of biological tissue and can offer noninvasive optical biopsy and delineate the tissue micromorphology. EOCT utilizes a mirror to scan the endoscopic laser. Its detailed functionalities can be found in many citations (Xie *et al.* 2003). One of its major components is the scanning mirror. Compared with a macroscale mirror, a micromirror is an ideal option due to its compact footprint and can conveniently fit into the slender endoscopes. The

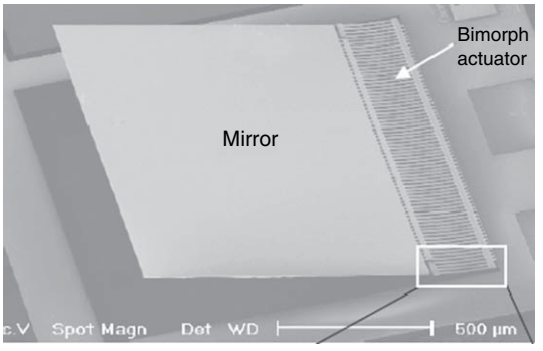


Figure 20 A scanning mirror with bimorph actuator. (Source: Xie T, Xie H, Fedder G K, Pan Y 2003 Endoscopic optical coherence tomography with new MEMS mirror. *Electron. Lett.* **39(21)**, 1535–6.)

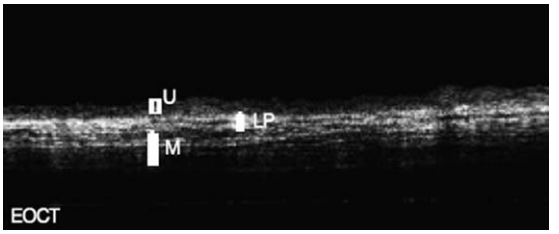


Figure 21 Optical coherence tomographic (OCT) image of a rabbit bladder. EOCT, endoscopic optical coherence tomography. (U = urothelium, M = muscular layer, LP = lamina propria). (Source: Xie T, Xie H, Fedder G K, Pan Y 2003 Endoscopic optical coherence tomography with new MEMS mirror. *Electron. Lett.* **39(21)**, 1535–6.)

MEMS thermally actuated micromirror-enabled EOCT can offer *in vivo* imaging diagnosis of early bladder cancers with good quality as shown in Figure 21.

Latching VOA and switch. Optical power management and network configuration are very important issues for broadband optical communication. The uniformity of optical power in the optical network is desirable to maintain the signal-to-noise ratio (SNR) and convenient for the network design. Dynamic power equalizer is utilized for this and can be built on the VOA array. Similarly, in order to dynamically reconfigure the optical network, programmable add/drop optical channel is essential. Optical switch plays a very important role in this, and tremendous work has been performed to develop MEMS optical switches or arrays. For both optical components, thermal actuators-based MEMS devices proposed by Que *et al.* (2002) have been developed to fulfill the demands.

In order to minimize the thermal power consumption, analog and digital latching mechanisms have

Figure 22 Analog latching variable optical attenuator (VOA). (Source: Que *et al*. 2003.)

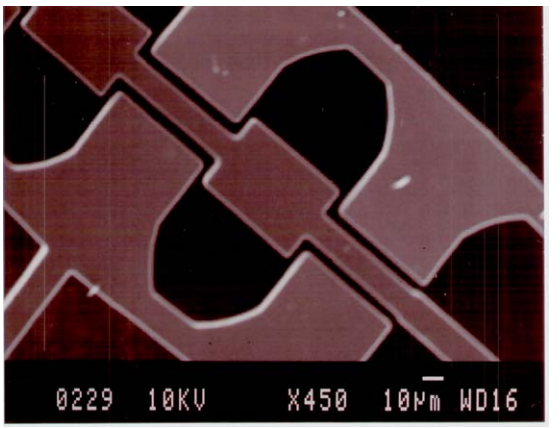


Figure 23 Close-up view of the analog latch region. (Source: Que *et al*. 2003.)

been designed and integrated in the VOAs and switches. The digital latching mechanism has been described previously. For the analog latching mechanism, the friction force between the holding structures and the central beam is the force offered to balance the mechanical restoring force of the central beam (Figures 22 and 23). Since the friction force can be generated continuously along the movable direction of the central beam, continuous latching can be realized. To achieve this, two control circuits for latching structure and central beam, respectively, are required. They should have certain phase difference to ensure proper operation. Typical response time of the VOA is 0.5 ms; power consumption is 200 mW with optimization. Once it is latched, power consumption is zero.

Pulse shaper for ultrafast optics. Ultrashort pulse shaping is used for coherent control of chemical reactions, nonlinear optics, and dispersion compensation in optical communication. Traditional approach

typically utilizes a pair of diffraction gratings and lenses to separate different spectra, after passing spatial light processor and then recombine them by the other pair of diffraction gratings and lenses to achieve the pulse shaping. The major drawbacks are its table-sized setup, tedious optical alignment, and many cautions to avoid dispersion effects through the lenses. A pulse shaper microchip, therefore, has been developed to address these issues (Udeshi *et al.* 2006).

This microchip consists of arrayed bent-beam actuators and chirped fiber Bragg gratings (CFBGs) (**Figure 24**). The principle of FBGs is Bragg reflection. When light propagates through the FBGs, it experiences periodic alternating regions of higher and lower refractive index, with a portion of the light being reflected at each interface. If the spacing between the interfaces satisfies the Bragg condition, the total reflection can be up to 100% for a particular wavelength. In order to reflect a spectrum of wavelengths rather than just a single wavelength, CFBGs are introduced. In CFBGs, the spacing between the interfaces linearly varies along the fiber with a chirped profile. For pulse shaping, the refractive index or the grating pitch should be changed. The integrated actuator array offers periodic localized forces up to millinewtons, thus compressive strain on the CFBGs, resulting in local changes of the refractive index of the fiber core along the fiber, thus the shaping of the reflected spectrum. By altering the applied force, the optical spectrum and the shape of the reflected pulse can be changed dynamically. Measurements found that the pulse width can be modulated from 1.05 to 4 ps with 500 mW power on the actuators (**Figure 25**).

2.03.3.2 Biomedical Microdevices and Microfluidics

In the field of the genomics and proteomics, miniaturized systems or components to manipulate biological samples and monitor the bioreactions are becoming very important and challenging. For instance, many different strategies have been explored for cell positioning, sorting, and separation, resulting in a very active research field. High-precision protein detection has become very important in the field of diagnosis and prognosis (Rickert *et al.* 1999). Previous efforts included resonant (Zhang and Kim 2003), electrochemical (Green 1987), and optical (Ekins 1998, Homola *et al.* 1999) methods. A method is proposed by exploiting the rigidity changing point of a nanomechanical system (Lee and Cho 2004), offering higher precision than the resonant method, and providing a simple and inexpensive approach without any labeling process and optical detection. In neural science and technology, MEMS technology also plays a very important role. Implantable neural microdevices are important for *in vivo* research and monitoring. One example is to monitor single- and multineuronal action using arrays of microelectrodes, which is the key to

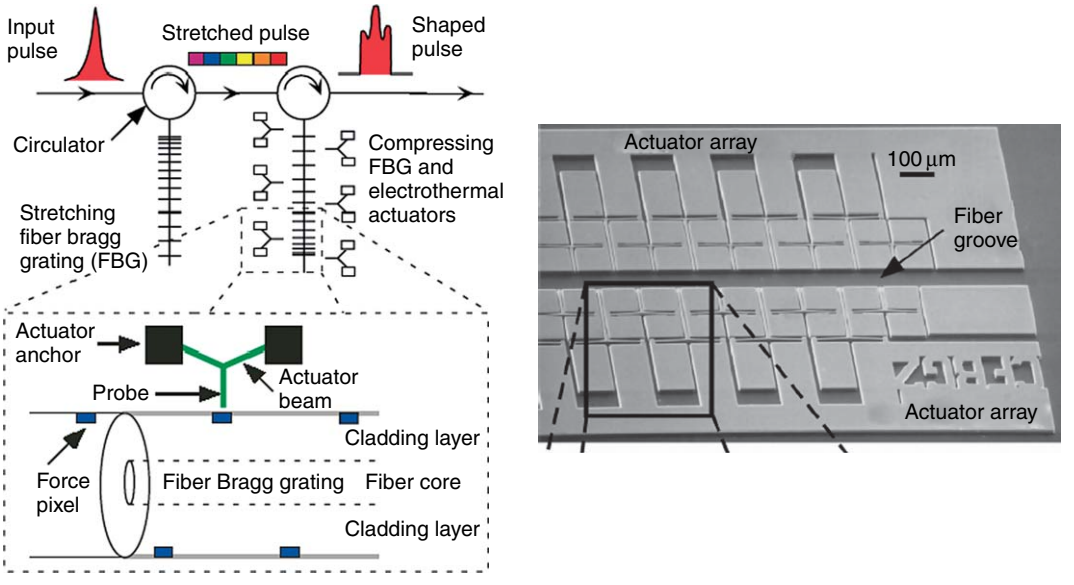


Figure 24 Sketch of the chip (*left*) and fabricated chip (*right*). (Source: Udeshi *et al*. 2006.)

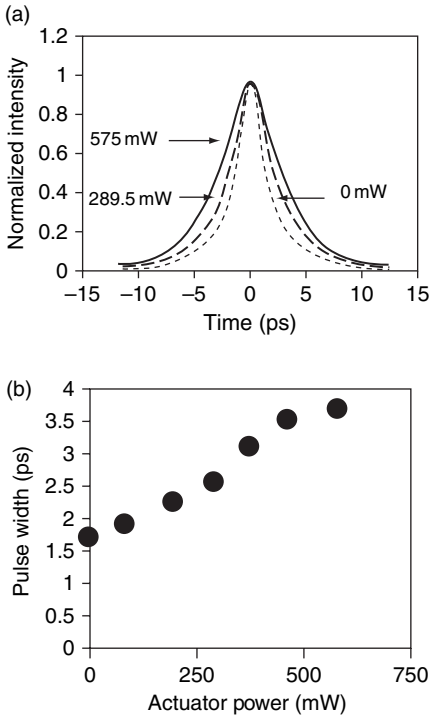


Figure 25 (a) Pulse shaping results and (b) pulse width and actuator power relationship. (Source: Udeshi *et al*. 2006.)

understanding their function, dysfunction, and restoration. In addition, one of the very important directions of bioMEMS technology is the miniaturized drug delivery technology (Reed and Lye 2004), which has been pursued extensively. The micromachined drug delivery devices can provide very accurate doses at nanoliter level, offer painless pinch, and usually are very cheap and thus could be disposable. Microscale hollow needles, liquid dispensers, and micropumps are typically the foundations of the devices in this category. Furthermore, in order to perform complete biochemical analysis at a much more cost-effective and efficient way, and achieve real point-of-care medical applications, emerging microfluidic systems definitely is an ideal technical platform. Many actuation schemes have been applied and adapted for the aforementioned biomedical applications. The thermal actuation becomes intriguing since it has a low actuation voltage, which is essential for most bioapplications, and it offers many attractive features, such as a high force and a small footprint, compared with other actuation mechanisms. In the following, some biomedical microdevices or microsystems based on thermal actuation will be described.

Cell manipulation. Biological samples manipulation in solution is important for genomics and proteomics. Optical tweezers (Ashkin 1997) or optoelectronic tweezers (Chiou *et al.* 2005) are powerful and non-invasive biomanipulation tools. However, they are bulky and expensive approaches. A micromachined SU8-based gripper has been proposed as a mechanical tool to grasp and transport the microscale biosamples to the desired location (Chronis and Lee 2005). Compared with other manipulation approaches (Kim *et al.* 1992), any optical or electrical interaction or interference with biosamples can be avoided. It has been demonstrated and can be operated in ionic environment; therefore, it can be used for single-cell manipulation and positioning, cell isolation, etc.

This SU8 microgripper consists of two pseudo-bimorphs mirror-imaged to each other (**Figure 26**). Its hot and cold arms are made from a thick SU8 layer with an embedded Cr/Au layer as the microheater and they are anchored to the silicon substrate. The microgripper is normally open. Gripping capability is achieved by applying the driving current. Experiments have demonstrated its ability to move a single Hela cell in solution (**Figure 27**). The high CTE of SU8 (52 ppm C^{-1}) is critical since the operation temperature changes can be maintained at less than 32°C and driving voltage can be kept below the electrolysis point. In addition, the high force offered by thermal actuation makes it feasible to operate in the liquid environment with big friction force, opening the door for any type of bioassay requiring sample manipulation.

Label-free biosensing. Label-free biosensing is advantageous when compared with labeled biosensing because the fluorophore is not required, and therefore, the properties of the biosamples cannot

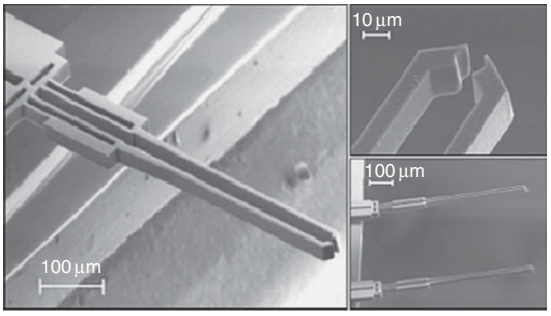


Figure 26 Fabricated microgripper using pseudo-bimorphs. (Source: Chronis N, Lee L 2005 Electrothermally activated SU-8 microgripper for single cell manipulation in solution. *IEEE/ASME J. Microelectromech. Syst*. **14**, 857–63.)

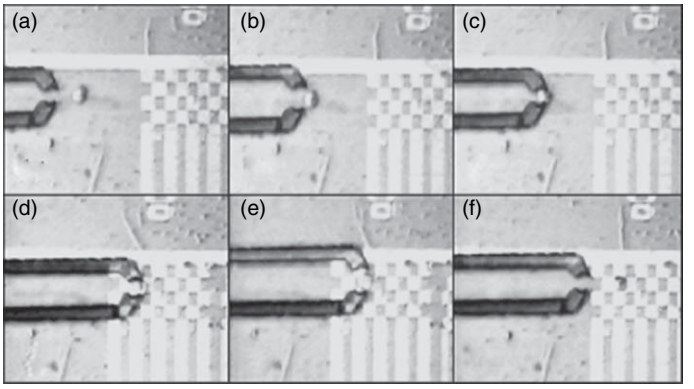


Figure 27 Optical micrographs showing the manipulation of the cell. (Source: Chronis N, Lee L 2005 Electrothermally activated SU-8 microgripper for single cell manipulation in solution. *IEEE/ASME J. Microelectromech. Syst*. **14**, 857–63.)

be perturbed or even modified. Many ways have been exploited for label-free biodetection, such as surface plasmon resonance (SPR), surface-enhanced Raman signal (SERS), and localized surface plasmon resonance (LSPR). Recently the mechanical properties of biomaterials have been exploited for label-free biosensing. By utilizing electrothermal nanogap actuators, the binding of biotin (the receptor) with streptavidin (target protein) can be monitored by the coordinate shift of the rigidity changing points (Lee and Cho 2004). The basic idea is illustrated in **Figure 29**. The device consists of two parallel plates with a nanoscale gap. One is attached to an actuator, and the other is fixed and anchored to the substrate. The receptors are immobilized on the gold-coated surfaces of the plates. The simplified model is given in **Figure 30**. The actuator pushes the push bar with a spring constant k_1. When the push bar touches receptors or the target proteins, the added spring constant k_{rp} changes the effective spring constant of the whole system, resulting in a coordinate shift of the rigidity. Therefore, the presence of the target proteins can be detected since they reduce the free distance of the push bar.

The schematic of the device is illustrated in **Figure 28** and the fabricated device is shown in **Figure 29(a)**. The sidewall surfaces of the nanogap plate, the protein sensing area, are coated with gold for the immobilization of the biotin. The fluorescence image of the nanogap region (**Figure 29(b)**) indicates the presence of the fluorescein isothiocyanate (FITC)-tagged streptavidin, suggesting the binding between the streptavidin and the biotin. The measured coordinate shift of the rigidity changing point (**Figure 29(b)**) is 14.0 ± 7.4 nm before and after the binding between the receptors and the target proteins, which is very close to the size of the streptavidin which is about 10 nm. The nanogap

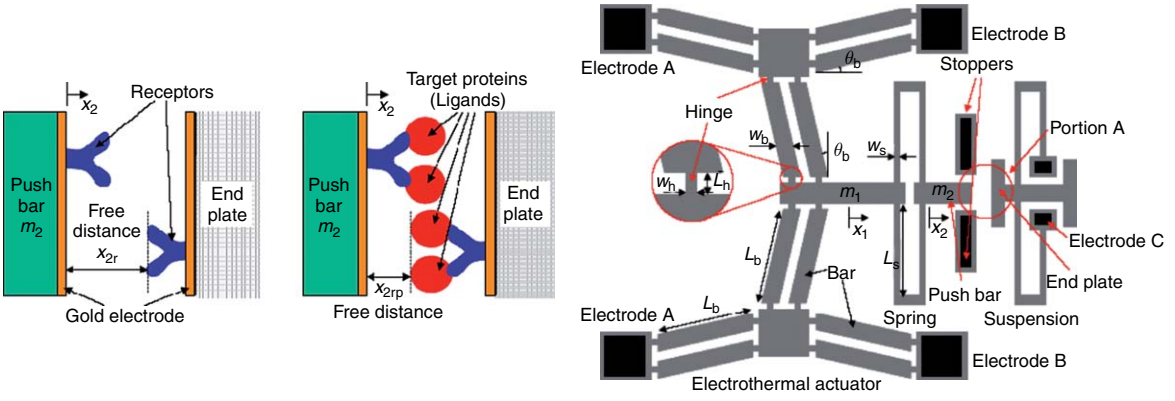


Figure 28 The basic principle of the protein sensor (*left*) and sketch of the whole device (*right*). (Source: Lee W, Cho Y H 2004 Nanomechanical protein detectors using electrothermal nano-gap actuators. *Proc. Int. Conf. MEMS*, Maastricht, The Netherlands, pp. 629–32.)

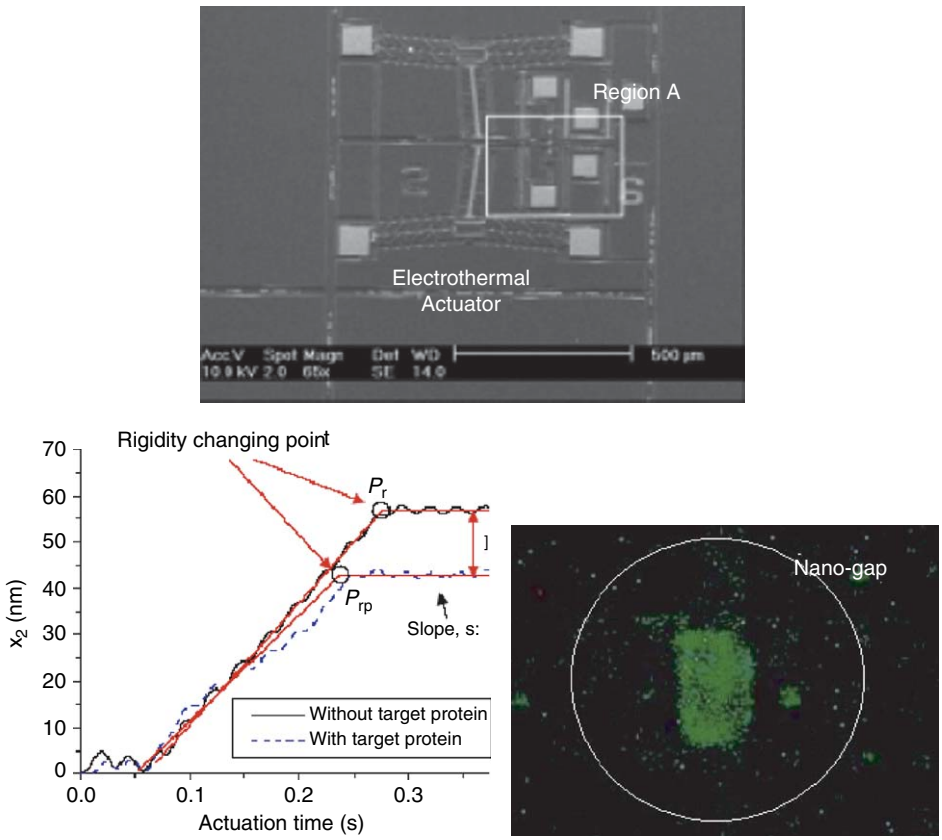


Figure 29 (a) Scanning electron micrograph (SEM) of the fabricated device and (b) measured rigidity changing point and fluorescence image in the nanogap. (Source: Lee W, Cho Y H 2004 Nanomechanical protein detectors using electrothermal nano-gap actuators. *Proc. Int. Conf. MEMS*, Maastricht, The Netherlands, pp. 629–32.)

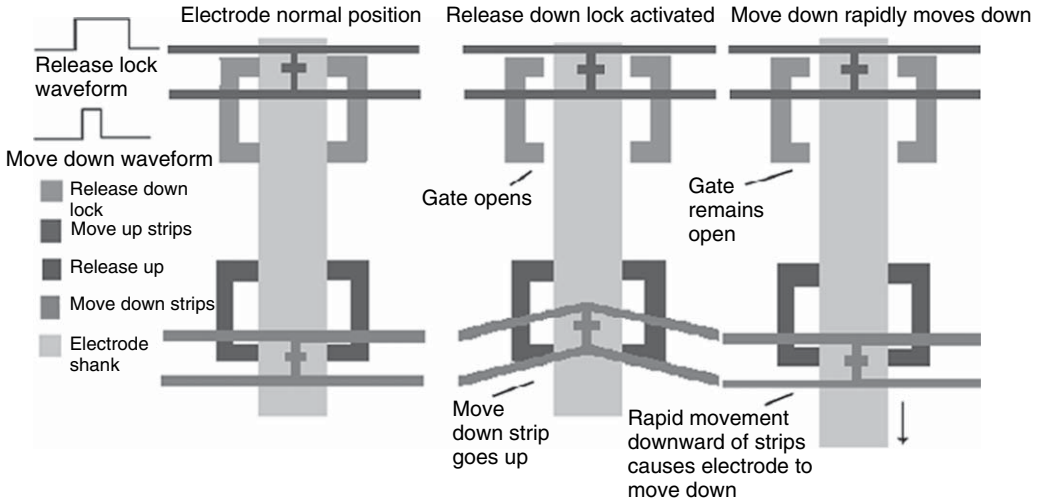


Figure 30 Basic principle of the neuron-monitoring device. (Source: Muthuswamy J, Okandan M, Gilletti A, Baker M S, Jain T 2005 An array of microactuated microelectrodes for monitoring single-neuronal activity in rodents. *IEEE Trans. Biomed. Eng.* **52(8)**, 1470–7.)

biosensing area thus offers an enabling technology and technical platform for high-accuracy protein detection.

Single-neuronal activity monitoring. The electrical activity of single neurons is the critical information to understand the neuronal function, dysfunction, and restoration. Micromachined multichannel fixed microelectrode array has been used to monitor a population of neurons simultaneously (BeMent *et al.* 1986, Wise *et al.* 1970). However, with this type of electrodes, each individual electrode cannot be positioned in the array and thus the number of neurons that need to be monitored cannot be maximized. Furthermore, the inability to position and reposition the electrodes makes this type of device undesirable since electrical contact with neurons could be lost due to either a gradual drift of microelectrodes or micromotion of brain tissue during long-term *in vivo* measurements when implanted in the brain of the animal.

The proposed microactuated microelectrodes array addresses the aforementioned issues (Muthuswamy *et al.* 2005). The basic principle is illustrated in **Figure 30**. In this design, the bent-beam actuators are used and coupled to a ratcheting system that drives a center shuttle, part of the microelectrode, up or down. The locking mechanism is similar to that of the digital latching mechanism described in Section 2.03.2. When microelectrode is to be pulled up, the ratchet pawls for the down-drive are disengaged from the shuttle, permitting its free motion. At the end of each actuation cycle, both sets of pawls are in place, therefore preventing the microelectrode from moving in either direction. A similar procedure is required when the microelectrode is to be pulled down. The motion resolution of the microelectrode is determined by the stroke of the actuator and the gap between the ratchet teeth on the shuttle. In summary, the basic sequence of the position and reposition of the microelectrodes includes unlocking, actuation in both directions, and locking of the structure, which is enabled by the use of bent-beam actuators attached to a ratchet system.

The bent-beam actuator offers several advantages over electrostatic actuated microelectrode devices. In addition to smaller footprint and hence higher packing density, it is much less susceptible to mechanical vibration and humidity variations. Much higher force and thus easier penetration of microelectrode in the rodent brain enabled by thermal actuators are also attractive features, when the microelectrodes array chip is implanted for chronic *in vivo* experiments.

The fabricated prototype device, neural probe chip (Muthuswamy *et al.* 2005), is shown in

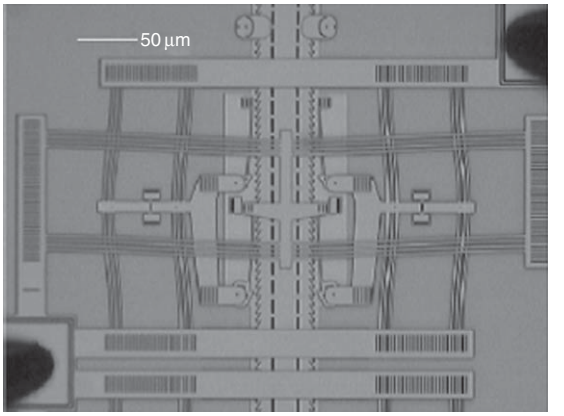


Figure 31 Scanning electron micrograph (SEM) of fabricated device showing microelectrodes with integrated actuators. (Source: Muthuswamy J, Okandan M, Gilletti A, Baker M S, Jain T 2005 An array of microactuated microelectrodes for monitoring single-neuronal activity in rodents. *IEEE Trans. Biomed. Eng.* **52(8)**, 1470–7.)

Figure 31. It has been used successfully to precisely position and reposition microelectrodes in the brain after implantation. Furthermore, it can be potentially used for long-term *in vivo* monitoring of single neurons and for sequential monitoring of multineurons with a single-implant microelectrode.

Micropump, microvalve, liquid dispenser, and thermal ink-jet print heads. Micropumps and microvalves are major components for functional integrated microfluidic networks. Different mechanisms and materials have been explored and implemented for these components. Thermal responsive polymers are attractive since their volumes are changeable, subject to the variation of the temperature. Hence, micropumps, microvalves, and dispensers with no movable parts can be made, resulting in their potential excellent reliability and long-time operation. Furthermore, the utilization of this category of polymer offers the possibility and capability for the integration of large-scale microfluidic systems.

Paraffin, polyNIPAAm, and PDMS mixtures are among those widely used polymers for these applications. In-line microvalves in **Figures 32** and **33** made by paraffin and polyNIPAAm have shown excellent performance (Mutlu *et al.* 2003, Selvaganapathy *et al.* 2003). Both of them are compatible with standard surface micromachining technology. Silicone elastomer composite, consisting of PDMS and expandable microspheres, has been employed to construct single-use drug delivery microdevices with integrated

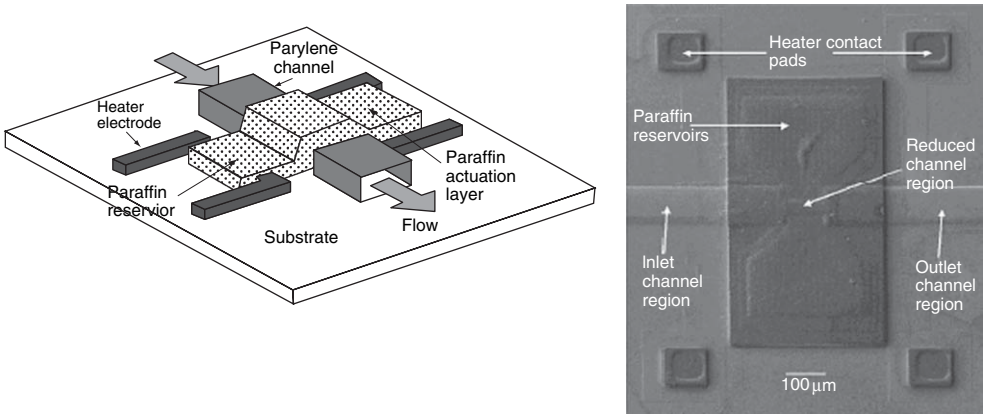


Figure 32 Schematic of (*left*) and fabricated (*right*) paraffin-based microvalve. (Source: Selvaganapathy P, Carlen E, Mastrangelo C H 2003 Electrothermally actuated inline microfluidic valve. *Sens. Actuators A***104**, 275–82.)

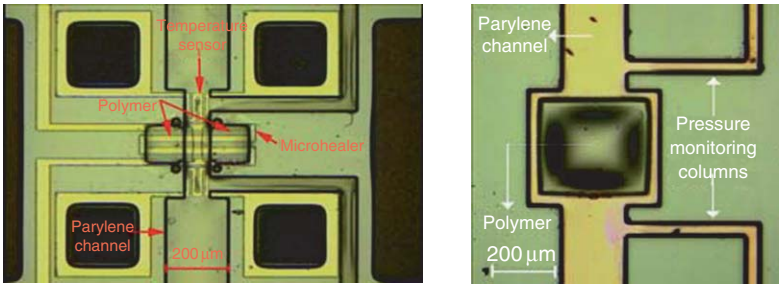


Figure 33 Thermal responsive polymer-based microvalve: overall (*left*) and close-up (*right*) view. (Source: Mutlu S, Yu C, Svec F, Mastrangelo C, Frechet J, Gianchandani Y B 2003 A thermally responsive polymer microvalve without mechanical parts photo-patterned in a parylene channel. *Int. Conf. Solid State Sensors, Actuators and Microsystems (Transducers '03)*, Boston, MA, USA, pp. 802–5.)

micropump and dispenser. The concept and fabricated device is shown in **Figure 34**. It consists of two subsystems. The microfluidic transdermal interface (MTI) creates the passages for the liquid through the outermost skin layer using hollow, side-opened microneedles; the dosing and actuation unit (DAU), pumps, and dispensers provide accurate amounts of liquid through MTI at appropriate times. The schematic and working principle of the micropump is illustrated in **Figure 34** (Roxhed *et al.* 2004, Samel *et al.* 2003). Integrated microheaters are used to elevate the temperature of the PDMS-XB composite locally, allowing the composite to be expanded locally without patterning it. The local expansion causes the liquid in the reservoir to be injected into the microchannel for the next step analysis. Thermal ink-jet print heads is another successful example enabled by thermal actuation (Klein 2004).

MEMS gas valve. Gas valve is important to control and manipulate the gas flow for many different applications. MEMS-based gas valve offers several advantages over traditional gas valves, such as compact size and low cost. However, it could experience the oscillation of the device in the presence of the high flow rate of the gas, resulting in degraded performance and even damaging of the devices. This issue has been addressed and solved by using the silicone strap as a damping structure described in the following.

The principle of a MEMS normally closed gas valve is illustrated in **Figure 35** (Knobloch *et al.* 2006). In the top structural layer, a slider in which rectangular slits are micromachined is connected with a bent-beam actuator. When the slider is moved, the gas flow through a series of flow channels in the handle wafer can be modulated. The fabricated

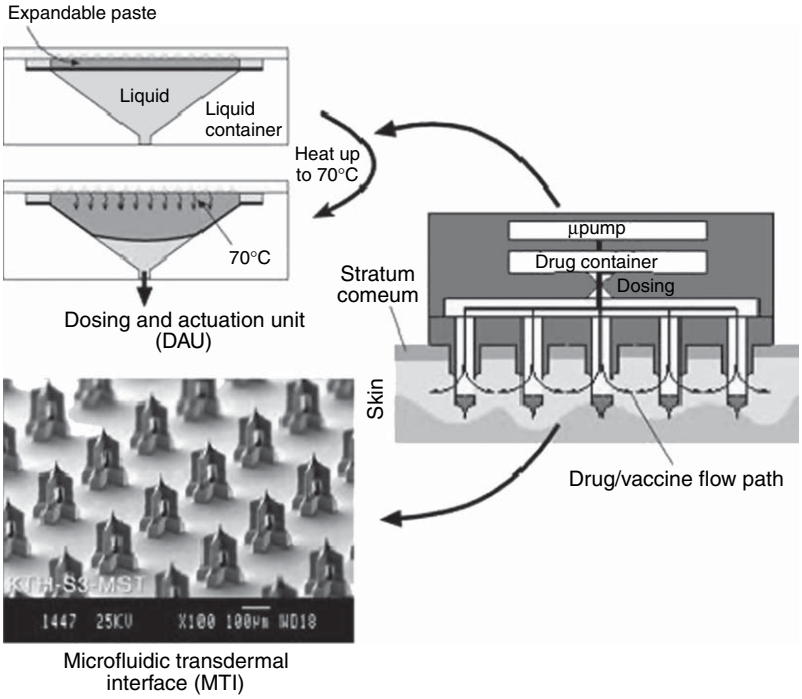


Figure 34 One-time-shot dispenser and microfluidic transdermal interface for drug delivery. (Sources: Samel B, Griss P, Stemme G 2003 Expandable microspheres incorporated in a PDMS matrix: A novel thermal composite actuator for liquid handling in microfluidic applications. *Int. Conf. Solid-State Sensors*, *Actuators and Microsystems (Transducers '03)*, Boston, MA, USA, pp. 1558–61; Roxhed N, Rydholm S, Samel B, van der Wijngaart W, Griss P, Stemme G 2004 Low cost device for precise microliter range liquid dispensing. *Tech. Dig. 17th IEEE Int. Conf. Micro Electro Mechanical Systems*, Maastricht, The Netherlands, pp. 326–9.)

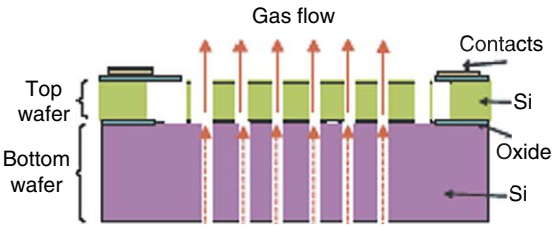


Figure 35 Cross-sectional view of the microelectromechanical systems (MEMS) gas valve at its open stage. (Source: Knobloch A, Seeley C, Mulay A, Saia R 2006 Flow structure instability prevention in a MEMS high flow gas valve. *Dig. Solid-State Sensors*, *Actuators*, *and Microsystems Workshop*, Hilton Head Island, SC, USA, pp. 288–91.)

device is shown in **Figure 36**. Experiments show that at high flow rate the gas microvalve experiences oscillation. In order to mitigate the resonance due to a fluid–structure interaction, silicone strap is used as a damping structure between the slider mechanism and the rigid supporting structure around the valve. Experiments found that the dampers show no failure up to 10 000 cycles actuation.

2.03.3.3 Wireless Communication

Many different micromachined components for wireless communication have been developed in the past. MEMS technology shows great promise and potential in this field since it usually offers small footprint and low cost while still maintaining super performance. Although most components utilize the electrostatic actuation for movable parts, significant portions of them use thermal actuation. Low actuation voltage is obviously one of the major advantages.

Specifically, compared to a conventional semiconductor switch and relay, a MEMS RF switch and relay are attractive because they have much better performance in terms of isolation, insertion loss, return loss, power consumption, and linearity. Even though its speed is not as fast as that of semiconductor switches, a MEMS RF switch is good enough for many applications. Similarly, for other RF components, such as voltage-controlled oscillator (VCO), resonator, and filter, their MEMS counterparts offer the similar advantages. Thermal actuation requires low voltage, thus making it very attractive for

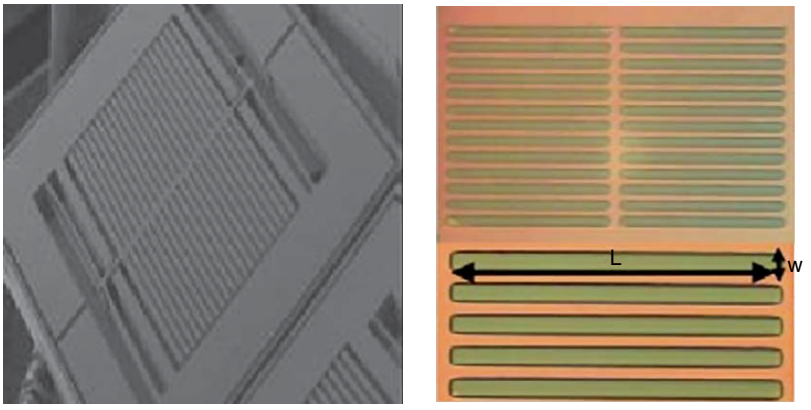


Figure 36 Perspective (*left*) and close-up (*right*) view of the gas valve. (Source: Knobloch A, Seeley C, Mulay A, Saia R 2006 Flow structure instability prevention in a MEMS high flow gas valve. *Dig. Solid-State Sensors, Actuators, and Microsystems Workshop*, Hilton Head Island, SC, USA, pp. 288–91.)

integration with RF components. However, power consumption could be a problem. Several ways have been implemented to abate or avoid this issue as mentioned earlier. In the following, thermally actuated RF switch, variable capacitor, VCO, and filter will be discussed.

RF relay and latching switch. One approach enables high-performance RF switch by combining bimorph actuator and electrostatic actuation (Robert *et al.* 2003). The thermal actuator is for the on/off function with low actuation voltage, whereas the electrostatic actuation is implemented as the latching mechanism ensuring low power consumption.

The sketch of the switch is illustrated in **Figure 37**. The bimorph actuator is formed by the aluminum layer and the silicon nitride layer. A TiN layer is sandwiched between two silicon nitride layers and serves as both the microheating resistors and the latching electrodes. The switch contact is formed by nitride etching and gold layer deposition. The fabricated device is shown in **Figure 38**. Its switching time is about 200 μs with 2 V actuation voltages. The latching voltage is <10 V. More details of its RF performance can be found in Robert *et al.* (2003).

Bistable RF switch and circuit. Bistable RF switch has been designed and implemented using bent-beam

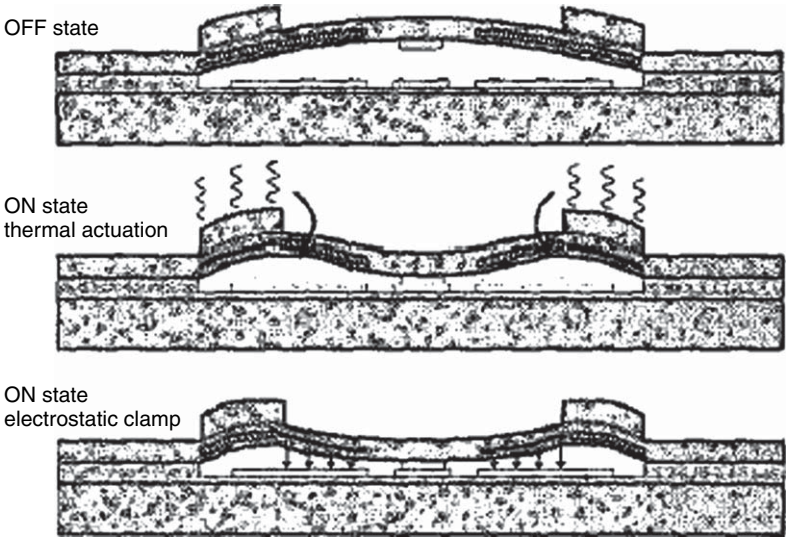


Figure 37 Schematic of the radio frequency (RF) switch combing thermal and electrostatic actuation. (Source: Robert P, Saias D, Billard C, Boret S, Sillon N, Maeder-Pachurka C, Charvet P L, Bouche G, Ancey P, Berruyer P 2003 Integrated RF-MEMS switch based on a combination of thermal and electrostatic actuation. *12th Int. Conf. on Solid-State Sensors, Actuators and Microsystems* vol. **2**. pp. 1714–17.)

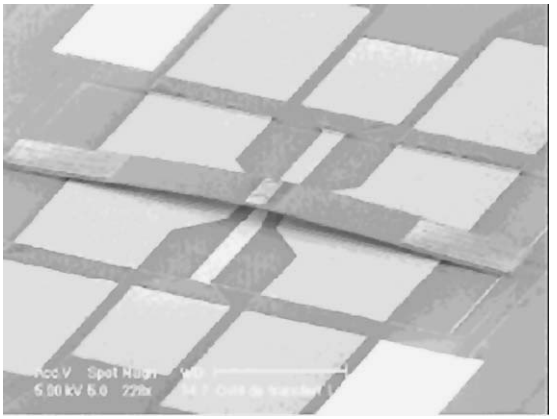


Figure 38 Scanning electron micrograph (SEM) of a fabricated device. (Source: Robert P, Saias D, Billard C, Boret S, Sillon N, Maeder-Pachurka C, Charvet P L, Bouche G, Ancey P, Berruyer P 2003 Integrated RF-MEMS switch based on a combination of thermal and electrostatic actuation. *12th Int. Conf. on Solid-State Sensors, Actuators and Microsystems* vol. **2**. pp. 1714–17.)

actuators as the driving engine (Howell 2001, Que *et al.* 2004). The simple structure and high impact force by the actuator offers unique advantages over other designs (Rebiez 2003).

The bistable structure consists of a suspended beam with a shallow lateral bend. A flexible element in the center offers longitudinal compliance; thus, it can be pushed laterally from side to side. A descriptive calculated potential energy curve as well as the reaction curve is shown in **Figures 39** and **40**, respectively (Que *et al.* 2004). It clearly shows two energy valleys corresponding to two stable positions. The reaction force of the bistable structure is highly

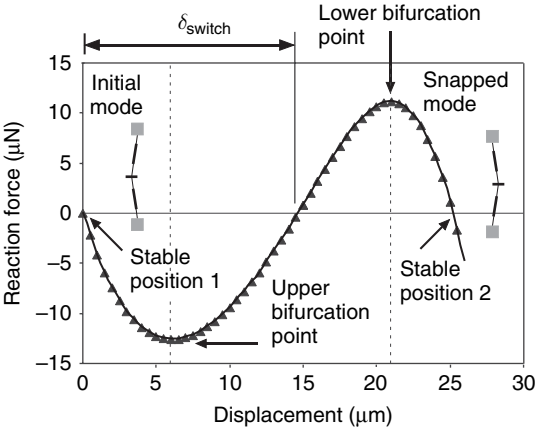


Figure 39 Energy curve showing two energy valleys corresponding two stable positions. (Source: Que *et al*. 2004.)

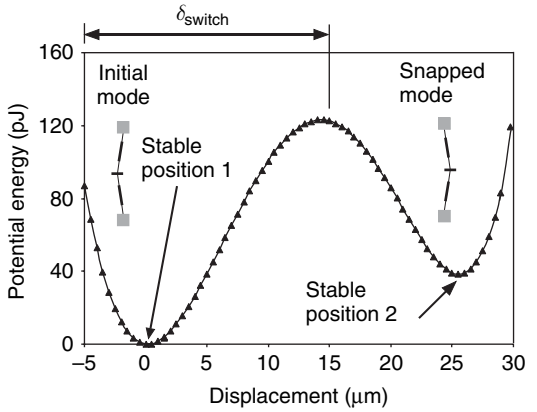


Figure 40 Impact force curve for the bistable structure. (Source: Que *et al*. 2004.)

nonlinear with the displacement. At two stable positions, the reaction force is zero. There are two ways to switch the bistable structure by the driving engine. One way is to impart sufficient kinetic energy so that bistable structure itself overcomes the potential energy peak and snaps to the other stable position. The other way is to gradually push the bistable structure with driving engine until it passes the unstable equilibrium point (energy peak), thereby falling into the other stable position.

The sketch of the designed switch is shown in **Figure 41** and the coplanar waveguide (CPW) transmission line with integrated switch is given in

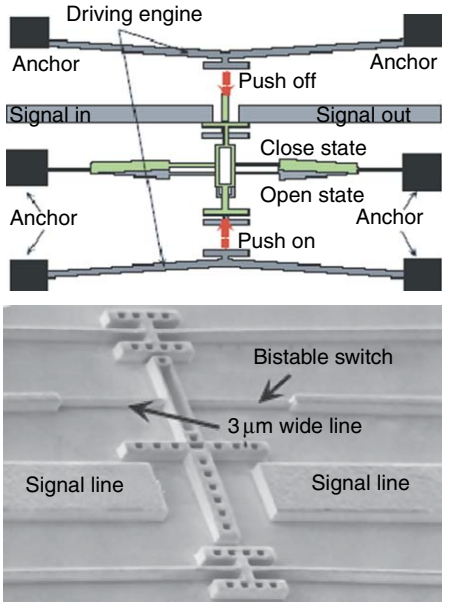


Figure 41 Sketch of the switch (*upper*) and close-up view of fabricated switch (*lower*). (Source: Que *et al*. 2004.)

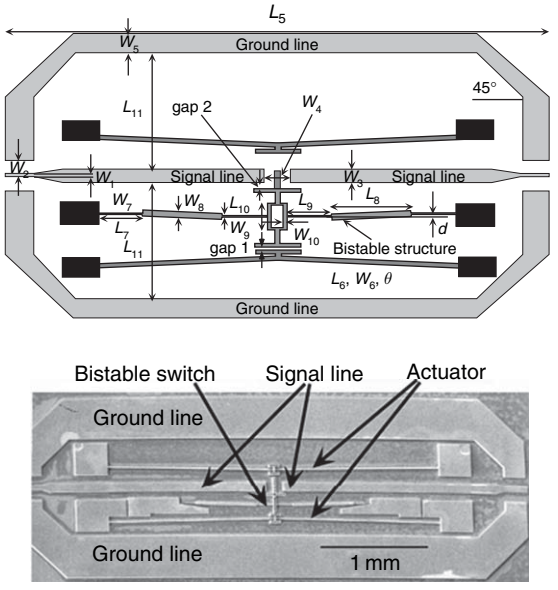


Figure 42 Sketch of the coplanar waveguide (CPW) transmission line with integrated switch (*upper*) and scanning electron micrograph (SEM) of fabricated device (*lower*). (Source: Que *et al*. 2004.)

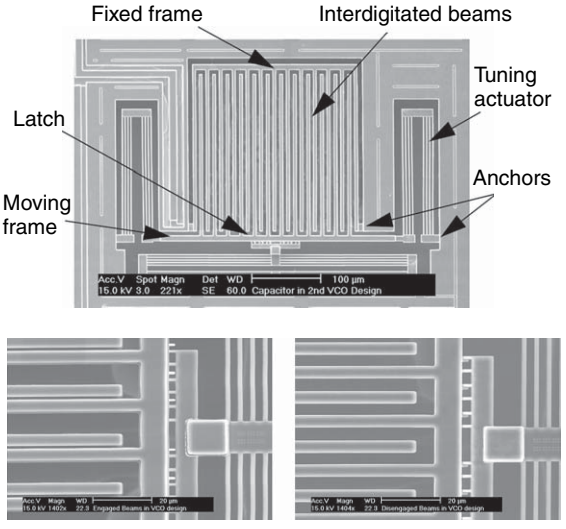


Figure 43 Microelectromechanical systems (MEMS) tunable capacitor: whole device (*upper*). Close-up views showing the latching mechanism at the maximum (*lower left*) and the minimum (*lower right*) capacitance stage, respectively. (Source: Ramachandran D, Oz A, Saraf V K, Fedder G, Mukherjee T 2004 MEMS-enabled reconfigurable VCO and RF filter. *Proc. 2004 IEEE Radio Frequency Integrated Circuits Symposium (RFIC '04)*, pp. 251–4.)

Figure 42 (Que *et al.* 2004). The driving engine is a bent-beam actuator, which can offer an impact force up to the range of millinewtons by scaling its dimensions. Measurements on these devices show excellent RF performance in terms of insertion loss and isolation, and their performance can be optimized readily. The switching speed can be around 0.1 ms if the pulsed driving voltage or current is used. Furthermore, the footprint can be minimized by optimizing the design.

This switch can be used as a basic component to design switching networks and other RF circuits while maintaining the advantages of the switch, such as microstrip line, single-pole N-through, N-pole N-through switch network, phase-shifter, phased array antenna, and tunable and reconfigurable antenna (Rebiez 2003).

Tunable capacitor, VCO, and RF filter. Reconfiguration of RF systems from one frequency to the other is of interest in multifunction radios. MEMS tunable capacitor plays an important role to design reconfigurable VCO and filter circuits by integrating with on-chip micromachined inductors and transistors. It has advantages of lower loss, larger tuning range, and more linear tuning characteristics over solid-state varactors. Significant efforts have been devoted to develop MEMS tunable capacitors during recent years (Rebiez 2003). In-plane tunable capacitor formed by interdigital beams has been developed using a CMOS process as shown in **Figure 43** (Ramachandran *et al.* 2004). The beams can be moved with respect to each other by the integrated thermal actuators, resulting in the changes of the gap between the beams, thereby the capacitance. In order to have zero-standby power operation, a lateral latch consisting of a peg in a slot is used and is shown in **Figure 43**. This capacitor has tuning ranges of 200% or even higher. A VCO is constructed with two variable capacitors, a center-tapped differential inductor at the top and a core circuit at the bottom as shown in **Figure 44**. A switching ability of more than 400 MHz has been demonstrated between frequencies (Ramachandran *et al.* 2004). Similarly, a filter is also built with one differential inductor and four tunable capacitors as shown in **Figure 43**. It has quality factor up to 9.5 and it can reject at least 7.8 dB of the undesirable frequency component.

2.03.3.4 Nanoinstrumentation and Nanotechnology

MEMS technology offers a powerful and flexible platform and infrastructure for nanoscience and nanotechnology research (Fujita 2003). It not only can be used as the infrastructure to implement

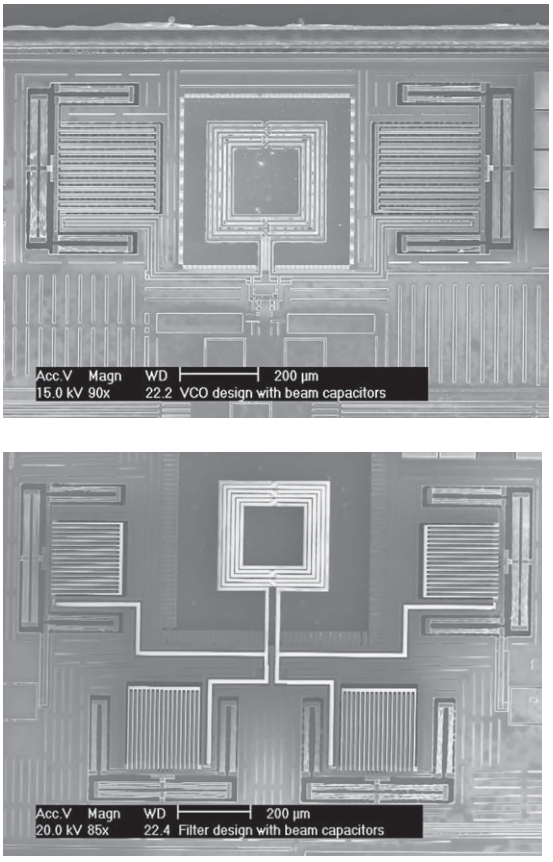


Figure 44 Scanning electron micrographs (SEM) of VCO (*upper*) and filter (*lower*). (Source: Ramachandran D, Oz A, Saraf V K, Fedder G, Mukherjee T 2004 MEMS-enabled reconfigurable VCO and RF filter. *Proc. 2004 IEEE Radio Frequency Integrated Circuits Symposium (RFIC '04)*, pp. 251–4.)

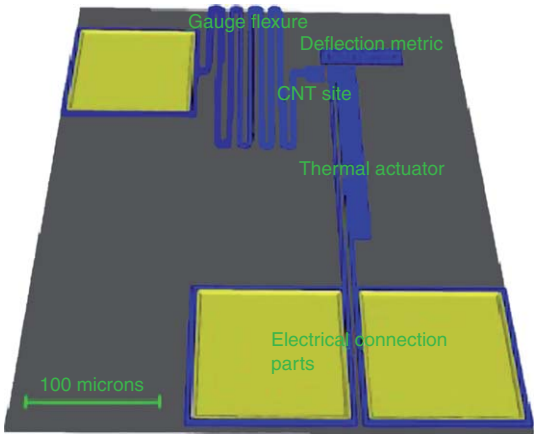


Figure 45 Tensile testing microdevice for carbon nanotubes (CNTs). (Source: Hartman A, Rice P, Finch D, Skidmore G, Bright V 2004 For-deflection characterization of individual carbon nanotubes attached to MEMS devices. *Proc. IEEE Int. Conf. on Microelectromechanical Systems,* Maastricht, The Netherlands, pp. 426–9.)

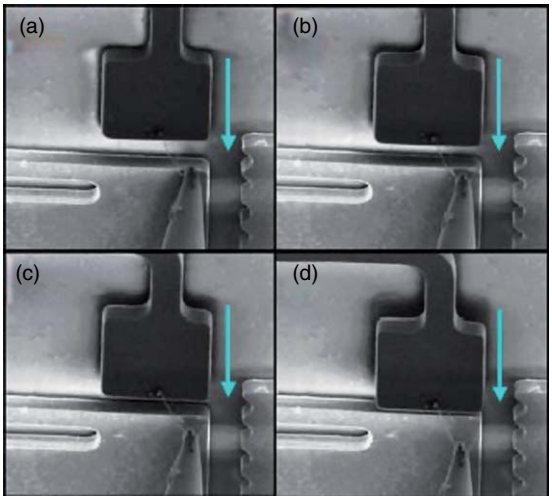


Figure 46 Microelectromechanical systems (MEMS) flexure deflected by an attached carbon nanotube (CNT). (Source: Hartman A, Rice P, Finch D, Skidmore G, Bright V 2004 For-deflection characterization of individual carbon nanotubes attached to MEMS devices. *Proc. IEEE Int. Conf. on Microelectromechanical Systems,* Maastricht, The Netherlands, pp. 426–9.)

nanoinstrumentation but also can be exploited as a platform for the characterization of advanced materials. Again, the thermal actuation plays a very important role in this field since it can offer a high force (up to several millinewtons), which is essential to deform nanoscale materials with high a Young's modulus, such as CNTs. Furthermore, micromachined structure serves as a bridge between the nanoscale domain and the macroscale world, enabling a much more convenient way to interface the input/output. In the following, several examples will be used to illustrate the MEMS-enabled nanoinstrumentation and technical platform.

Mechanical characterization of CNTs. In nanoscience and nanotechnology community, sophisticated and expensive equipment such as transmission electron microscopy (TEM) and AFM has been used to analyze the properties of CNTs. It is a complicated and time-consuming process. In order to address this issue, an *in situ* manipulation, attachment technique of individual CNT to MEMS devices, has been explored.

One approach is to exploit the micromachined pseudo-bimorph as the tensile force generator to characterize the properties of CNTs (Hartman *et al.* 2004). The tensile test device in **Figures 45** and **46** consists of a gauge flexure, a deflection meter, and a

thermal actuator. The individual CNT is attached to the two separated MEMS surfaces (the gauge flexure and the thermal actuator). When the actuator is powered, the CNT is stretched due to the tensile force. The deflection of the gauge flexure, therefore, is measured and is used to evaluate the mechanical properties of CNT.

Nanopositioner. Nanoscale positioning is essential for many instruments and applications such as scanning microscopy, optical switch, wireless communication, and data storage. Some nanopositioners use electrostatic, magnetic, and piezoelectric transduction mechanisms (Guckel *et al.* 1998, Hirano *et al.* 1998), whereas several efforts focused on thermal actuation-based approaches (Chu and Gianchandani 2003, Tuantranont *et al.* 2000).

A positioner that offers subnanometer–resolution positioning is used as an example (Chu and Gianchandani 2003). The schematic of a 2D positioner is shown in **Figure 47**. It consists of the bent-beam actuator, the amplifier, and built-in capacitive position sensor. The device in **Figure 47** is fabricated into a SOI wafer made with fusion wafer bonding. For this particular design, it can achieve 19 μm displacement along each axis at subnanometer resolution (~0.3 nm).

MEMS probes for nanolithography. Maskless lithography is sometimes more flexible in defining nanostructures than is traditional masked lithography. Maskless lithography includes E-beam litho, interference litho, micromirror array-based litho, and AFM tip litho. However, all of them involve coating of photoresist as the mask and the final features are formed by etching processes. Recently, a photoresist-free maskless lithography technique, called dip pen nanolithography (DPN) (Piner *et al.* 1999), has been developed. DPN utilizes scanning probes to deposit chemical and biological materials directly on a substrate. The thermally actuated probe is used to achieve dynamical DPN (Wang *et al.* 2004).

In order to enhance its throughput, a silicon probe array with tips of 100 nm radius of curvature has been fabricated. Schematic and SEM of the fabricated silicon probe array are shown in **Figures 48** and **49**, respectively. Each probe can be addressed with a bimorph thermal actuator on its cantilever, enabling the flexibility of nanoscale pattern generation. It demonstrated parallel writing of different octadecanethiol (ODT) patterns with a sub-50 nm linewidth on the gold substrate.

2.03.3.5 Compliant Mechanism and Other Applications

Owing to space constraints, other thermal actuator-enabled applications are not covered. Compliant micromechanism is a structure deforming elastically to transmit a force or displacement. It will be illustrated since it is the basis for many integrated micromechanical systems and has been explored extensively (Ananthasuresh *et al.* 1994, Kota *et al.* 1999). One example is the mechanical amplifier implemented using compliant micromechanism.

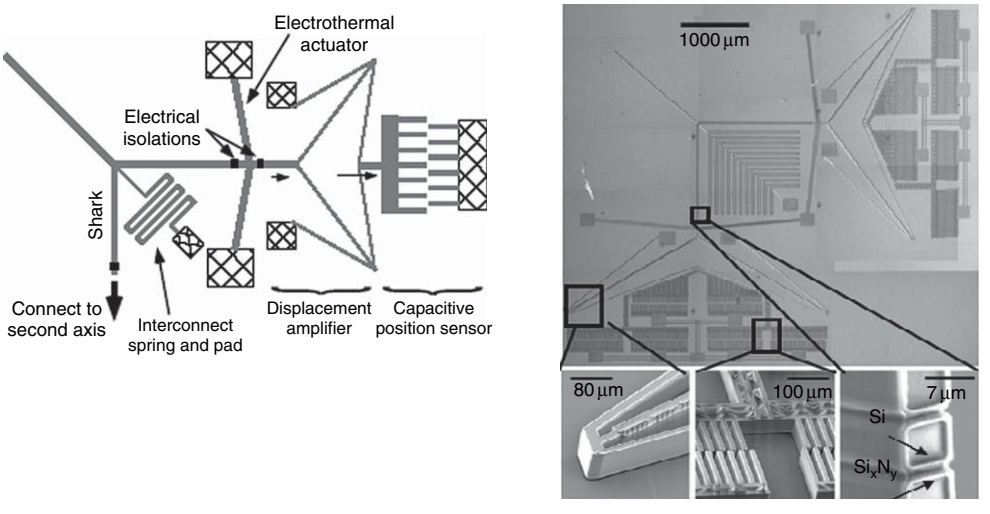


Figure 47 Nanopositioner with integrated position sensing operated by a bent-beam actuator. (Source: Chu and Gianchandani 2003.)

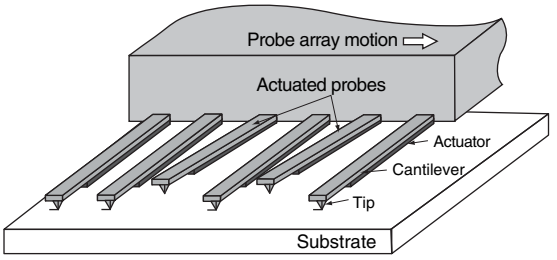


Figure 48 Schematic of arrayed thermal actuated probe for dip pen nanolithography (DPN). (Source: Wang X, Bullen D, Zou J, Liu J, Mirkin C 2004 Thermally actuated probe array for parallel dip-pen nanolithography. *J. Vac. Sci. Technol. B* **22(6)**, 2563–7.)

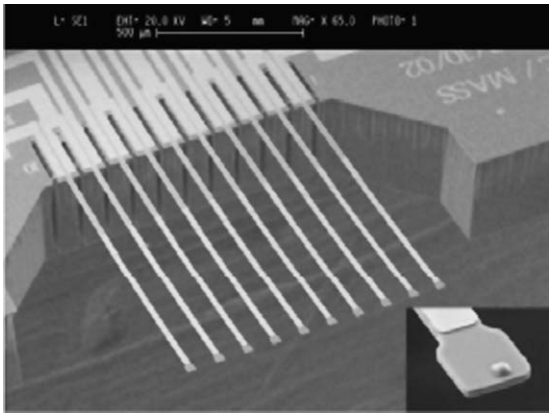


Figure 49 Scanning electron micrograph (SEM) of an active silicon-probe array. (Source: Wang X, Bullen D, Zou J, Liu J, Mirkin C 2004 Thermally actuated probe array for parallel dip-pen nanolithography. *J. Vac. Sci. Technol. B* **22(6)**, 2563–7.)

Rectilinear motion amplifier. The mechanical amplifier is implemented with a bent-beam actuator to modify the force–displacement relationships (Chu *et al.* 2002). The topology synthesis using truss elements and dimensional optimization on beam elements are two major design stages. The ultimate goal is to maximize the blocking force and unloaded output displacement under all relevant design constraints.

One optimized design of a rectilinear microtransmission with two inputs is illustrated in **Figure 50(a)**. The optical micrograph of the corresponding fabricated nickel device is shown in **Figure 50(b)**, which is fabricated using single-crystal p^{++} silicon by the dissolved wafer process. The dimensions for this design are $L_0 = 2450\,\mu m$, $L_1 = 2415\,\mu m$, $L_2 = 1750\,\mu m$, $L_3 = 1400\,\mu m$, $L_4 = 1000\,\mu m$, $L_5 = 1210\,\mu m$, $W = 70\,\mu m$, whereas the thickness is $H = 11.5\,\mu m$. For this design, rectilinear nonresonant output displacements up to 100 μm and displacement amplification factors of about 17 were reported. Both displacements and amplification factors can readily be scalable by modifying the dimensions of the device.

Other applications. Thermal actuation has also been widely used for many other different applications, such as microaccelerator, resonator, micromotor (Park *et al.* 2001), tunable electrostatic converter, and thermal oscillator (Udeshi and Gianchandani 2004). Owing to space constraints, only the thermal oscillator is described. The detailed information of

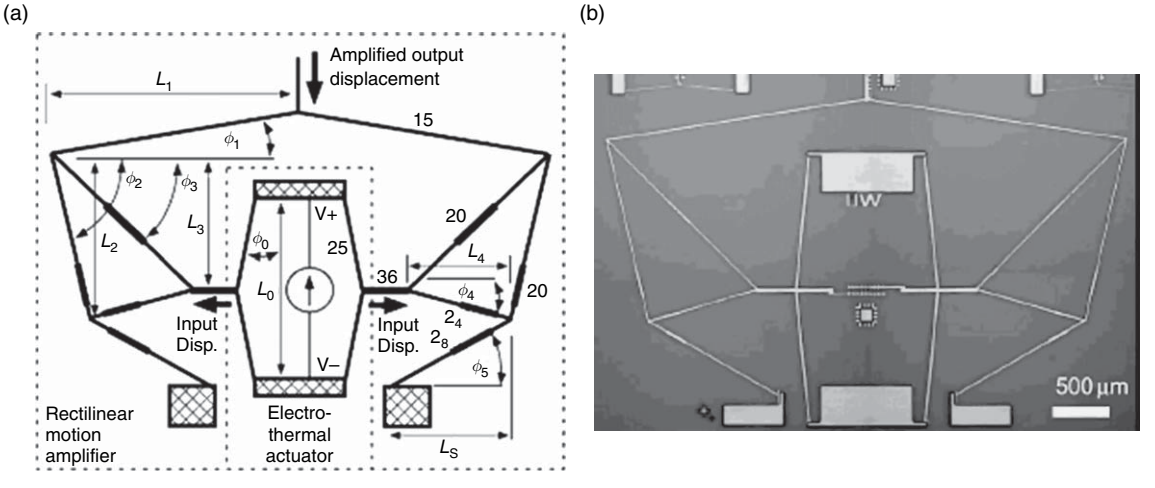


Figure 50 (a) Schematic of an optimized rectilinear microtransmission design and (b) optical micrograph of fabricated device. (Source: Chu *et al.* 2002.)

other devices can be found in their respective citations (Park *et al.*, 2001).

Some micromachined devices either require or exhibit improved performance when driven by a modulating signal, such as strain sensors and pressure sensors (Burns *et al.* 1995), accelerometers, and gyroscopes (Putty and Najafi 1994, Roessig *et al.* 1997). The modulating signal is usually generated by oscillators, typically implemented using standard microelectronic circuits. However, this approach severely limits their voltage and power-handling capacity, making it unsuitable for any application demanding even moderately high voltages or power levels. In addition, even those oscillators that use MEMS components employ electronics to provide the feedback of an amplified signal (Dec and Suyama 2000, Nguyen 1997).

A fully mechanical micromachined oscillator operated using a single DC power source, without the aid of any electronic components, has been developed to address these issues (Udeshi and Gianchandani 2004). The schematic of a mechanical oscillator and its operation procedure is shown in **Figures 51** and **53**. The SEM of the fabricated device is shown in **Figure 52**.

The oscillator is enabled by the bent-beam actuator. The actuator opens a normally closed switch when actuated, resulting in cutting its own current supply. Subsequent cooling of the actuator causes the closure of the switch, thus the cycle to repeat, resulting in oscillations (**Figure 53**). However, hysteresis must be introduced in the system in order to keep the switch open or closed in a controllable manner, allowing the system to remain in unstable equilibrium for a predefined amount of time. Two types of the oscillators have been implemented using different hysteresis mechanisms.

In the Type I oscillator, hysteresis is based on the bistable toggle mechanism (Jensen *et al.* 2001). Initially, the switch is closed and the thermal actuator undeflected. When the voltage is applied, the thermal actuator is heated up, resulting in an increasing force

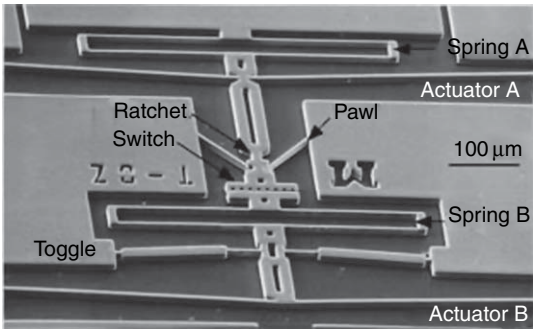


Figure 52 Scanning electron micrograph (SEM) of fabricated oscillator. (Source: Udeshi and Gianchandani 2004.)

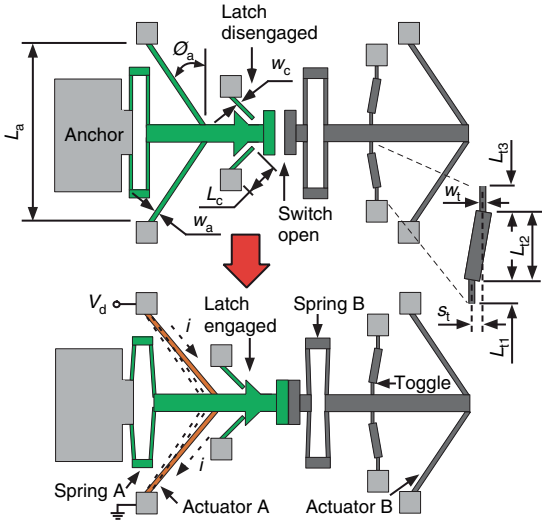


Figure 51 Device configuration: fabricated (*top*) and after priming (*bottom*). Priming is permanent and achieved by driving actuator A. (Source: Udeshi and Gianchandani 2004.)

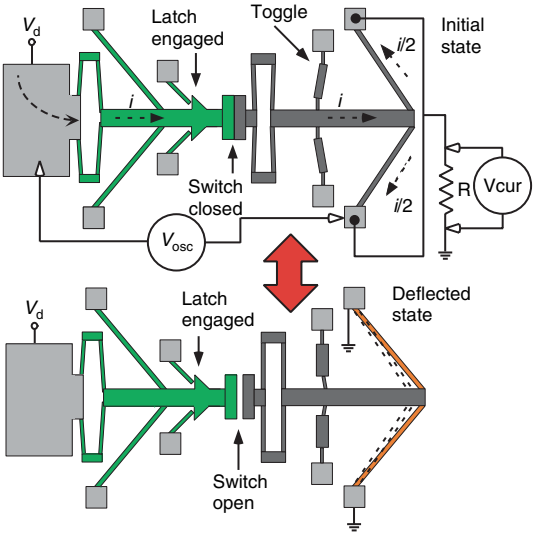


Figure 53 Device operation: current flow when switch is closed (*top*) causes displacement of electrothermal actuator (*bottom*), opening the switch. Subsequent cooling of the actuator closes the switch, permitting repetition. (Source: Udeshi and Gianchandani 2004.)

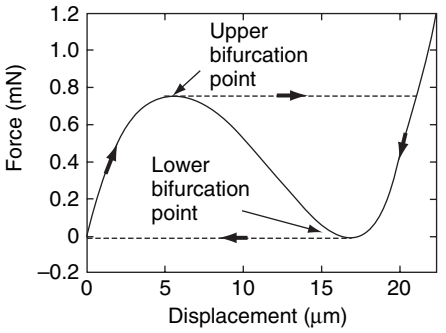


Figure 54 Finite element analysis (FEA) load–displacement response of the bistable toggle element. (Source: Udeshi and Gianchandani 2004.)

on the toggle. Once the force exceeds the force corresponding to the upper bifurcation point of the toggle (**Figure 54**), the toggle snaps through, generating a large displacement, opening the switch, and cutting the current to the actuator, resulting in its deflected state. Without power supply, the actuator cools down due to heat dissipation, decreasing the force applied on the toggle. The toggle snaps back to its original position when the applied force falls below that of the lower bifurcation point, closing the switch and restoring the current to the actuator. The device returns to its initial state, resulting in the repetition of the cycle, and thus oscillation (**Figure 53**).

In the Type II oscillator, the function of bistable element is achieved by the intrinsic thermal hysteresis of actuator B (**Figure 51**), resulting from the duration of heating and cooling time constants of an electrothermal actuator.

Measurements found that the Type I device generates a signal with a relatively constant frequency of about 39 Hz at different power levels, with a variable duty cycle from 0.3 to 0.6. The Type II oscillator is tunable, with a frequency range from 200 to 1200 Hz and a duty cycle from 0.3 to 0.7. Less than 0.5 V is required for the oscillators. The footprint of both devices is measured at about 1 mm × 1 mm (Udeshi and Gianchandani 2004).

2.03.4 Summary and Future Trend

The theory, materials, fabrication, issues, and applications of thermal actuation have been discussed and summarized in this chapter. Thermal actuation is based on thermal expansion or shrinkage of the structural materials, which could be gas, liquid, or solid. A basic mathematical model has been developed and implemented using finite difference method. It is applicable to all types of thermal actuators. The major issues related to thermal actuators are power consumption, response time, and thermal isolation. In order to overcome these issues, specific designs and approaches have been explored, proposed, and implemented. The power consumption can be significantly reduced by cutting the effective thermal mass, using microscale latching mechanisms or implementing more efficient actuation schemes, including pulsed or burst-pulsed actuation. Similarly, all these schemes are also useful for the response speed improvement. In addition, another way to enhance the response is to integrate the cooling microfluidic network with the thermal actuators, accelerating the heat dissipation during the cooling cycle of thermal actuation. As a critical issue for many applications, thermal isolation has been achieved by sandwiching the low-heat-conduction materials between the actuators and other parts of the microsystems. From biomedical imaging, label-free biosensing to cell manipulation and drug delivery; from microfluidics, single-neuron monitoring to optical and wireless communication; from microoptics, ultrafast optical pulse shaping to nanoinstrumentation and nanoscale material characterization; thermal actuation has found a variety of applications and has proven to be an enabling technology.

A new type of thermal actuation that offers high force, consumes low power, and operates faster, however, is still needed. Besides the device design optimization and possible integration of cooling network, exploration of new advanced materials probably is an ultimate way to achieve this goal even though it could be quite challenging.

References

Ananthasuresh G K, Kota S, Kikuchi N 1994 Strategies for systematic synthesis of compliant MEMS. *Proc. 1994 ASME Winter Annu. Meet., Symp. MEMS, Dynamics Systems and Control,* DSC-Vol. 55-2, Chicago, IL, USA, pp. 677–86

Ashkin A 1997 Optical trapping and manipulation of neutral particles using lasers. *Proc. Natl. Acad. Sci. USA* **94**, 4853–60

BeMent S L, Wise K D, Anderson D, Najafi K, Drake K L 1986 Solid state electrodes for multi channel multiplexed intracortical neuronal recording. *IEEE Trans. Biomed. Eng.* **BME-33**, 230–40

Bifano T, Johnson H, Bierden P, Mali P K 2002 Elimination of stress-induced curvature in thin-film structures. *J. Microelectromech. Syst.* **11(5)**, 592–7

Burns D, Zook J, Horning R, Herb W, Guckel H 1995 Sealed-cavity resonant microbeam pressure sensor. *Sens. Actuators A. Phys.* **48(3)**, 179–86

Chiou P Y, Ohta A, Wu M C 2005 Massively parallel manipulation of single cells and microparticles using optical images. *Nat. Lett.* **436**, 370–2

Chronis N, Lee L 2005 Electrothermally activated SU-8 microgripper for single cell manipulation in solution. *IEEE/ASME J. Microelectromech. Syst.* **14**, 857–63

Chu L, Gianchandani Y B 2003 A micromachined 2D positioner with electrothermal actuation and sub-nanometer capacitive sensing. *J. Micromech. Microeng.* **13**, 279–85

Chu L, Hetrick J, Gianchandani Y B 2002 High amplification compliant microtransmissions for rectilinear electrothermal actuators. *Sens. Actuators A* **97–98**, 776–83

Chu L, Que L, Oliver A, Gianchandani Y B 2006 Lifetime studies of electrothermal bent-beam actuators in single crystal silicon and polysilicon. *IEEE/ASME J. Microelectromech. Systs.* **15(3)**, 498–506

Comtois J H, Bright V M 1997 Application for surface-micromachined polysilicon thermal actuators and arrays. *Sens. Actuators A* **58**, 19–25

Comtois J H, Michalicek M A, Barron C C 1997 Characterization of electrothermal actuators and arrays fabricated in a four-level, planarized surface-micromachined polycrystalline silicon process. *Int. Conf. Solid-State Sensors and Actuators (Transducers '97)*, Chicago, IL, USA, pp. 769–72

Cowan W D, Bright V M 1997 Vertical thermal actuators for micro-opto-electro-mechanical systems. *Proc. SPIE* **3326**, 137–46

Cragun R, Howell L 1999 Linear thermomechanical microactuators. *MEMS 1999 ASME International Mechanical Engineering Congress and Exposition*, Nashville, TN, USA, pp. 181–8

Dec A, Suyama K 2000 Microwave MEMS-based voltage-controlled oscillators. *IEEE Trans. Microwave Theory Tech.* **48(11)**, 1943–9

Ekins R P 1998 Ligand assays: From electrophoresis to miniaturized microarrays. *Clin. Chem.* **44**, 2015–30

Fujita H (*ed*.) 2003 *Micromachines as Tools for Nanotechnology.* Springer, Berlin

Green M J 1987 New approaches to electrochemical immunoassays. *Biosensors: Fundamentals and Applications*. Oxford University Press, New York, NY, Chap. 4

Guckel H, Klein J, Christenson T, Skrobis K, Laudon M, Lovell E G 1992 Thermomagnetic flexure actuators. *Tech. Dig. Solid State Sensor and Actuator Workshop*, Hilton Head, SC, USA, pp. 73–5

Guckel H, Fischer K, Stiers E 1998 Closed-loop controlled, large throw, magnetic linear microactuator with 1000 μm structure height. *IEEE Int. Conf. Micro-electro-mechanical systems (MEMS '98)*, Heidelberg, Germany, pp. 414–18

Hartman A, Rice P, Finch D, Skidmore G, Bright V 2004 For-deflection characterization of individual carbon nanotubes attached to MEMS devices. *Proc. IEEE Int. Conf. on Microelectromechanical Systems,* Maastricht, The Netherlands, pp. 426–9

Hirano T, Fan L S, Lee W Y, Hong J, Semba T, Imaino W, Pattanaik S, Chan S, Webb P 1998 A micro-actuator for a hard-disk drive fine tracking servo. *Proc. ASME Int. Mech. Eng. Congress and Exposition*, New York, NY, USA, pp. 253–60

Homola J, Yee S, Gauglitz G 1999 Surface plasmon resonance sensors: Review. *Sens. Actuators* **B54**, 3–15

Howell L 2001 *Compliant Mechanism*. John Wiley and Sons, Inc., Hoboken, NJ

Huang S, Liao B, Zhang X 2006 Elimination of stress-induced curvature in microcantilever infrared focal plane array. *Sens. Actuators A* **130–131**, 331–9

Jensen B, Parkinson M, Kurabayashi K, Howell L, Baker M 2001 Design optimization of a fully-compliant bistable micro-mechanism. *ASME Int. Mech. Eng. Congress and Exposition,* New York, NY, USA, pp. 357–63

Kawada H, Yoshida H, Kamakura M, Yoshida K, Saitou M, Kawahito K, Tomonari S 2003 Thermally driven microactuator containing thermal isolation structure with polyimide and its application to microvalve. *12th Int. Conf. Solid State Sensors, Actuators and Microsystems (Transducers '03)*, Boston, MA, USA, pp. 1935–8

Kim C J, Pisano A P, Muller R S 1992 Silicon-processed overhanging microgripper. *J. Microelectromech. Syst.* **1**, 31–6

Klein R J 2004 Thermal and inkjet printer, *US Pat.* 7: 170–538

Knobloch A, Seeley C, Mulay A, Saia R 2006 Flow structure instability prevention in a MEMS high flow gas valve. *Dig. Solid-State Sensors, Actuators, and Microsystems Workshop*, Hilton Head Island, SC, USA, pp. 288–91

Kota S, Hetrick J, Li Z, Saggere L 1999 Tailoring unconventional actuators using compliant transmissions: Design methods and applications. *IEEE/ASME Trans. Mechatronics* **4**, 396–408

Lee W, Cho Y H 2004 Nanomechanical protein detectors using electrothermal nano-gap actuators. *Proc. Int. Conf. MEMS*, Maastricht, The Netherlands, pp. 629–32

Lin L, Pisano A P, Lee A P 1991 Microbubble powered actuator. *Int. Conf. Solid-State Sensors and Actuators (Transducers '91)*, San Francisco, CA, USA, pp. 1041–4

Lott C, Mclain T, Harb J, Howell L 2002 Modeling the thermal behaviour of a surface-micromachined linear-displacement thermomechanical microactuator. *Sensors and Actuators A* **101**, 239–50

Lu S, Panchapakesan B 2005 Optically driven nanotube actuators. *Nanotechnology* **16**, 2548–54

Ma C, Wang R, Sun Q, Zohar Y, Wong M 2000 Frequency response of TiNi shape memory alloy thin film microactuators. *IEEE Int. Conf. Micro-electro -mechanical systems (MEMS '00)*, Miyazaki, Japan, 370–4

Muthuswamy J, Okandan M, Gilletti A, Baker M S, Jain T 2005 An array of microactuated microelectrodes for monitoring single-neuronal activity in rodents. *IEEE Trans. Biomed. Eng.* **52(8)**, 1470–7

Mutlu S, Yu C, Svec F, Mastrangelo C, Frechet J, Gianchandani Y B 2003 A thermally responsive polymer microvalve without mechanical parts photo-patterned in a parylene channel. *Int. Conf. Solid State Sensors, Actuators and Microsystems (Transducers '03)*, Boston, MA, USA, pp. 802–5

Najafi K, Wise K D, Najafi N 1994 Integrated sensors. In: Sze S M (ed.) *Semiconductor Sensors.* John Wiley and Sons, Inc., New York, pp. 473–530

Nguyen C 1997 High-Q micromechanical oscillators and filters for communications. *1997 IEEE Int. Symp. Circuits and Systems*, Hong Kong, China, pp. 2825–8

Oliver A D, Vigil S R, Gianchandani Y B 2003 Photothermal surface-micromachined actuators. *IEEE Trans. Electron. Devices* **50(4)**, 1156–7

Pan C S, Hsu W 1997 An electro-thermally and laterally driven polysilicon microactuator. *J. Micromech. Microeng.* **7**, 7–13

Park J, Chu L, Oliver A, Gianchandani Y B 2001 Bent-beam electrothermal actuators – Part II: Linear and rotary microengines. *IEEE J. Microelectromech. Syst.* **10(2)**, 255–62

Piner R D, Zhu J, Xu F, Hong S, Mirkin C A 1999 Dip pen lithography. *Science* **283**, 661

Putty M, Najafi K 1994 A micromachined vibrating ring gyroscope. *Solid State Sensor and Actuator Workshop*, Hilton Head, SC, USA, pp. 213–20

Que L 2000 Micromachined sensors and actuators based on bent-beam suspensions. Ph.D. thesis, University of Wisconsin-Madison

Que L, Gianchandani Y B 2005 MOEMS actuation and sensing. In: Motamedi E (ed.) *Micro-opto-electro-mechanical Systems (MOEMS)*, pp. 121–210

Que L, Park J S, Gianchandani Y B 1999 Bent-beam electrothermal actuators for high force application. *Proc. IEEE Int. Conf. MEMS (MEMS '99)*, Orlando, FL, USA, pp. 31–6

Que L, Otradovec L, Oliver A, Gianchandani Y B 2001a Pulse and DC operation lifetimes of bent-beam electrothermal actuators. *Proc. IEEE Int. Conf. MEMS (MEMS '01)*, Switzerland, pp. 570–3

Que L, Park J, Gianchandani Y B 2001b Bent-beam electrothermal actuators – Part I: Single beam and cascaded devices IEEE. *J. Microelectromech. Syst.* **10(2)**, 247–54

Que L, Lim M, Fan R 2003 Latching mechanism for MEMS actuator and method of fabrication, *US Pat.* 6 549–1107 B2

Que L, Udeshi K, Park J, Gianchandani Y B 2004 A bi-stable electrothermal RF switch for high force applications. *Proc. IEEE Int. Conf. on Microelectromech. Systs.* Maastricht, The Netherlands, pp. 797–800

Ramachandran D, Oz A, Saraf V K, Fedder G, Mukherjee T 2004 MEMS-enabled reconfigurable VCO and RF filter. *Proc. 2004 IEEE Radio Frequency Integrated Circuits Symposium (RFIC '04)*, Fort Worth, TX, USA, pp. 251–4

Rebiez G 2003 *RF MEMS: Theory, Design, and Technology*. John Wiley and Sons, Inc., Hoboken, NJ

Reed M, Lye W 2004 Microsystems for drug and gene delivery. *Proc. IEEE* **92(1)**, 56–75

Rickert J, Wessa T, Gopel W 1999 Sensors for biomolecular studies. *Microsystem Technology: A Powerful Tool for Biomolecular Studies*. Birkhauser, Switzerland, Chap. 12

Robert P, Saias D, Billard C, Boret S, Sillon N, Maeder-Pachurka C, Charvet P L, Bouche G, Ancey P, Berruyer P 2003 Integrated RF-MEMS switch based on a combination of thermal and electrostatic actuation. *12th Int. Conf. on Solid-State Sensors, Actuators and Microsystems* vol. **2**. pp. 1714–17

Roessig T, Howe R, Pisano A, Smith J 1997 Surface-micromachined resonant accelerometer. *Int. Conf. Solid-State Sensors and Actuators, Proc.*, Chicago, IL, USA, **2**, pp. 859–62

Roxhed N, Rydholm S, Samel B, van der Wijngaart W, Griss P, Stemme G 2004 Low cost device for precise microliter range liquid dispensing. *Tech. Dig. 17th IEEE Int. Conf. Micro Electro Mechanical Systems*, Maastricht, The Netherlands, pp. 326–9

Samel B, Griss P, Stemme G 2003 Expandable microspheres incorporated in a PDMS matrix: A novel thermal composite actuator for liquid handling in microfluidic applications. *Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '03)*, Boston, MA, USA, pp. 1558–61

Schiller P, Polla D L 1993 Integrated piezoelectric microactuators based on PZT thin films. *Int. Conf. Solid-State Sensors and Actuators (Transducers '93)*, Yokohama, Japan, pp. 154–7

Selvaganapathy P, Carlen E, Mastrangelo C H 2003 Electrothermally actuated inline microfluidic valve. *Sens. Actuators A* **104**, 275–82

Shimamura Y, Udeshi K, Que L, Park J, Gianchandani Y 2006 Impact behavior and energy transfer efficiency of pulse-drive bent-beam electrothermal actuators. *IEEE J. Microelectromech. Syst.* **15(1)**, 101–10

Singh-Gasson S, Green R D, Yue Y, Nelson C, Blattner F, Sussman M, Cerrina F 1999 Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotechnol.* **17**, 974–8

Sniegowski J 1993 A microactuation mechanism based on a liquid-vapor surface tension. Late news paper. *Int. Conf. Solid State Sensors and Actuators (Transducers '93)*, Yokohama, Japan, pp. 12

Tang W C, Nguyen C, Howe R T 1989 Laterally drive polysilicon resonant microstructures. *Sens. Actuators* **20**, 25–32

Tuantranont A, Bright V M, Zhang J, Zhang W, Neff J, Lee Y C 2000 MEMS-controllable microlens array for beam steering and precision alignment in optical interconnect systems. *Proc. Solid-State Sensor and Actuator Workshop,* Hilton Head, SC, USA, pp. 101–4

Udeshi K, Gianchandani Y 2004 A DC-powered, tunable, fully mechanical oscillator using in-plane electrothermal actuator. *Proc. IEEE Int. Conf. Micro Electro Mechanical Systems*, Maastricht, The Netherlands, pp. 502–5

Udeshi K, Liao K, Que L, Gianchandani Y B, Galvanauskas A 2006 Programmable optical wave form shaper on a microchip. *Appl. Phys. Lett*. **89(3)**, 31120–3

Wang Y, Li Z, McCormik D, Tien N C 2002 Low-voltage lateral-contact microrelays for RF applications. *15th IEEE Int. Conf. Micro-electro-mechanical Systems*, Las Vegas, NV, USA, pp. 645–8

Wang X, Bullen D, Zou J, Liu J, Mirkin C 2004 Thermally actuated probe array for parallel dip-pen nanolithography. *J. Vac. Sci. Technol. B* **22(6)**, 2563–7

Wise K D, Angell J B, Starr A 1970 An integrated circuit approach to extracellular microelectrodes. *IEEE Trans. Biomed. Eng.* **BME-17**, 238–46

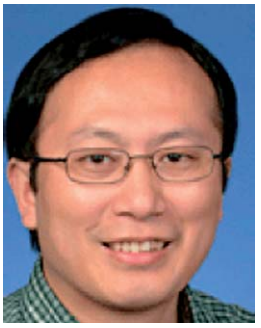
Xie T, Xie H, Fedder G K, Pan Y 2003 Endoscopic optical coherence tomography with new MEMS mirror. *Electron. Lett.* **39(21)**, 1535–6

Zdeblick M J, Anderson R, Jankowski J, Kline-Schoder B, Christel L, Miles R, Weber 1994 Thermopneumatically actuated microvalves and integrated electro-fluidic circuits. *The Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, pp. 251–5

Zhang J, Kim E S 2003 Vapor and liquid mass sensing by micromachined acoustic resonator. *Proc. 16th IEEE Int. Conf. Micro Electro Mechanical Systems*, Kyoto, Japan, pp. 470–3

Zhang Y W, Zhang Y X, Marcus R B 1999 Thermally actuated microprobes for a new wafer probe card. *IEEE J. Microelectromech. Syst.* **8(1)**, 43–9

Biography



Long Que received his undergraduate and graduate education in Physics and Communication from Peking University, Beijing, China. He earned his Ph.D. degree in Electrical Engineering focusing on MEMS/nanotechnology from the University of Wisconsin-Madison in 2000.

He is presently with GE Global Research Center as a research staff member, task leader, and project leader. Prior to this, he was with the Center for Nanoscale Materials at Argonne National Laboratory, Argonne, IL, where his research focused on nanoscience and nanotechnology. Prior to this, he was with the Electrical Engineering and Computer Science (EECS) Department and the Center for Wireless Integrated Microsystems (WIMS) of the University of Michigan at Ann Arbor. His research interests are in bioMEMS/optical/RF MEMS, nanoscience, and nanotechnology. He has coauthored a textbook *MOEMS (SPIE* '05), published more than 30 papers in major MEMS/nanotechnology journals and conferences, and has been awarded five US patents, and has more than 10 patents pending.

Dr. Que won Invention Awards from GE in 2006 and 2007, a National Research Award from the Chinese Academy of Sciences in 1997, and the Vilas Professional Development Fellowship from the University of Wisconsin in 2000. He is a member of IEEE and SPIE.

2.04 Pressure Sensors

Andrew DeHennis, Germantown, MD, USA

Junseok Chae, Tempe, AZ, USA

2.04.1 Introduction

Micromachined pressure sensors are one of the most successfully commercialized microdevices. As one of the first developed MEMS devices, for more than four decades, MEMS pressure sensors have matured to significantly impact all commercial sectors. In 2005, the volume of MEMS pressure sensors exceeded 200 million units, dominating the unit volume and representing over 90% of units and 50% of revenues for pressure sensors in markets spanning automotive, medical, industrial, consumer, and military applications (Bryzek *et al.* 2006). A few of the more prominent automotive applications for pressure sensors include engine manifold monitoring, tire pressure monitoring, and both oil and brake fluid pressures. Monitoring of these parameters then improves the fuel economy and automobile safety. High-end automobiles are now equipped with more than 10 pressure sensors. Medical applications cover intraocular, intracranial, arterial pressure measurement, and disposable blood pressure sensors. Disposable blood pressure sensors became one of the most attractive medical applications to lower costs and maintain the level of sterility standards needed in heath care facilities. Other than automotive and medical applications, MEMS pressure sensors provide solutions to applications that demand small size, lightweight, low cost, and low-power consumption. Performance of MEMS pressure sensors and their interface circuitry have improved as manufacturing technologies advance optimizing their sensitivity, dynamic range, accuracy, linearity, and reducing temperature dependence. Device integration has enabled the ability to provide smart pressure sensors that are integrated on a monolithic substrate with their readout electronics. This chapter focuses on the developments that have been made in the realization and application of silicon-based pressure transducers spanning the basic diaphragm functionality to wireless system integration.

2.04.1.1 Historical Perspective

The first silicon micromachined pressure sensor was presented by Gieles (1969). At that time, the standard technologies were minisize pressure transducers that had freestanding diaphragms 2 mm in diameter and 30 μm in thickness. Gieles' device was implemented using microspark erosion and electrochemical etching to obtain a 5- to 25-μm-thick n-type epitaxial silicon membrane as shown in **Figure 1**. The membrane had embedded piezoresistors to sense deflection under pressure. The diameter of the membrane was defined by the size of the electrode used for microspark erosion. A microspark erosion process has been optimized to achieve a surface roughness on the order of 1 μm. To minimize the roughness, electrochemical etching was performed in a fluoric acid bath with a platinum electrode. The etch rate of the electrochemical etching is about 2 μm per minute. The device-to-device uniformity of this process showed a sensitivity variation within 5%.

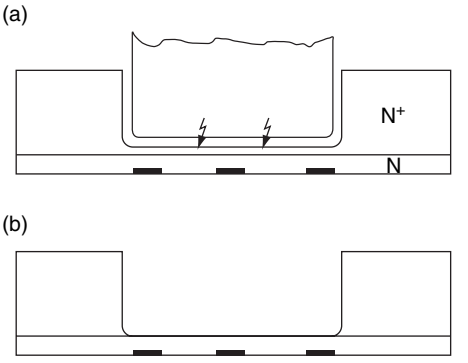


Figure 1 (a) Spark erosion machining and (b) electrochemical etching steps that formed the first micromachined pressure sensor presented by A C M Gieles at ISSCC in 1969. (Source: Gieles A C M 1969 Subminiature silicon pressure transducer. *International Solid-State Circuits Conference (ISSCC)*, Philadelphia, PA, USA, pp. 108–9.)

Another milestone in pressure sensor development was the merging of a sensor and its interface circuitry on a monolithically fabricated die. This was first implemented by Borky and Wise in 1980 with their development of a micromachined piezoresistive pressure sensor integrated into a triple-diffused bipolar circuit process (Borky and Wise 1979). **Figure 2** shows an integrated pressure sensor.

The single-crystal silicon diaphragm is defined by both anisotropic and isotropic chemical etches to achieve a nominal diaphragm thickness between 7 and 10 μm. Owing to the integrated electronics, the pressure sensor requires only two leads for the measurement and has a die size less than 1.2 mm.

2.04.1.2 Performance Parameters

There are a number of parameters that are needed to define the performance of micromachined pressure sensors. Parameters have been developed for

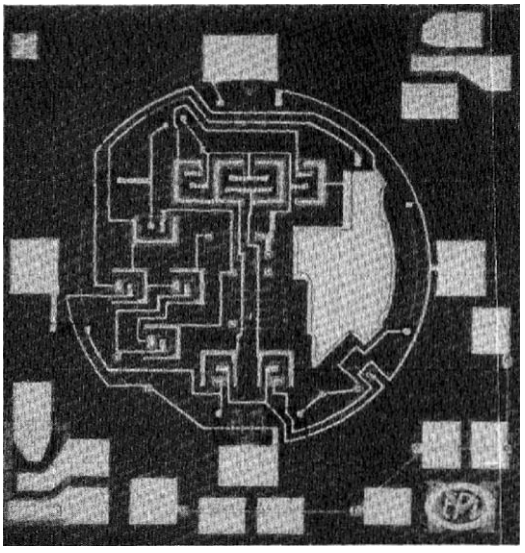


Figure 2 Piezoresistive pressure sensor integrated with signal conditioning circuits on a single chip developed by Borky and Wise. The transducer uses a circular diaphragm transducer with a diameter of 40 mils. (Source: Borky J, Wise K D 1979 Integrated signal conditioning for silicon pressure sensors. *IEEE Trans. Electron Devices* **ED-26**, 1906–10.

both the device and the system level to describe operation. At the device level, a transducer's performance is mainly defined by its sensitivity and dynamic range. The sensitivity of a device describes its conversion from pressure to an electrically measurable signal. Sensitivity is defined as:

$$S = \frac{1}{\vartheta}\frac{\mathrm{d}\vartheta}{\mathrm{d}P} \qquad [1]$$

where ϑ is the transducer's output, $\mathrm{d}\vartheta$ is the change in the output, and $\mathrm{d}P$ is the pressure change differential over which the output change occurred. The transducer's dynamic range states the functional pressure range for which a given set of parameters can describe the sensor's performance. To some extent, sensitivity and dynamic range can be applied at the system level as well, since merging of a transducer with its interface circuitry can affect these parameters. Higher-order parameters such as full-scale output (FSO), temperature coefficient of offset (TCO), and temperature coefficient of sensitivity (TCS) all apply at both the device and the system level. FSO describes the linearity of the output of a sensor by taking maximum percentage offset between the sensor's output and the linear fit of the data. A sensor's TCO is defined as:

$$\mathrm{TCO} = \frac{1}{\vartheta_\mathrm{o}}\frac{\partial\vartheta_\mathrm{o}}{\partial T} \qquad [2]$$

where T is temperature and ϑ_o is the offset extracted from a fit of the sensor's output. The TCO then defines the dependence of the extracted offset on temperature. A sensor's TCS is the dependence of the transducer's output sensitivity on temperature and is defined as:

$$\mathrm{TCS} = \frac{1}{S}\frac{\partial S}{\partial T} \qquad [3]$$

Two of the critical system-level parameters that define a sensor's performance are resolution and accuracy. A sensor's resolution defines the noise floor for a given transducer and its readout circuitry in producing a quantified data point. Sensor accuracy builds on the resolution by incorporating calibration information to determine the preciseness with which a sensor can determine pressure.

2.04.1.3 Pressure Ranges

Many different pressure units have been used in literature although the standard MKS unit is pascal. However, different applications have settled on different standards in their literature. To provide continuity between the various examples in the chapter, **Figure 3** shows a conversion chart between the various pressure units and the pressure ranges for different applications. Most developed micromachined pressure sensors fall into the 10–1000 kPa operating range to accommodate automotive and biomedical applications. Other transducer technologies that are explored in this chapter apply to measurement in a vacuum environment that typically operates in the range between 0.1 and 10 Pa.

Building on the basis of pressure sensor functionality that has been presented defining the parameters used to describe sensor performance, the remainder of the chapter focuses on the actual implementation at both the device and the system level. Section 2.04.2 covers the mechanics that govern the behavior of diaphragm-based transducers. This is then followed by a section that discusses the various transduction mechanisms that can be implemented with silicon-based sensors. Section 2.04.4 discusses issues involved in extending this technology to the actual applications. Section 2.04.5 then presents the chapter's conclusions.

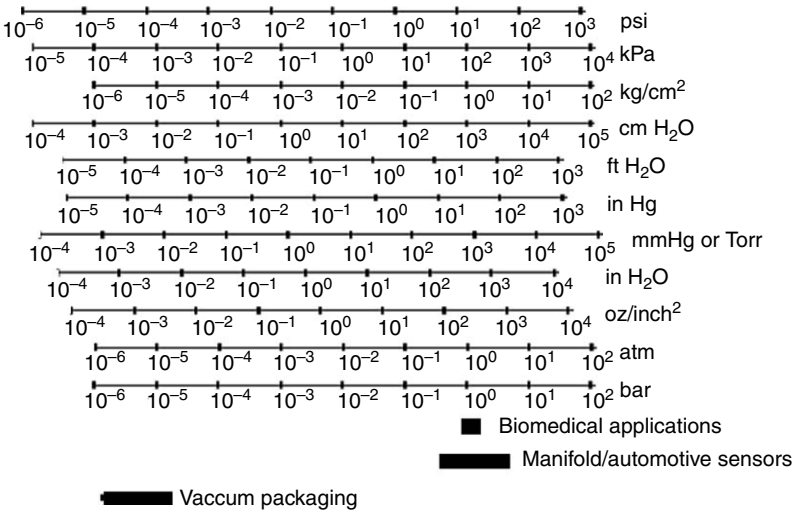


Figure 3 Various pressure units and application ranges.

2.04.2 Device Design

MEMS-based pressure sensors take a variety of forms in terms of fabrication technology and transducer type. However, a majority of the devices are based on the use of a deflectable diaphragm that enables the transduction. The stresses induced within a diaphragm and the corresponding displacement for an incident pressure can be modeled to incorporate the various levels of complexity involved in the device. The fundamental theory that governs the mechanics of diaphragm devices is put forth by Timoshenko (1959). This theory is developed for diaphragms in a variety of configurations and boundary conditions, from which MEMS-based sensor models can then be built.

2.04.2.1 Diaphragm Mechanics

The deflectable diaphragm model begins by defining the diaphragm incident shape and load of the device as well as the boundary conditions for diaphragm deflection. A generic diaphragm is described in **Figure 4**, where a is the radius, h is the thickness, ω is the deflection, with clamped edges supporting the diaphragm's perimeter. The deflection in the center of the diaphragm is defined as ω_c, which is the point of highest deflection for unbounded diaphragm movement.

For a clamped diaphragm under a uniform load, such as pressure, the boundary conditions can then be defined such that the angle of deflection, φ, is equal to zero at the center ($r = 0$) and at the edge ($r = a$) of

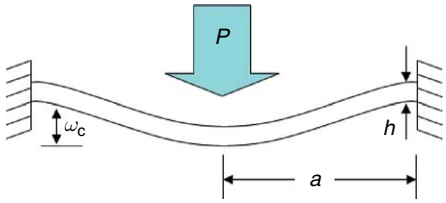


Figure 4 Generalized model for a deflectable diaphragm illustrating the clamped boundary conditions and the diaphragm design parameters.

the diaphragm. The model then builds in the geometry of the device including the layout and the resulting diaphragm thickness. The most straightforward diaphragm structures utilize either square or circular diaphragm designs, both of which have benefits and trade-offs in their implementation.

2.04.2.1.1 Square diaphragm

Due to the availability of anisotropic etchants for bulk silicon (Seidel *et al.* 1990, Tellier and Brahim-Bounab 1994), there are a number of fabrication processes that have been developed and tightly controlled to produce rectangular diaphragm sensors. The deflection of a rectangular diaphragm under incident pressure can be modeled by (Chau and Wise 1987):

$$\Delta P = \frac{Eh^4}{(1-\upsilon^2)a^4}\left[4.20\frac{w_c}{h} + 1.58\frac{w_c^3}{h^3}\right] \qquad [4]$$

where υ is the Poisson's ratio, E is the Young's modulus, and ΔP is the pressure difference across the diaphragm. This equation can be separated into two

regions for small and large deflections. For small diaphragm deflections, where $w_c < h$, the relationship between the differential pressure and the deflection at the center of the diaphragm is relatively linear. For large deflections, where $w_c > h$, the relationship is dominated by the second term in the equation and the pressure shows a third-order response with respect to the deflection of the center of the diaphragm. Utilizing a square or rectangular device will maximize the sensitivity of the diaphragm per unit area for a given sensor design.

2.04.2.1.2 Circular diaphragm

Continuing with the boundary conditions and approximations used in the models for square diaphragms, the deflection of a circular diaphragm's response to pressure, P, is defined by (Eaton and Smith 1997, Timoshenko 1959):

$$w(r) = \frac{3Pa^4}{16Eb^3}\left(1-v^2\right)\left[1-\left(\frac{r}{a}\right)^2\right]^2 \qquad [5]$$

More simply, the center deflection can be modeled in the small deflection region by (Timoshenko 1959):

$$w_c = \frac{3Pr^4}{16Eb^3}\left(1-v^2\right) \qquad [6]$$

Circular diaphragms do have the benefit of a more even stress distribution throughout the device since they eliminate the areas of high stress that occur in the corners of square diaphragms. Circular diaphragms can also provide the platform for building more advanced diaphragm-based devices such as the use of a center boss or operation of the sensors in the touch mode (Chavan and Wise 2001, Ko and Wang 1999). These additional features can then be used to further customize the response of the diaphragm to the needed specifications. From a layout perspective, the parameters available to the designer are only the length and the width of the diaphragm. However, a major factor in the design is developing the fabrication process used to build the device. The process will have the most dramatic effect on the sensor performance and variability in defining the thickness of the diaphragm due to its third-order relationship with the deflection. The diaphragm materials and fabrication processes used for the diaphragm will then define the Young's modulus, the Poisson's ratio, as well as the residual stress. These factors can be utilized to optimize the device performance, but also to introduce higher-order effects in the response of diaphragm-based devices.

2.04.2.2 Nonlinear Diaphragm Effects

The mechanical nonlinearities of a diaphragm need to be addressed to understand their impact on sensor performance. Sensor linearity is an essential parameter when addressing the device at a system level. Given that the transducer's interface circuitry must process the signal and transform it back into a pressure reading, closing this loop is where the linearity of the sensor will have a direct impact on the accuracy of a sensing system. In building on the basic models that have been presented, two main higher-order effects will impact the overall performance of the device: diaphragm stress and trapped gas effects. Diaphragm stress contributes both to the residual stress of the diaphragm's materials and to the stress effects that occur due to the bending and stretching of the diaphragm. The stretching of the diaphragm transforms the linear relationship between pressure and center deflection to a third-order relationship in eqn [4].

2.04.2.2.1 Residual stress

In general, the analytical modeling for diaphragm deflection is done with the assumption that there is no residual stress within the diaphragm. However, practically, fabricated devices all have some finite level of residual stress in the operational diaphragm. A few of the more prominent origins of a diaphragm's residual stress are intrinsic properties of the diaphragm material, thermal mismatch for silicon-on-glass devices, or stress from device packaging. The residual stress that remains on a functional diaphragm has the potential to drastically alter the device's overall performance. **Figure 5** shows the dependence of the diaphragm's sensitivity on the residual stress for both the compressive and the tensile cases.

The generalization of the two cases is that a tensile diaphragm will have decreased sensitivity, whereas a diaphragm left in compression will have increased sensitivity, but has the potential for buckling.

One of the predominate sources of residual stress in some of the bulk micromachined devices is the stress induced from the dopant that is used to form the etch stop in fabricating the device for either electrochemical or etch stop devices. Boron-doped silicon diaphragms exhibit large amounts of tensile

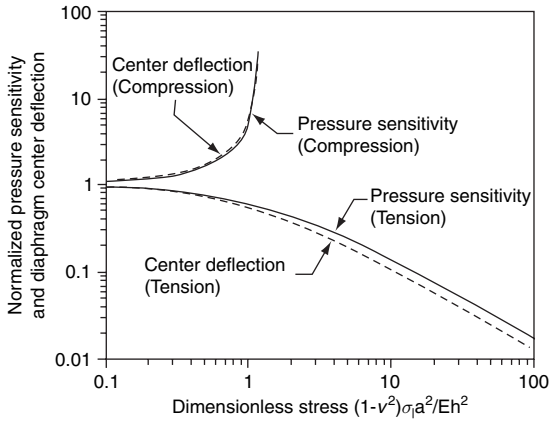


Figure 5 Normalized plot from Cho *et al.* showing the relationship between pressure sensitivity, diaphragm center deflection, and their dependence on intrinsic stress in both tensile and compressive diaphragms. (Source: Chau H-L, Wise K D 1987 Scaling limits in batch-fabricated silicon pressure sensors. *IEEE Trans. Electron Devices* **ED-34**, 850–8.)

stress ranging from 40 to 150 Mpa, the theory for which is that the doping causes irregularities in the lattice due to the smaller radii of the boron atoms when compared with silicon (Smirnov 1977). The residual diaphragm stress can then be compounded by the compressive stresses that develop from the use of high-temperature oxidation steps that drive in and activate the dopants. However, optimization of the fabrication process can use this combination of tensile and compressive stresses to stress-compensate the diaphragm for optimal performance.

2.04.2.2.2 Stress compensation

Stress compensation techniques can be used in device design by the selection of diaphragm materials and fabrication processes. Stacking different materials with both compressive and tensile layers enables building in the resulting stress that will govern the final device. However, attempting to achieve the zero-stress state has the potential to build in device-to-device variability depending on whether the residual stress in the diaphragm is slightly compressive or slightly tensile. Designing a moderate amount of tensile stress, between 5 and 50 MPa, does dampen the sensitivity in the diaphragm; it, however, produces more consistent results for device performance. Composite diaphragm stress can be modeled by taking into account the contribution from each material. The resulting stress of a composite diaphragm can be calculated as (Cho *et al.* 1992):

$$\sigma_c h_c = \sum_m \sigma_m h_m \qquad [7]$$

where σ_m is the composite stress, h_c is the thickness of the stacked layers, and σ_m and h_m are the respective stresses and thicknesses of the m layers in the stack. The ceiling that sets the maximum stress for a device is the fracture stress of the diaphragm material. This ceiling must include both the residual and the operational stresses for the device. For silicon, the fracture stress has been reported by Najafi *et al.* (1990) as 2000 MPa. Cho *et al.* (1992) found that the dependence of the diaphragm's sensitivity to capacitive sensors on the tensile intrinsic stress, σ_i, of a diaphragm can be defined as:

$$S_{\text{tension}} = \frac{a^2}{8\sigma_i hd} \qquad [8]$$

where d is the gap between electrodes. In general, a stress reduction factor of six can be utilized to increase the sensitivity of a diaphragm before membrane buckling becomes a serious concern (Cho *et al.* 1992).

2.04.2.2.3 Trapped gas effects

Trapped gas effects must be considered to assess the performance of absolute sensors that function with a sealed cavity. When combining a sealed cavity along with a deflectable diaphragm, a back-pressure effect is produced from the function of the sensor and the trapped gases temperature dependence. This effect can be modeled by factoring in the sealing pressure and temperature of the device to the TCO, which is defined as given below (Puers *et al.* 1990):

$$\text{TCO}_{\text{gas}} = -\left(\frac{P_{\text{seal}}}{T_{\text{seal}}}\right) \qquad [9]$$

and the TCS, which is defined by:

$$\text{TCS}_{\text{gas}} = -\frac{1}{2}\left(\frac{P_{\text{seal}}}{P_{\max}}\right)\left(\frac{1}{T_{\text{seal}}}\right) \qquad [10]$$

where $P_{\max}$ is the maximum pressure in the device's range, and P_{seal} and T_{seal} are the sealing pressure and temperature, respectively. The effects of the trapped gas can also result from the volume reduction of the sealed cavity due to the deflection of the diaphragm, which can be modeled with the ideal gas law providing matching within 7% compared with finite element modeling (Park *et al.* 2002). Utilizing a

vacuum cavity as the reference pressure for absolute sensors will minimize the effects of any trapped gas. This can be done using a combination of vacuum sealing of the cavity as well as processes that provide gettering of residual air in the cavity or from outgassing of the transducer materials. The residual gas in the cavity can be caused from the finite vacuum levels used during the sealing process. Even with completely sealed bond lines, outgassing has also been shown to occur through the cavity walls (Henmi *et al.* 1994). For sensors that are fabricated on glass substrates, the area of the glass exposure within the cavity should be minimized. This can be done by using an evaporated metal getter, such as titanium, to coat the majority of the exposed glass within the cavity and help maintain the integrity of the vacuum (Chavan and Wise 2002).

2.04.3 Transducer Technologies

There have been a wide variety of micromachined structures that have been designed to read out pressure using various different phenomena. Diaphragm-based devices have been the most pervasive, though these structures then branch out into various ways of converting their deflection into an electrical signal. The more common methods are piezoresistive and capacitive transducers. However, transduction of the diaphragm deflection through an optical readout proves to be advantageous for various applications. These types of devices are then complemented by resonant sensors and Pirani gauges that can be optimized to work in a vacuum environment. **Table 1** presents an overview of silicon micromachined transducer technologies and their system-level application. Although this table does not limit any of the technologies, it does highlight some of the strong points of the various transduction methods.

2.04.3.1 Piezoresistive Sensors

Piezoresistive transduction is one of the earliest demonstrations of mechanisms suitable for microdevices. Piezoresistive sensors are the most widely used sensors in industry today, partially due to their straightforward interface circuitry and the ease of process integration. Piezoresistive transduction utilizes resistive changes that occur due to the mechanical stress or strain of a material. The change in resistance comes from two components: one is by the physical change of the structures due to deformation, the other is by mobility change due to the quantum–physical phenomena (Geyling and Forst 1960). The former does not change the resistivity but changes the resistance while the latter changes resistivity. Strictly speaking, the former does not utilize piezoresistive effects but does add to the sensitivity of the device. The main effect on a thin-film metal resistor's change in response to strain is due to the shape of deformation and not the mobility change. In terms of transduction materials, the most commonly used piezoresistive material in MEMS is silicon due to its high gauge factor, which is the ratio of resistance change to dimensional change, as defined by Liu (2005):

$$\mathrm{GF} = \frac{\Delta R/R}{\Delta l/l} = \frac{\Delta R}{\varepsilon R} \qquad [11]$$

where R is the resistance, l is the length, and ε is the strain. The gauge factor is one of the most important parameters to design piezoresistive devices. Single-crystal silicon has a high gauge factor, which makes it well suited to enable high-sensitivity piezoresistive devices. However, the gauge factor of single-crystal silicon must also take into account the orientation dependence; if piezoresistors are placed on different orientations, a design needs to take into account the respective gauge factors. The π matrix piezoresistance coefficients, which can be found in material

Table 1 Overview of silicon micromachined pressure transduction technologies

Transducer type	*Enables monolithic integration with CMOS*	*Requires on-site circuitry*	*Enables ultralow power systems*	*Enables high-temperature measurement*	*Linear response*	*Enables vacuum measurement*
Piezoresistive	√				√	
Capacitive	√	√	√			
Resonant	√	√			√	√
Pirani gauge	√					√
Optical				√	√	

databases (Gad-el-Hak 2002), correlate stress and resistivity change as defined by:

$$[\Delta\rho/\rho] = [\pi][T] \qquad [12]$$

where ρ is resistivity, $\Delta\rho$ is the change in resistance, and T is the stress matrix. Therefore, the gauge factor is simply the product of piezoresistance coefficient and the Young's modulus.

Piezoresistive interface electronics often use the Wheatstone bridge configuration as shown in Figure 6. This configuration is particularly useful when a piezoresistive sensor has a differential piezoresistor configuration. For instance, when only one piezoresistor is present, the output V_x is defined by:

$$V_x = \left(\frac{\Delta R}{2R_s + \Delta R}\right) V_s \propto \Delta R \qquad [13]$$

which assumes $R_1 = R_s + \Delta R$, $R_2 = R_3 = R_4 = R_s$. The output is not linear to the change of the piezoresistors. On the other hand, for differential piezoresistors, the output V_x is defined as:

$$V_x = \left(\frac{\Delta R}{R_s}\right) V_s \propto \Delta R \qquad [14]$$

which assumes $R_1 = R_3 = R_s + \Delta R$, $R_2 = R_4 = R_s - \Delta R$. Thus, it is always preferable to have differential piezoresistors for a linear response. As mentioned in the previous sections, one of the drawbacks of the piezoresistive detection mechanism is the dependency on temperature. The dependence is minimized by using the resistors in the Wheatstone bridge configuration. Assuming $R_1 = R_2 = R_3 = R_4 = R$, the variation at the output can be approximated as:

$$\Delta V_x \approx \left(\frac{\Delta R_1 - \Delta R_2 + \Delta R_3 - \Delta R_4}{4R}\right) V_s \qquad [15]$$

As shown above, the temperature dependence of piezoresistors cancels each other out at least for the first-order approximation.

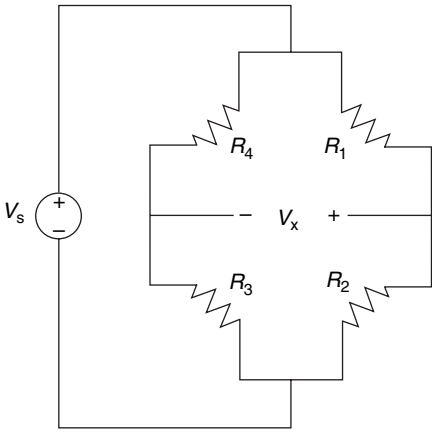


Figure 6 Wheatstone bridge configuration.

2.04.3.1.1 Single-crystal silicon piezoresistors

Piezoresistive pressure sensors using a single-crystal silicon diaphragm have been successfully researched and commercialized over the past four decades. One of the earliest works is a pressure sensor developed by Samaun *et al.* (1971) for biomedical instrumentation applications including cardiovascular catheterization. A 50-μm-thick silicon substrate was used to fabricate a single-crystal silicon diaphragm 1.2 mm in diameter and 5 μm in thickness as shown in Figure 7.

Four resistors are diffused into the diaphragm to form piezoresistors, and the thin diaphragm is formed by an anisotropic wet etch. They also developed a technique to precisely define the thickness of the membrane, within 1 μm thickness, which is shown in Figure 7(b). This technique uses self-terminating detents that are located on to top of the <100> oriented wafer and timed etch stops when the anisotropic etch from the bottom reaches the indicator. One of the critical limitations of piezoresistive-type devices is their sensitivity to temperature. In this work, the temperature sensitivity has been reduced by temperature compensation circuits. The output of the sensor is very linear and the size, critical to biomedical instrumentation applications, can be reduced to as small as 0.8 mm.

To build a completely stand-alone system, signal conditioning circuits are often integrated with pressure sensors. One of the earliest integrated MEMS device was developed by Borky and Wise (1979). A cross section of this device is shown in Figure 8.

This integrated pressure sensor needs only two leads, which makes the sensor very attractive for use in a flexible tube. The thickness of the single-crystal silicon diaphragm is 10 μm and bipolar triple-diffused technology has been used to fabricate the signal conditioning circuitry. The triple-diffused technology provides p+ diffusion as an isolation block, which is critical to the performance of the sensor. The isolation and integrated circuit configurations on the pressure sensor diaphragm are shown in Figure 8. The integrated pressure sensor has

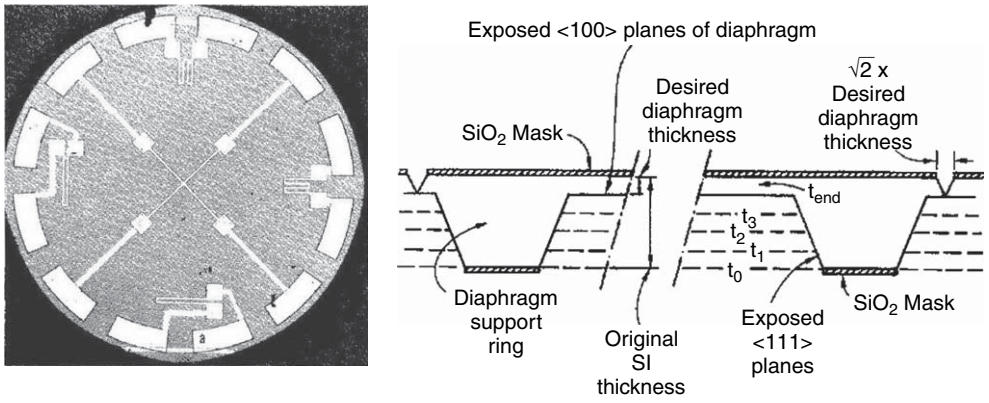


Figure 7 (a) Single-crystal silicon diaphragm device developed by Samaun *et al.* Measures 1.2 mm in diameter and 5 mm in thickness. (b) Process cross section. (Source: Samaun S, Wise K D, Nielsen E, Angell J 1971 IC piezoresistive pressure sensor for biomedical instrumentation. *IEEE International Solid-State Circuits Conference*, Philadelphia, PA, USA, pp. 104–5.)

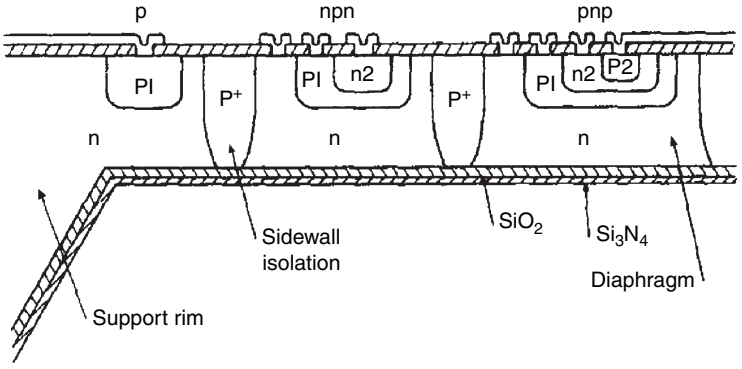


Figure 8 Cross section of the device developed by Borky and Wise in 1979 that integrated a pressure transducer and its interface circuitry on a deflectable diaphragm. (Source: Borky J, Wise K D 1979 Integrated signal conditioning for silicon pressure sensors. *IEEE Trans. Electron Devices* **ED-26**, 1906–10.)

1 mmHg resolution, 0–250 mmHg pressure range, and 1% linearity.

Piezoresistive elements can be easily formed in both bulk crystal silicon and deposited polysilicon. Early work on micromachined pressure sensors utilized piezoresistors to sense the induced strain of a thin membrane. Tufte *et al.* (1962) diffused doping impurities into a single-crystal silicon diaphragm to form piezoresistors as shown in **Figure 9**.

The 7/8″ silicon wafer is lapped and etched to be a thin membrane, ~183 μm, and the diffused strip is located in the middle of the wafer. Depending upon the piezoresistor's location on either convex or concave sides, the sensitivity of the diaphragm to external pressure varies; however, the output resistance change is fairly linear between 0 and 200 mmHg. Since this early work, forming piezoresistors on a single-crystal silicon diaphragm has been very popular for low-cost and low-temperature

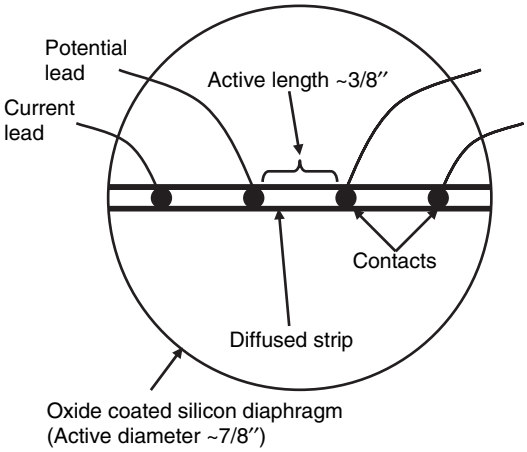


Figure 9 Single-crystal silicon piezoresistors. (Source: Tufte O, Chapman P, Long D 1962 Silicon diffused-element piezoresistive diaphragms. *J. Appl. Phys.* **33**, 3322–7.)

applications. Doping impurities used to define the piezoresistors can be embedded either by diffusion or by ion implantation. Taking advantage of a high gauge factor, single-crystal silicon piezoresistors have been well appreciated in many applications except high-temperature applications.

2.04.3.1.2 Polysilicon piezoresistors

Though single-crystal silicon diaphragm has a number of advantages including low cost, compactness, and a high piezoresistive effect, there are some disadvantages of single-crystal devices in that they suffer from higher temperature sensitivity and limited operating temperature. The temperatures are limited due to the p–n junction isolation scheme for which operating temperatures cannot exceed 120°C. By using polysilicon piezoresistor with silicon dioxide as an isolation layer, the limit for operating temperatures can be increased to over 200°C. One notable material parameter of polysilicon piezoresistors is that the piezoresistive effects of polysilicon is at a maximum 60–70% of the effect on single-crystal silicon.

Schafer *et al.* (1989) developed the polysilicon diaphragm piezoresistive pressure sensor that is shown in **Figure 10**. On a silicon substrate, 0.4-μm-thick polysilicon strain gauges are formed on top of a 0.2-μm-thick silicon dioxide. The polysilicon gauges are doped with boron by ion implantation, and annealed at a high temperature (between 950°C and 1100°C) for 30 min. After interconnecting the gauges by metal layers (W/Ti/Au), a membrane is formed by microspark erosion and a chemical etching from the backside of the wafer.

Piezoresistors are located at the maximum stress points on a membrane, either at the very center or at the edge. However, doing so may introduce high sensitivity to process variation. In order to minimize the sensitivity to any process variation, smart design and placements of piezoresistors have also been pursued. One suggested method is to distribute piezoresistors along the perimeter of the membrane and to have differential measurements to compensate any process variation.

In many cases where polysilicon piezoresistors are used, a polysilicon diaphragm is also implemented. Polysilicon is an attractive material to form a diaphragm for pressure sensors due to many advantages including great controllability of the film thickness, availability of a dielectric film for superior isolation, and suitability for high-temperature applications. As discussed in the previous sections, polysilicon offers a higher operating temperature than does single-crystal silicon for piezoresistive sensors. Availability of an excellent dielectric film such as silicon dioxide allows polysilicon to be used as a material for the diaphragm in piezoresistive sensors.

Kalvesten *et al.* (1998) presented a piezoresistive pressure sensor for cardiovascular pressure measurement (**Figure 11**). Although capacitive pressure sensors can achieve high sensitivity and low temperature drift they need on-chip detection circuitry, which may become difficult when fitting in 0.36 mm diameter of the guide wire to measure blood pressure inside the body. This is the major motivation for them to choose a piezoresistive detection mechanism for their device. The piezoresistive sensor has a polysilicon membrane with silicon dioxide as the sacrificial layer. The sensitivity of the pressure sensor is 2 $\mu V\,V^{-1}\,mmHg^{-1}$.

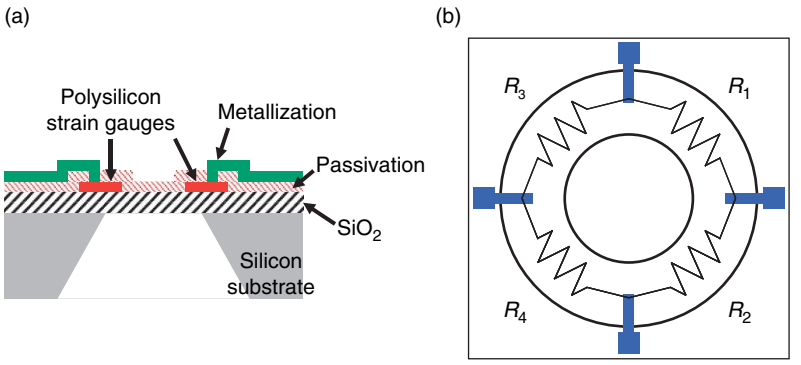


Figure 10 (a) Cross section and (b) equivalent circuit for the polysilicon diaphragm-based piezoresistive pressure sensor. (Source: Schafer H, Graeger V, Kobs R 1989 Temperature-independent pressure sensors using polycrystalline silicon strain gauges. *Sens. Actuators A*, **17**, 521–7.)

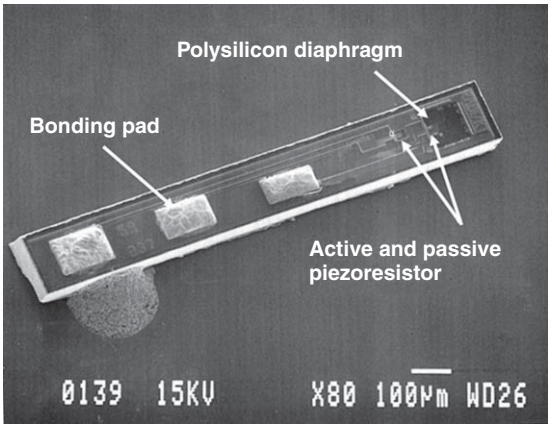


Figure 11 Pressure sensor using polysilicon diaphragm developed by Kalvesten *et al.* in 1998. (Source: Kalvesten J E, Smith L, Tenerz L, Stemme G 1998 First surface micromachined pressure sensor for cardiovascular pressure measurements. *Proc. 1998 IEEE 11th Annu. Int. Workshop Micro Electro Mechanical Systems*, January 25–9, 1998, Heidelberg, Germany, 1998, pp. 574–9.)

2.04.3.1.3 Other micromachined piezoresistive transducers

Although MEMS pressure transducers have mainly focused on the use of polysilicon and bulk silicon, there are other materials that can be integrated into transducers that enable beneficial device properties for some applications. Silicon carbide has many attractive features including excellent mechanical properties, chemical inertness, and radiation resistance, all suitable for harsh environment operation. Although polysilicon piezoresistors increase the operating temperature up to 200°C, many applications require higher operating temperature. For instance, monitoring combustion processes needs a pressure sensor operating at ~500°C. Wu *et al.* (2006) presented a bulk micromachined silicon carbide pressure sensor that is capable of operating up to 400°C. This work used poly-SiC as piezoresistors and silicon nitride as an isolation layer, as shown in **Figure 12**.

The low gauge factor of poly-SiC can be improved upon by a factor of nine by using 3C–SiC. The fabricated device was tested at 400°C, and showed the gauge factor drops by a factor of more than two, which indicates further developments are still needed to further optimize the device.

One of the drawbacks of a piezoresistive detection mechanism is smaller sensitivity compared with a capacitive detection mechanism. Pramanik and Saha (2006) introduced a porous silicon membrane to increase sensitivity by a factor of three over that of

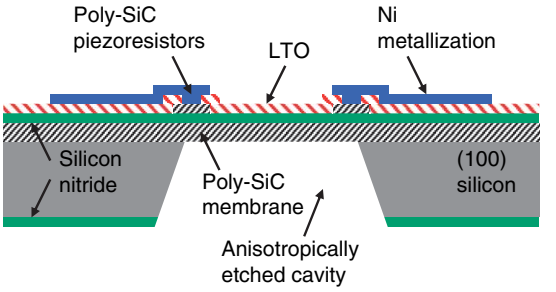


Figure 12 Fabrication process flow of a silicon carbide piezoresistors pressure sensor. (Source: Wu C, Zorman C, Mehregany M 2006 Fabrication and testing of bulk micromachined silicon carbide piezoresistive pressure sensors for high temperature applications. *IEEE Sens. J.* **6**, 316–24.)

a conventional bulk silicon membrane. **Figure 13** shows scanning electron microscope (SEM) images of the porous silicon. The porous silicon is formed by an anodic etching method in a two-pond cell, compatible with conventional micromachining techniques. The fabricated porous silicon (63% porosity) has piezoresistivity coefficient 50% higher than for single-crystal silicon, which is believed to be due to quantum confinement in the porous silicon nanostructure.

French *et al.* (1992) presented a silicon-on-oxide pressure sensor using a unique fabrication process to isolate piezoresistors by thermal oxide. **Figure 14** illustrates a cross section of the isolated piezoresistors. Instead of using a conventional silicon-on-insulator (SOI) substrate such as separation by implantation of oxygen (SIMOX), they have developed a unique fabrication process to completely isolate piezoresistors by thermal oxide and filled with polysilicon. The diamond-shaped cavity is first etched by hydrazine at 55°C and uses the <111> planes as the etch stop. Then, an oxide film is thermally grown to form silicon islands. This thermal oxide completely isolates silicon islands from the substrate. Finally, the rest of the cavities are filled with polysilicon to finish the fabrication. The fabricated device shows very linear fractional change in resistance ($\Delta R/R$) up to 300°C.

2.04.3.2 Capacitive Sensors

The implementation of silicon-based capacitive pressure sensors has been explored since the 1970s (Clark and Wise 1979, Frobenius *et al.* 1973, Fryer 1975). Overall, the use of capacitive transduction has many

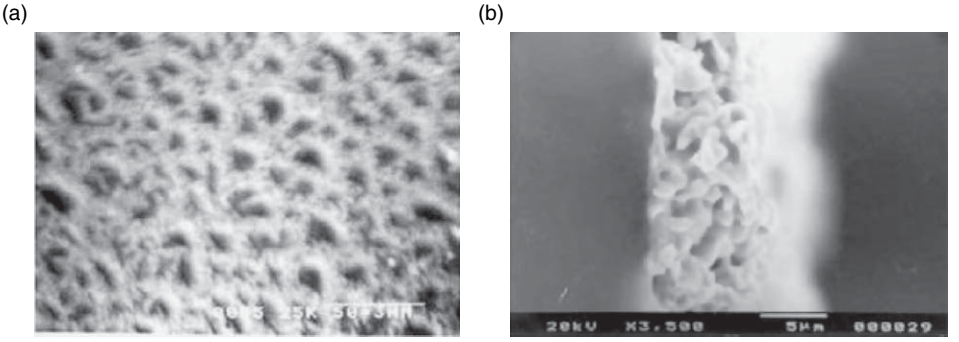


Figure 13 (a) Scanning electron micrograph (SEM) of porous silicon fabricated and (b) SEM of cross-sectional view of the porous silicon sample developed by Pramanik and Saha in 2006. (Source: Pramanik C, Saha H 2006 Piezoresistive pressure sensing by porous silicon membrane. *IEEE Sens. J.* **6**, 301–9.)

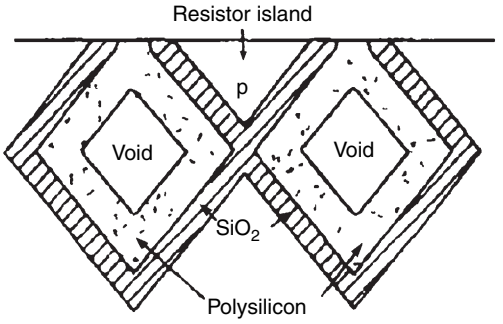


Figure 14 Silicon-on-oxide piezoresistors developed by French *et al.* in 1992. (Source: French P J, Muro H, Shinohara T, Nojiri H, Kaneko H 1992 SOI pressure sensor. *Sens. Actuators A Phys.* **A35**, 17–22.)

benefits from both a fabrication and a system design perspective. Capacitive sensors enable lower-power systems over their piezoresistive counterparts since the transducer itself consumes no active power. They also benefit from lower cross-domain sensitivity to temperature since their dependence is purely due to mechanical effects in the transducer structure. In general, capacitive transducers are diaphragm based and utilize the deflectable diaphragm as one of the electrodes of the sensors. There have been many perturbations of the transducer structure that optimize various aspects of its performance. An illustration of the basic transducer is shown in **Figure 15** along with configurations that utilize a bossed diaphragm and the diaphragm's operation in the touch mode.

A model for the transducer begins by defining its capacitance as:

$$C = \frac{\varepsilon_0 \varepsilon_r A}{d} \qquad [16]$$

where ε_0 is the permittivity of free space, ε_r is the relative permittivity, A is the area of the capacitor, and d is the gap between the electrodes. When configured using the deflected diaphragm as one of the electrodes, the capacitance must be integrated over the effective area of the electrodes. Integrating over the electrode then defines the capacitance as (Eaton and Smith 1997):

$$C = \int\int \frac{\varepsilon_0 \varepsilon_r}{d - \omega(r)} r \mathrm{d}r \mathrm{d}\theta \qquad [17]$$

where $\omega(r)$ is the diaphragm deflection across the diaphragm determined from the diaphragm mechanics, which is presented in eqn [5] for the circular case.

Inherent in the device structure is the trade-off between the device sensitivity and the needed dynamic range for the application. One of the main advantages of capacitive sensors is their ability to change their capacitance by 50–100% over their dynamic range (Chavan and Wise 2001), which contrasts with the typical 5% seen in piezoresistive devices (Puers *et al.* 1990). Since their early stages of development, capacitive transducers have claimed a 20-fold increase in implementation (Sander *et al.* 1980) over piezoresisitive sensors with equivalent gauge factors that are 5–10 times greater than silicon strain gauges (Frobenius *et al.* 1973). However, the benefits of the high response of capacitive transducers have to be weighed against the increased nonlinear response and dynamic range limitations. In addressing the nonlinearity of the flat diaphragm, Crescini *et al.* (1997) designed a snail shell-shaped bottom electrode in a capacitive transducer. This layout then tapers the response toward the periphery of the diaphragm and reduces the linearity error from

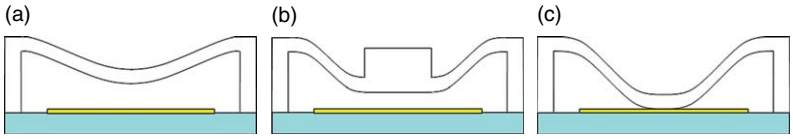


Figure 15 (a) Flat; (b) bossed; and (c) touch mode diaphragm structure and functionality for implementing capacitive transducers.

12% to 2% when compared with a standard electrode. However, this type of linearity compensation does have a trade-off with the overall sensitivity of the transducer. In optimizing these trade-offs, the device can be designed for use in the touch mode or by building in a boss in the deflectable diaphragm.

2.04.3.2.1 Bossed diaphragm sensors

The diaphragm mechanics with the implementation of a boss needs to then satisfy the boundary condition of having the angle of deflection equal to zero at $r = c$, where c is the radius of the center boss. The relationship between the pressure and the center deflection, which is then the deflection of the entire bossed area, is defined by (Giovanni 1982):

$$P = \frac{Eb^3}{A_p a^4}\omega_c + B_p \frac{Eb}{a^4}\omega_c^3 \qquad [18]$$

where:

$$A_p = \frac{3(1-v^2)}{16}\left(1 - \frac{c^4}{a^4} - \frac{4c^2}{a^2}\log\frac{a}{c}\right) \qquad [19]$$

$$B_p = \frac{\frac{7-v^2}{3}\left(1+\frac{c^2}{a^2}+\frac{c^4}{a^4}\right) + \frac{(3-v)^2 c^2}{(1+v)a^2}}{(1-v)\left(1-\frac{c^4}{a^4}\right)\left(1-\frac{c^2}{a^2}\right)} \qquad [20]$$

For a given fabrication process, the response of the bossed diaphragm can be controlled by defining both the radius of the boss and the area of the diaphragm. Chavan and Wise (2001) have implemented an array of these structures where the dynamic range of individual sensors in an array has been designed to have peak sensitivities for the various devices on the chip in a different pressure range. The transducers were realized using a dissolved-wafer, silicon-on-glass process. The diaphragms were defined using deep and shallow boron diffusions to form a concentration-dependent etch stop. The resulting deflectable part of the diaphragm is 3.5 μm and has a linear tape-open (LTO) layer that provides stress compensation as well as electrical isolation. This type of array then enables a system that maintains a resolution of 25 mtorr over the 300 torr dynamic range, with each of the segmented sensors in the array having a sensitivity of 24 fF/torr with a 50 torr dynamic range. A high level of sensitivity was achieved using a combination of pre-etching the cavity to a depth of 12 μm and then vacuum sealing the device to provide a functional gap between 300 and 800 nm. **Figure 16** shows a photograph of this device.

2.04.3.2.2 Touch mode sensors

Touch mode operation provides another option to enhance the linearity of capacitive transducers. If the distance between the diaphragm and the corresponding electrode is designed correctly and a passivation layer is used between the two, there are many benefits with operation of the device in the touch mode. The benefit of having a close to zero gap between the electrodes can be used to enhance the overall device sensitivity in the touch mode region. Ko and Wang (1999) reported the performance of circular diaphragm touch mode sensors with a sensitivity of 86 fF per psi with a sensor that has a diaphragm diameter of less than 400 μm. **Figure 17** shows the capacitive response across the various regions of operation of a capacitive transducer. The normal region is where the capacitive response is governed by the third-order equations defined in Section 2.04.2.1. The transition region is where the diaphragm first touches down. The linear region then shows the desired response of operation of the sensor in the touch mode. Guo *et al.* (2000) have reported implementing touch mode sensors with an integrated capacitance-to-frequency conversion interface circuit to measure 0.63% nonlinearity in the sensor readout across the linear touch mode region of operation. The saturation region is where the major part of the diaphragm touches the opposite electrode and only fringing effects remain. Modeling of the actual diaphragm itself then needs a more complex definition since the boundary condition defining where the angle of deflection equals zero is changes throughout the dynamic range of the sensor. This type of mechanical modeling is more suited for the use of finite element modeling or spreadsheet tools, which can provide designs that match the transducers response within 5% (Meng and Ko 1999).

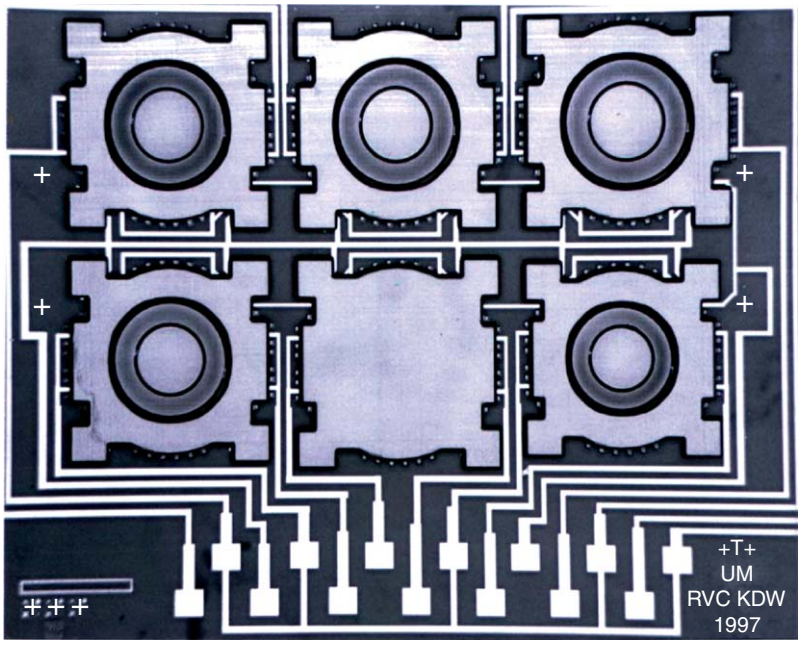


Figure 16 Top-side view of the bossed diaphragm capacitive sensor array that uses a five-transducer array along with a reference capacitor. (Source: Chavan A V, Wise K D 2001 Batch-processed vacuum-sealed capacitive pressure sensors. *IEEE J. Microelectromech. Syst.* **10**, 580–8.)

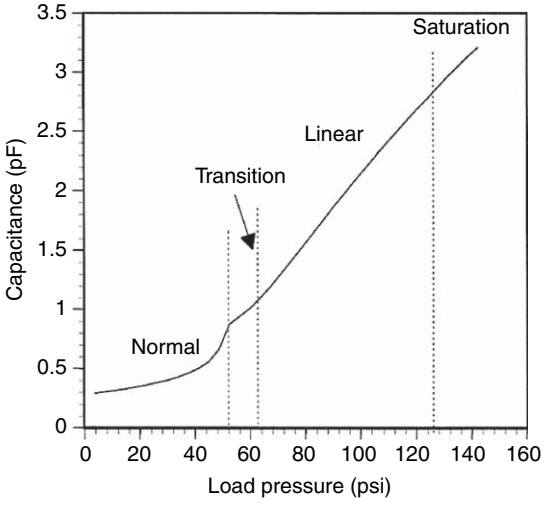


Figure 17 Functional regions of a diaphragm-based capacitive transducer. This plot highlights the linearity that is achieved operating the transducer in the touch mode region. (Source: Ko W H, Wang Q 1999 Touch mode capacitive pressure sensors. *Sensors and Actuators A (Physical)*, **A75(3)**, 242–51.)

2.04.3.2.3 Device noise and scaling limits

Fundamentally, the scaling of a sensor is limited by its noise floor and the sensitivity trade-offs incurred when scaling the size. Capacitive pressure sensors are exponentially dependent on the equivalent radius of the diaphragm. This then makes capacitive sensors application specific in addressing the needed signal-to-noise ratio that defines their scaling limit. Scaling limits for capacitive pressure sensors have been developed to address both device- and system-level issues. The front end of the sensing noise floor was defined by Chau and Wise (1987) as:

$$\overline{p_n^2} = \frac{kT\left(C_0 + C_p\right)}{\left(V_s C_0 S_{cap}\right)^2} \qquad [21]$$

where C_0 is the sense capacitance, C_p is the parasitic capacitance, S_{cap} is the capacitive sensitivity, V_s is the supply voltage, k is the Boltzmann's constant, and $\overline{p_n^2}$ is the mean square input pressure noise. For some applications, the transient response of the diaphragm must be considered both in the time and in the frequency domains. Diaphragm vibrations from the sampling frequency of the interface circuitry can occur if the sampling frequency is near one of the various modes of vibration for the diaphragm. The device noise does provide some detail as to the achievable resolution and accuracy of a device; however, a system-level analysis must address the noise and sensitivity coupling the device with its readout circuitry. The system-level interface circuitry can

also implement noise reduction techniques such as correlated double sampling, which lowers the $1/f$ noise of the amplifier offset and thus further optimizes the performance of the device (Chavan and Wise 2002).

2.04.3.2.4 Bulk diaphragm fabrication for capacitive sensors

Fabricating of devices from bulk silicon leverages the various isotropic and anisotropic micromachining techniques that have been developed. It also allows the use of the crystalline structure of silicon in forming the device. Various silicon-etching methods have been implemented to form micromachined pressure sensors, for many of which process simulation tools that go down to the molecular level of interaction with bulk silicon have been investigated and developed (Elwenspoek 1993, Kakinaga *et al.* 2005, Seidel *et al.* 1990). Overall, the goal is to achieve a thin, flexible, and reproducible silicon diaphragm. The three predominant methods of achieving this are the use of concentration-dependent etch stops, material- or junction-isolated etch stops, or a timed etch. Deep reactive ion etching (DRIE) is an alternative to forming for micromachining silicon structures and would fall into the category of a time-dependent etch stop.

The use of concentration-dependent etch stops enabled the silicon-on-glass structure developed by Chau and Wise (1988). The paper describes the used of sequential deep and shallow boron diffusions to define the diaphragm structure into a lightly doped silicon wafer. This silicon wafer is then anodically bonded to a patterned and metallized glass wafer. The diaphragm is then etched out using ethylene-diamine pyrocatechol (EDP) to dissolve the bulk silicon and leave only the highly boron-doped diaphragm, the metal interconnect, and the glass substrate. This method can achieve a high level of device-to-device consistency and provides a device structure that has been developed further to provide sensor circuit integration. The repeatability is mostly a factor of the EDP having an etch rate of lightly doped silicon, which is 50 times faster than the etch rate of a boron-diffused etch stop at concentration greater than $7 \times 10^{19}\,\mathrm{cm}^{-3}$.

Kloeck *et al.* (1989) utilized a reverse-biased p–n junction to define a diaphragm with an electrochemical etch stop. This device uses a <100> silicon wafer and a 40% potassium hydroxide (KOH) aqueous solution. This process has the potential to achieve a diaphragm tolerance of $\pm 0.2\,\mu\mathrm{m}$ (Kloeck *et al.* 1989). The development of an *in vivo* pressure sensor using a junction-isolated etch stop was done by El-Bahar in 1998 (El-Bahar *et al.* 1998). For this device they used TMAH to do the etching of the p-doped silicon and stop on the n-type junction. This etch then stops when the etchant reaches the biased n-type silicon, which is then passivated by an anodically formed oxide. This passivation prevents further etching of the silicon. Since an electrical setup is needed in the etch solution as well as in a sealed chamber that exposes only one side of the silicon wafer, the setup of this etch can be somewhat complicated. However, this release method does provide an opportunity to reduce intrinsic stress over methods that use a concentration-dependent etch stop, which has a high level of tensile stress in the silicon.

Timed etching of silicon has been implemented to form pressure sensors in various forms. It has been used by itself both as a dry and as a wet etch. It has been used for the pre-etch to form the recess in a silicon wafer as well as in the release step when etching all the way through the substrate and defining a sensor diaphragm. KOH has been used as the pre-etch for square diaphragms in silicon because as a timed etch its etch depth can be carefully controlled and its anisotropic etching properties provide geometrical control of the etch (Tellier and Brahim-Bounab 1994). DRIE has also been implemented as the recess etch for concentration-dependent etch processes. The use of DRIE also provides the ability to machine large corrugations in a diaphragm to further optimize the device (Chen *et al.* 2003).

2.04.3.2.4.(i) Wafer bonding It is typical to use a secondary substrate wafer for bulk micromachined devices since in most cases the majority of the silicon is removed from the primary wafer. The bonding of two wafers together provides a batch fabrication method that enables these types of structures. Since pressure sensors need the formation of a hermetically sealed cavity, inorganic materials are typically preferred over some of the polymer bonding options. Processes that bond two silicon wafers as well as bond a silicon wafer to a glass wafer have been developed (Schmidt 1998). In the case of much of the piezoresistive work, this secondary silicon wafer is used to form the sealed cavity. This bond can be formed either directly, with methods such as silicon fusion bonding, or indirectly with the use of an interstitial layer such as gold or glass that will form a eutectic or anodic bond to attach the two wafers. For direct silicon fusion bonding, there are strict

requirements needed for each of the polished silicon surfaces needing a roughness of below 10 Å and a bow of less than 5 Å (Schmidt 1998). The surface of the wafers can also be treated and cleaned to improve the quality of the bonding surface. The two silicon wafers are then elevated above 1000°C to create a bond that is comparable in strength to the silicon itself. The use of an intermediate layer when bonding two silicon wafers does relax the surface and temperature requirements to some extent. Silicon–gold eutectic bonding, when combined in the correct ratio, takes advantage of the drop in the reflow temperature to 363°C (Wolffenbuttel and Wise 1994).

Anodic bonding between a silicon and a glass wafer has been implemented for various transducer types and presents advantages both in device performance and in packaging and manufacturability of the device. Many of these are built into devices that have been implemented with the dissolved wafer process, which basically merges anodic bonding and a concentration-dependent etch in EDP. The glass substrate can then serve as a substrate for the electrical connections. These can be accessed either by wire-bonding to the routed metal traces or by more novel ways such as creating through-wafer vias in the glass substrate, which can also serve to facilitate the lead transfer to the sealed cavity (Chae *et al.* 2005a, Esashi *et al.* 1990).

2.04.3.2.4.(ii) Vacuum sealing and lead transfer The ability to form a vacuum-sealed cavity in the device enables the reference needed for absolute sensors as well as minimizes trapped gas effects on the device's performance. Various MEMS-based sensors have been developed to include a vacuum-sealed cavity and provide absolute pressure measurement. The technology that provides vacuum sealing at the wafer level also has the potential to reduce the overall device costs and provide device-to-device sameness across a batch of sensors. Chavan and Wise (2001) developed a bulk micromachined capacitive transducer that is fabricated using a silicon-on-glass dissolved wafer process. In order to enable vacuum sealing and lead transfer to the opposite electrode on the glass, a polysilicon-to-glass anodic bonding process was developed. Implementing this process needs a reduction in the surface roughness of the deposited polysilicon layer. Chemical–mechanical polishing was used to reduce the surface roughness below 50 nm and to enable a hermetic seal for the vacuum cavity. Using a polysilicon bond, the glass contact is routed in the bonded polysilicon layer, while contact to the bulk silicon is made outside the sealing ring. **Figure 18** illustrates lead transfer in the two cross sections of the device structure.

Park and Gianchandani (2003) developed a novel way to eliminate the need to provide electrical interconnect into a vacuum cavity with their implementation of the skirt electrode sensor. For this device, the fabrication process is simplified to three masks and the capacitive transduction takes place around the perimeter of the sealed cavity. **Figure 19** shows the mechanics of the structure that provide the sensor's output from a small diaphragm deflection by the touch mode operation of the device.

2.04.3.3 Resonant Sensors

Resonant structures have been implemented for pressure measurement by taking advantage of the following:

(1) Q, the quality factor, of the resonant structure changes
(2) Tension of the resonator changes over the pressure

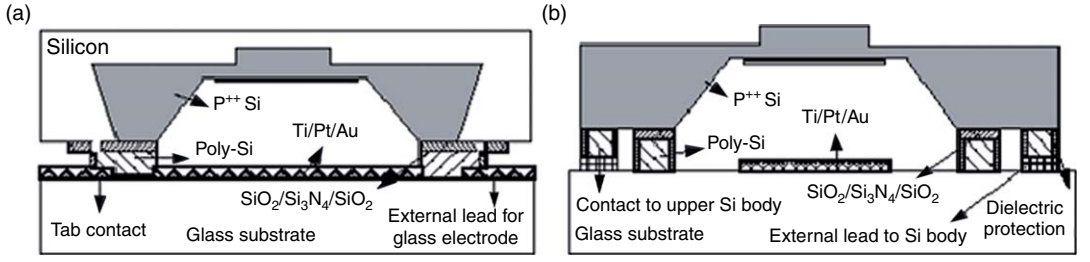


Figure 18 Cross section of the absolute sensors implemented by Chavan and Wise showing (a) the continuous polysilicon seal the perimeter of the device and (b) the isolation of the glass and silicon interconnect. (Source: Chavan A V, Wise K D 1997 A batch-processed vacuum-sealed capacitive pressure sensors. International Conference on Solid-State Sensors and Actuators. Digest of Technical Papers, Transducers 97', 1997, pt. 2, vol. 2, pp. 1449–52.)

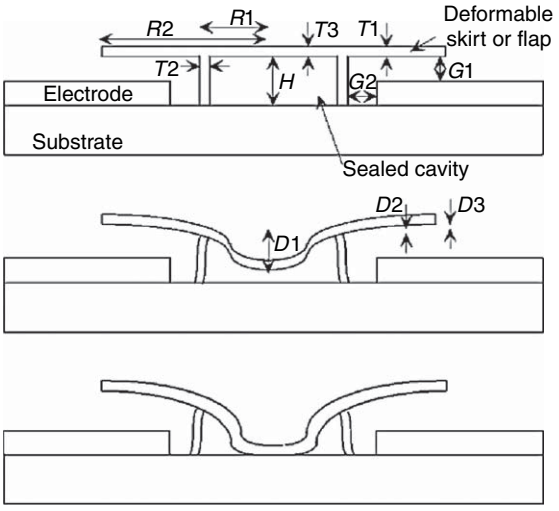


Figure 19 Capacitive skirt electrode structure implemented by Park and Gianchandani eliminates the need for electrical interconnect in a vacuum-sealed capacitive transducer structure. (Source: Park J-S, Gianchandani Y B 2000 A capacitive absolute-pressure sensor with external pick-off electrodes. *J. Micromech. Microeng.* **10**, 528–33.)

Resonant pressure sensors provide higher sensitivity than do piezoresistive sensors, and also offer better accuracy and stability, especially temperature stability. In this section, we first discuss the resonant pressure sensor using the Q factor and then a sensor utilizing the tension change over the applied pressure.

The quality factor (Q) of a resonant structure is highly dependent on its surrounding air pressure. The quality factor remains constant at a very low pressure as an intrinsic Q based on the device structure. The quality factor, Q, then begins to roll off at a certain pressure as shown in **Figure 20** (Newell 1968). The dependence Q has on air damping is due to individual air molecules exchanging their momentum with the resonant structure and inducing energy dissipation, which then results in a reduction of Q. The Q reduction stops when the molecules behave more as a viscous film, which is independent of pressure. Of course, one may want to use resonant structures as a pressure sensor at the momentum-damping region.

In the region, where Q is dependent on the momentum damping the air molecules with a temperature of 300 K, Q can be approximated by (Newell 1968):

$$Q \approx 93\left(\frac{d}{L}\right)^2 \frac{\sqrt{Y\rho}}{P} \qquad [22]$$

where d and L are the thickness and the length, respectively, of a resonator; Y is the Young's modulus; and ρ is the mass density.

By using the pressure sensitivity of microresonators, devices have been developed to include pressure measurements of enclosed cavities. Chang-Chien and Wise have demonstrated comb-drive resonators to monitor pressure integrity of micropackages (Chang-Chien 2002). The Q of the resonator versus the pressure characterization curve is shown in **Figure 21**. Using the characteristic curve, they evaluated micropackage vacuum integrity to maintain vacuum pressures below 50 mtorr.

Instead of utilizing fluidic damping effects (determining Q of a resonator), many micromachined resonators have been used as pressure sensors by taking advantage of the structural tension coupled with a diaphragm (Burns *et al.* 1995, Ikeda *et al.* 1990, Petersen *et al.* 1991, Stemme and Stemme 1990). One of the earliest examples used single-crystal silicon resonators etched by anisotropic etchants for absolute pressure sensors (Greenwood 1984).

Figure 22 shows the resonator and the response of the resonator over the pressure ranging from 0 to 60 kPa. The resonant frequency of a stretched string changes with the square root of the tension, and tension is proportional to the applied pressure; thus, the resonant frequency of the resonator varies with the square root of the pressure. The resonator is excited electrostatically on one side, with a bias voltage of 50 V and a drive signal of less than 1 V, and sensed by field effect transistor (FET) on the other side.

In order to minimize Q reduction due to damping by surrounding fluids, vacuum-sealed integrated resonators with differential measurements have been introduced (Ikeda *et al.* 1990). Ikeda *et al.* have developed a magnetically excited vacuum-sealed integrated resonant pressure sensor as shown in **Figure 23**.

The resonant beams are integrated in a diaphragm, one at the center and the other at the corner. Since these resonant beams are sealed at the vacuum, the pressure sensor has no Q reduction and no need to optimize the amplitude response of the resonator over the pressure measurement. The resonant frequency of integrated resonant beams can be described as:

$$f = \frac{4.73^2 h}{2\pi l^2}\left(\frac{E}{12\rho}\left(1 + 0.2366(l/h)^2\varepsilon\right)\right)^{1/2} \qquad [23]$$

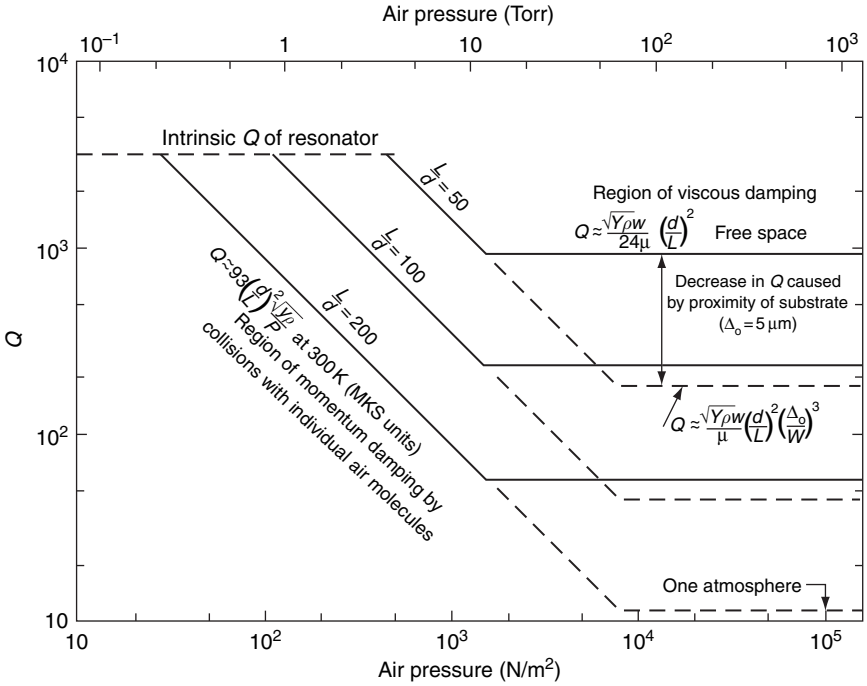


Figure 20 Quality factor (*Q*) of a resonator and its dependence on air pressure. (Source: Newell W 1968 Miniaturization of tuning forks. *Science* **161**, 1320–6.)

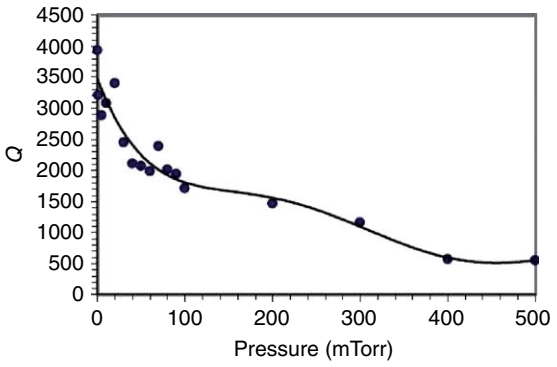


Figure 21 Quality factor (*Q*) versus pressure for a polysilicon comb-drive resonator. (Source: Chang-Chien P P L, Wise K D 2001 Wafer-level packaging using localized mass deposition. *Proc. 11th Int. Conf. Solid State Sensors and Actuators Transducers '01/Eurosensors XV.* Digest of Technical Papers. Springer-Verlag, Munich, Germany, June 10–14, 2001, Vol. 1, pp. 182–5, BN – 3 540 42150 5.)

where f is the natural frequency of a fundamental oscillating mode, l is the resonator length, h is the resonator thickness, E is the Young's modulus, ρ is the density, and ε is the strain.

The strain caused by pressure changes the resonant frequency of the resonant beams, and by differential measurement between center and corner gauges the maximum fitting error is only about 0.05% of the full-scale range. The temperature coefficient is also measured using the integrated resonators as approximately $-40\,\text{ppm}\,\text{K}^{-1}$, which is nearly equal to the theoretical value that is half the sum of the thermal expansion coefficient and the temperature coefficient of elasticity.

Kim *et al.* (2005) demonstrated the stability of microresonators over time and thermal hysteresis. Their microresonators are encapsulated by a thick ($\sim$20 μm) epitaxial polysilicon film in a very controlled environment, and were monitored for performance change over time and thermal hysteresis. Over 8000 h of operation, the resonant frequency was maintained with less than $0.5\,\text{ppm}\,\text{year}^{-1}$ drift, which is equivalent to that of a temperature compensated crystal oscillator (TCXO). For thermal cycling between $-50°\text{C}$ and 80°C for more than 400 cycles, the frequency stability is below 0.77 ppm.

2.04.3.4 Pirani Sensors

Pressure sensors typically involve physical movement of a thin diaphragm or minute structures. They generate very accurate and sensitive signals; however, moving structures are sometimes not

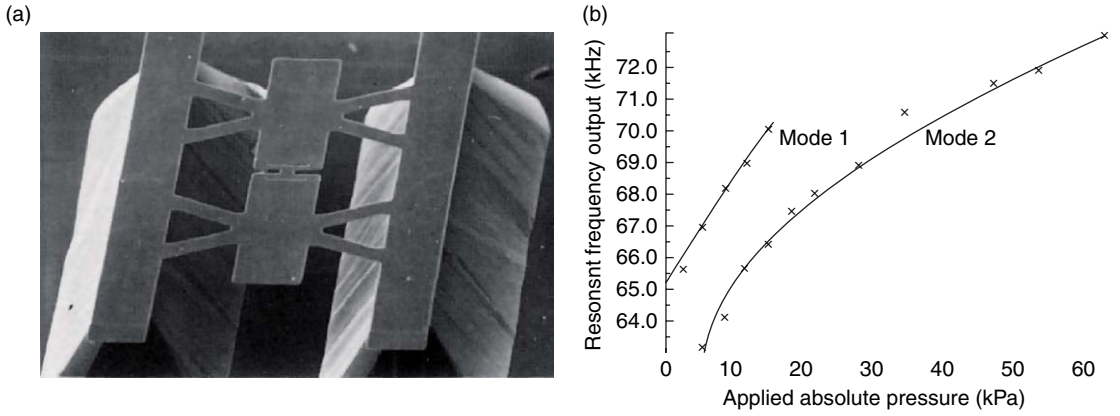


Figure 22 (a) SEM image of a micromachined resonator implemented for frequency-sensitive pressure transduction; and (b) its resonant frequency change over absolute pressure. (Source: Greenwood J C 1984 Etched silicon vibrating sensor. *J. Phys. E (Sci. Instrum.)* **17**, 650–2.)

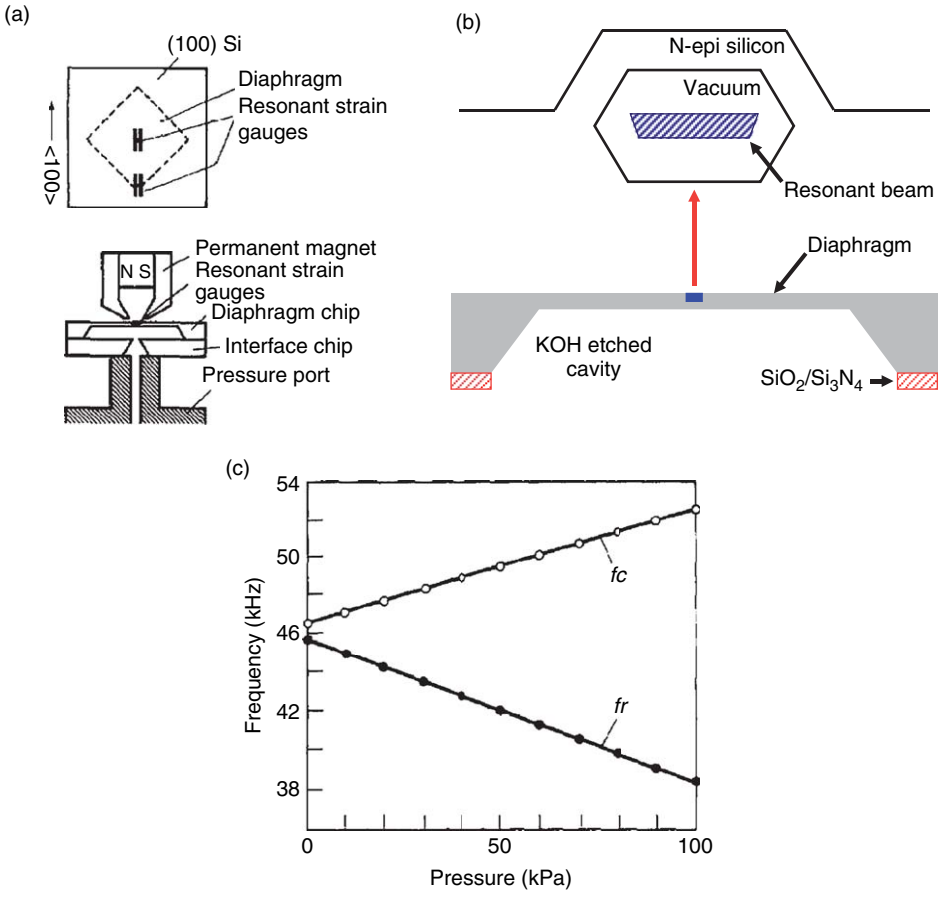


Figure 23 Vacuum-sealed integrated resonant pressure sensor: (a) Top-side view; (b) vacuum-sealed resonant beam in a diaphragm; and (c) output frequencies (f_c, center gauge frequency; F_r, corner gauge frequency) versus pressure. (Source: Ikeda K, Kuwayama H, Kobayashi T, Watanabe T, Nishikawa T, Yoshida T, Harada K 1990 Silicon pressure sensor integrates resonant strain gauge on diaphragm. *5th Int. Conf. Solid-State Sensors and Actuators and Eurosensors III*, *Sens. Actuators A, Phys. (Switzerland).* Montreux, Switzerland, Vol. A21, pp. 146–50.)

preferred for certain applications where structural moving causes undesired effects; e.g., cross-over distortion, harmonics. A Pirani sensor has no moving parts to transduce pressure and instead measures thermal conductance changes of the ambient. Since Pirani sensors were invented in 1906 (Pirani 1906), many miniaturized Pirani sensors that take advantage of micromachining technology have been introduced.

The operation of a Pirani gauge is based on heat transfer from a suspended heater to a heat sink through a gas. The thermal conductance through the gas is a function of its pressure. Depending on the Knudsen number, K_n, the gas can be modeled in the continuum regime ($K_n << 1$) at high pressure or in the molecular regime ($K_n >> 1$) at low pressure. An illustration showing these regimes of operation is shown in **Figure 24(a)**.

With reasonable assumptions and approximations, heat flux, which is a function of ambient pressure for all K_n, can be modeled as (Mastrangelo 1991):

$$\dot{Q}(P) \approx \dot{Q}(\infty)\left(\frac{P/P_o}{1+P/P_o}\right) \qquad [24]$$

where P_o is an empirical transition pressure. This $\dot{Q}(P)$ has a linear dependence on P at low pressure and is limited to a constant P_o at high pressure, which determines the upper limit of the dynamic range. Assuming that the heater is a lossless thermal conductor, the total thermal conductance of the heater is the sum combining solid conduction, gaseous conduction, gaseous convection, and radiation. The dependence of these forms of conduction on pressure is illustrated in **Figure 24(b)**. Radiation can be ignored at low temperature, and gaseous convection can be neglected because the Pirani gauge is usually placed inside a package where no external forced gas convection exists.

The sensitivity of the Pirani gauge is then the slope of the total thermal conduction (G_{Total}) versus pressure as shown in **Figure 24(b)**. The sensitivity can be obtained as:

$$\frac{\partial G_{gas}(P)}{\partial P} \approx \frac{G_{gas}(\infty)}{P_o}\frac{1}{(1+P/P_o)^2} \cong \frac{G_{gas}(\infty)}{P_o} \text{ for } (P/P_o) << 1 \qquad [25]$$

This configuration of fabricated micro Pirani gauges is shown in **Figure 25**. These gauges have two heat sinks instead of the conventional one heat sink configuration. The dual heat sink Pirani gauge is to improve sensitivity and dynamic range. Vertical configuration has a thin-film metal heater underneath of p++ silicon and glass substrate and p++ silicon as heat sinks whereas p++ silicon itself serves as a heater and as dual heat sinks for lateral configuration. The vertical sensor performs better than the lateral one as shown in **Figure 26**, but with added fabrication complexity of a six-mask process for the vertical structure over a two-mask process for lateral sensors.

These Pirani gauges are integrated inside a micropackage, size of 2 mm × 2 mm × 0.6 mm, to monitor package integrity over the time. The base pressure is ~33 torr without a getter inside the package, and the

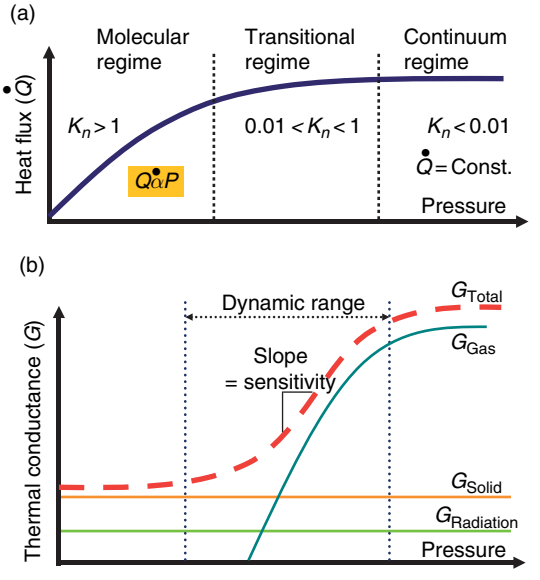


Figure 24 (a) Heat flux; and (b) thermal conductance versus pressure. (Source: Chae J, Stark B, Najafi K 2005 A micromachined Pirani gauge with dual heat sinks. *IEEE-CPMT Adv. Packaging Micro/Nano-Scale Syst.* **28**, 619–25.)

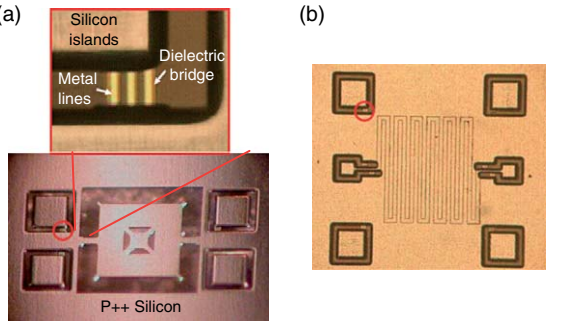


Figure 25 Photographs of configurations of fabricated Pirani gauges. (a) Vertical and (b) lateral. (Source: Chae J, Stark B, Najafi K 2005 A micromachined Pirani gauge with dual heat sinks. *IEEE-CPMT Adv. Packaging Micro/Nano-Scale Syst.* **28**, 619–25.)

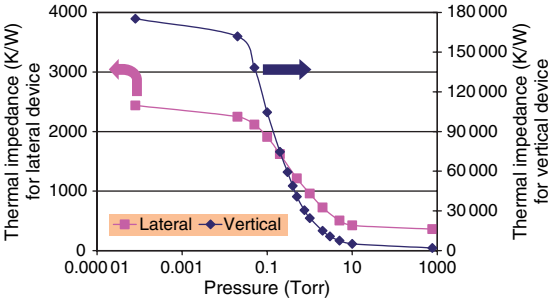


Figure 26 Pirani sensor characterization versus pressure for the lateral and vertical configuration. (Source: Chae J, Stark B, Najafi K 2005 A micromachined Pirani gauge with dual heat sinks. *IEEE-CPMT Adv. Packaging Micro/Nano-Scale Syst.* **28**, 619–25.)

pressure is maintained relatively constant for more than 4 months (Chae *et al.* 2005a).

2.04.3.5 Optical Sensors

The development of technology for the telecom industry has also been leveraged in various sensor applications. The use of fiber-optic technology has led to the advance of sensors systems that benefit from off-site interface electronics. Capacitive- and piezoresistive-based devices run into problems at higher-temperature applications ($T > 250°C$). These temperatures cause problems with the on-site circuitry for capacitive sensors and increase the leakage currents in bulk silicon piezoresistive sensors. Sensors that are interfaced through an optical fiber eliminate these issues since the electronics can be placed off-site and the sensor is independent of the electrical characteristics of the diaphragm. Pulliam *et al.* (2002) have developed a diaphragm-based device mounted on an optical fiber, which uses extrinsic Fabry–Perot interferometry. This device is designed to work in temperatures up to 600°C for applications within a turbine engine. As shown in **Figure 27**, this transducer detects deflection of the diaphragm by measuring the distance between the diaphragm and the end of the optical fiber.

The readout for this type of device can use either a single- or a dual-wavelength system for interrogation. Optical readout of the diaphragm deflection can be interrogated using a single wavelength of light; however, this interface suffers from nonlinearity issues when the diaphragm displacement needs a dynamic range of $\lambda/2$ due to the sinusoidal interference pattern. Pulliam *et al.* (2002) have implemented a readout for this device that uses a dual-wavelength

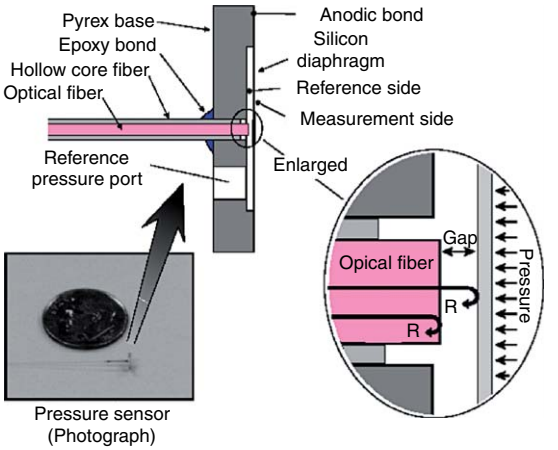


Figure 27 Illustration of the functionality of an optical transduction scheme that implemented a silicon diaphragm along with a fiber optic-based optical interface converting diaphragm deflection into a Fabry–Perot interference pattern. (Source: Pulliam W, Russler P, Fielder R 2002 High-temperature, high-bandwidth, fiber optic, MEMS pressure sensor technology for turbine engine component testing. *Proc. SPIE – The International Society for Optical Engineering: Fiber Optic Sensor Technology and Applications 2001*, Newton, MA, USA, October 30–November 1, 2001, pp. 229–38.)

interferometric output that overlaps the wavelengths to avoid the nonlinearities of the single wavelength readout.

Even though the circuitry has been removed from the environment in which the diaphragm is located, there is still a need to provide temperature compensation in some applications due to the temperature dependence of the diaphragm. In addressing this issue, Wolthuis *et al.* (1993) developed a dual-function sensor that measures both pressure and temperature and can be optically interrogated through a single optical fiber. This is done by using light emitting diodes (LEDs) of two different wavelengths to interrogate each sensor type at the readout end of the optical fiber. At the transduction end of the fiber, a dichroic mirror is placed after the temperature sensor that serially separates it from the pressure sensor. To implement this architecture, the temperature sensor has to respond to one LED for transduction as well as for passively transmitting light from the LED that is interrogating the pressure sensor. For temperature transduction, Wolthuis *et al.* utilized an absorptive long pass filter that has a temperature-dependent cutoff wavelength. The temperature signal is transduced using a 674 nm LED, whose reflection is marked by the temperature-dependent absorptive

filter. With the pressure sensors working with light in the infrared (IR) region, it is unaffected by the filters that interact with the smaller wavelength and just passes through to the pressure-sensitive diaphragm at the tip of the device. Displacements in the cavity below the diaphragm alter the reflectivity of each LED wavelength, which can be detected at the opposite end of the fiber.

2.04.4 Applications

Some of the major bottlenecks for successful commercialization are the efforts in packaging, testing, and calibration, which are all needed for final use of the devices in their intended application. Almost all MEMS packaging, testing, and calibration need to be completely redesigned for individual devices. Unlike in integrated circuits, the lack of standard/automated packaging, calibration, and testing capability strongly impedes yield improvement and cost reduction. This section first addresses packaging for one of the first commercialized pressure sensors in the case study of the manifold absolute pressure (MAP) sensor. Rough structures of the MAP sensor package and issues are described. This section is then followed by a section that addresses the advantages and disadvantages of sensor circuit integration. Interface circuitry is presented to discuss various readout schemes for the different transducer types that then convert the transduced parameter into an electrical signal. This is followed by a subsection that discusses calibration and temperature compensation techniques, which are essential issues in addressing the system-level transducer operation. This section is then concluded with a discussion on merging MEMS-based pressure sensors within a wireless system.

2.04.4.1 Packaging Technology

Packaging is considered to be the single most critical element in successful commercialization of microsystem products. The cost of packaging ranges between 20% and 95% of the total production cost depending upon packaging requirements. There are many factors related to the high costs including the following:

(1) No (or very little) standard automated assembly tools are available.
(2) High diversity of products, which demands different requirements of packaging from one to another.
(3) Packaging processes being used for integrated circuits often cause undesirable effects on microsystems including stiction, geometry distortion, and delamination of the packaged devices.

Packaging of micromachined pressure sensors is not an exception to these aforementioned issues. In this section, we introduce an example of packaging techniques for pressure sensors.

Packaging of pressure sensors needs to provide access for an external pressure source unlike for other more conventional hermetic packaging requirements. This attribute specification often makes pressure sensor packaging complicated and costly. Nevertheless, developments of pressure sensors have matured to penetrate markets, and one package that has been successfully commercialized is the pressure sensor described in this section (Monk 2002).

Figure 28 illustrates the Freescale MAP sensor package (Freescale 2005, 2007a). The MAP sensor needs a reference pressure to measure absolute pressure, which is provided by a backplate to form a vacuum cavity (Freescale 2007b in press). The sensor is bonded to the backplate by wafer-level, glass frit bonding. Due to on-chip conditioning electronics, the bond temperature needs to be low to maintain the thermal budget. Unlike anodic bonding, glass frit bonding can tolerate nonperfectly flat bond interface while the bond is performed at low temperature (400–500°C). The bonded die is then placed inside the premolded thermoplastic case, attached to the

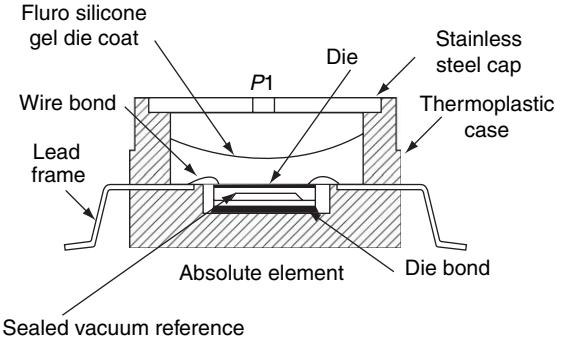


Figure 28 Freescale manifold absolute pressure (MAP) sensor package. (Source: Freescale 2007 Data sheet, MPXA6115A, High temperature accuracy integrated silicon pressure sensor for measuring absolute pressure, on-chip signal conditioned, temperature compensated, calibrated).

bottom of the case by an adhesive wire bonded to a lead frame, and covered with a low-stress (fluoro) silicone gel. The die is attached to a square pad of adhesive (silicone gel), which needs a 150°C oven cure. The adhesive was chosen not to transfer a large stress to the die. The low-stress (fluoro) silicone is to protect the sensor from environments while allowing pressure measurements at the same time. It has a low modulus and provides passivation against corrosion. A stainless-steel cap with a pressure access hole covers the plastic case. The sensor has a mechanical diaphragm and signal-conditioning electronics monolithically integrated with the MEMS device. The monolithic integration can achieve a smaller overall device size and improved interconnect reliability.

2.04.4.2 Monolithic Integration

Sensor circuit integration is one of the main benefits in micromachined transducers made out of silicon. This attribute has been sought after since the early days of micromachining pressure sensors in silicon. As discussed earlier, the first monolithically integrated sensor and circuit was developed by Borky and Wise in 1979 (Borky and Wise 1979). Compatibility with the current circuit fabrication processes has been difficult to maintain for this level of integration, since many of the sensor and circuit processes need some level of compromise. For example, concentration-dependent etch stops cannot be used in the circuit areas so a certain level of care must be taken to ensure that the circuit areas are protected in other ways. Chavan and Wise (2002) successfully developed a monolithic device that merges the bulk micromachined silicon-on-glass sensor array with a 3 μm bipolar complementary metal oxide semiconductor (BiCMOS) process. This then provides the needed on-site electronics for the capacitive transducers as well as enables the first-level encapsulation of the system. A cross section of this device is shown in **Figure 29**.

Surface micromachined versions of monolithic devices have implemented a combination of polysilicon diaphragm and silicon dioxide or phosphosilicate glass (PSG) film as a sacrificial layer. Guo *et al.* (2000) presented a capacitive pressure sensor that uses polysilicon as a membrane and PSG as a sacrificial layer as shown in **Figure 30**. The PSG sacrificial layer is deposited between n-/p-channel source/drain implantation and polysilicon deposition. The sacrificial layer is etched using hydrofluoric acid (HF) to release the polysilicon membrane, and the cavity is sealed by LTO to form a reference pressure.

Success in monolithic integration has now come to the point where it is offered as a foundry service utilizing bulk devices fabricated using a electrochemical-dependent etch stop and surface micromachined structures (Foundries 2005a, b).

2.04.4.3 Interface Circuits

Interface circuits are an essential part of a sensing system providing the signal conversion to an analog or digital signal. There have been very extensive developments on the readout/feedback circuitry for micromachined pressure sensors. In this section,

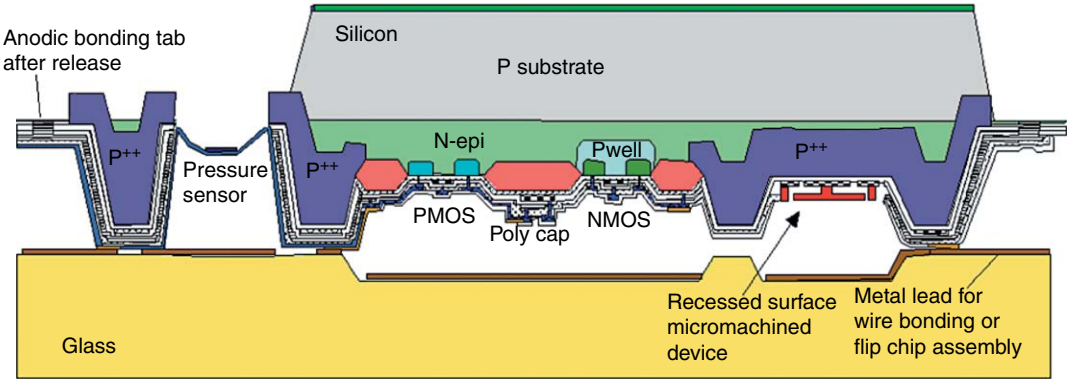


Figure 29 Cross section of the integrated device developed by Chavan and Wise merging pressure sensors formed with the silicon-on-glass dissolved wafer process with a bipolar complementary metal-oxide semiconductor (BiCMOS) fabrication process. (Source: Chavan A V, Wise K D 2002 A monolithic fully-integrated vacuum-sealed CMOS pressure sensor. *IEEE Trans. Electron Devices* **49**, 164–9.)

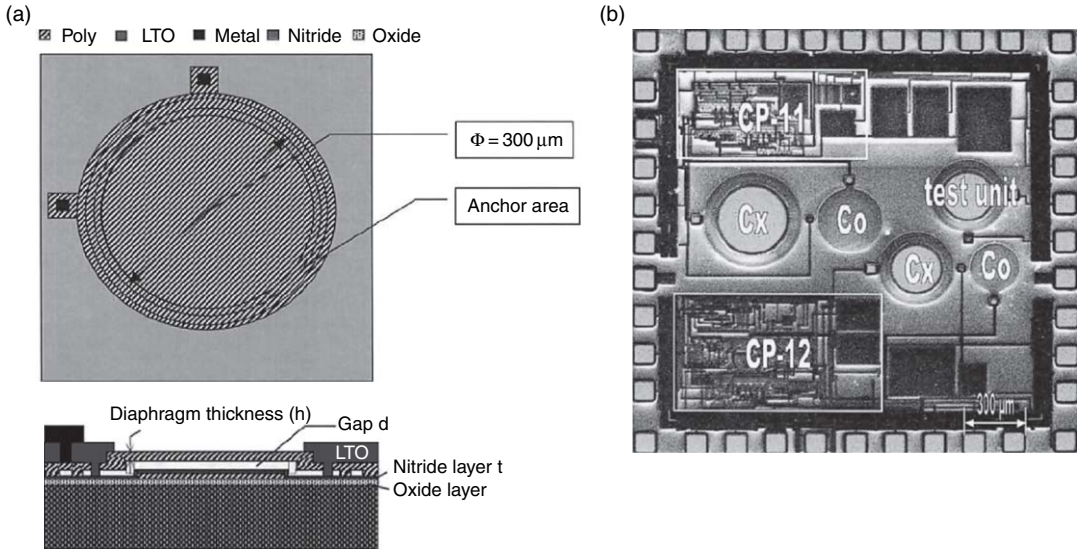


Figure 30 Integrated polysilicon diaphragm-based pressure sensors. (a) Layout of cross section and (b) integrated chip shot. (Source: Guo S, Guo J, Ko W H 2000 A monolithically integrated surface micromachined touch mode capacitive pressure sensor. *Sens. Actuators A Phys.* **A80**, 224–32.)

interface electronics for pressure sensors are briefly introduced. The interface electronics described in this section are separated into piezoresistive, capacitive, resonant, and thermal detection for readout of their respective transducer types.

2.04.4.3.1 Piezoresistive interface

Building interface circuitry into the output of the Wheatstone bridge provides the first stage of signal conversion. The output of Wheatstone bridge is analog, which is not immune to changes in the signal path and noises as a time-domain signal; thus converting an analog signal to a digital form is highly desirable. Spencer *et al.* introduced a voltage-controlled duty-cycle oscillator (VCDCO). This straightforward approach to digital signal conversion is shown in **Figure 31** (Spencer and Angell 1990). This oscillator generates a pulse-width-modulated digital output using simple bipolar junction transistor (BJT) circuits. The oscillation cycle and the voltages of notable nodes are shown **Figure 31(b)**.

Analysis of this circuit enables defining the voltage charge on one side of the capacitor as:

$$\Delta V_C = V_{E2} - V_{E1} = (V_{CC} - 0.7V) - (V_{CC} - RI_{EE} - 0.7V) = RI_{EE} \quad [26]$$

where V_{E1} and V_{E2} are the voltages at the emitters of Q_1 and Q_2, respectively, and V_C is the voltage across the capacitor. The output of the oscillator is then taken at the collector of Q_2 and is shown in the V_{C2} waveform. The positive feedback transistors then create a multivibrator with the time for each phase of the charging cycle defined as:

$$t_2 \approx C\frac{\Delta V}{I} = C\frac{2RI_{EE}}{I_{C3}} \quad [27]$$

and

$$t_1 \approx C\frac{2RI_{EE}}{I_{C4}} \quad [28]$$

If Q_1 and Q_2 transistor pairs are matched and switch abruptly, then the current ratio can be defined as:

$$\frac{I_{C3}}{I_{C4}} = \exp(v_{in}/V_T) \quad [29]$$

This then enables the dependence of the oscillator timing on the input voltage from the Wheatstone bridge as:

$$v_{in} = V_T \ln\left(\frac{t_1}{t_2}\right). \quad [30]$$

2.04.4.3.2 Capacitive interface

Capacitive sensor readout provides a number of advantages that build on the simple transducer structure such as low noise and low power consumption. There are three main implementations of capacitive readout: capacitance–frequency (C–F) converter, AC

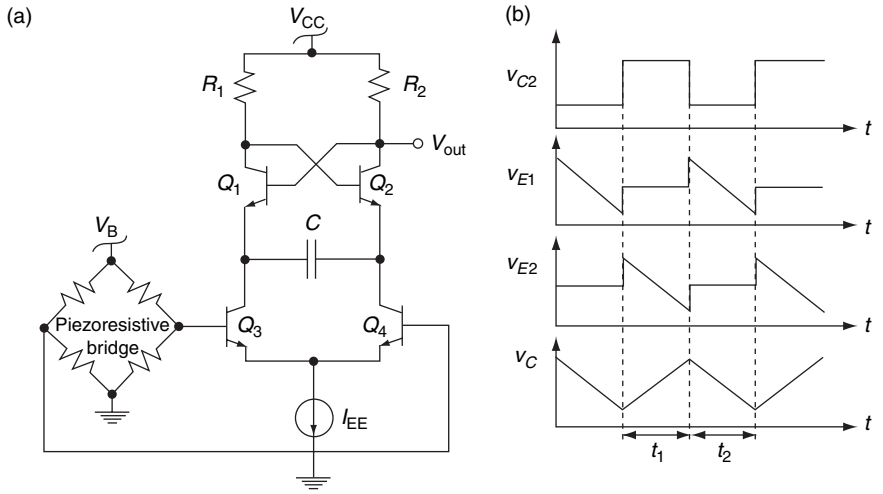


Figure 31 (a) Voltage-controlled duty-cycle oscillator (VCDCO) and (b) waveforms of VCDCO. (Source: Spencer R R, Angell J B 1990 A voltage-controlled duty-cycle oscillator. *IEEE J. Solid-State Circuits* **25**, 274–81.)

bridge, and switched-capacitor (SC) interface. C–F converter is simple, but generally suffers from parasitics and temperature effects. AC-bridge-type interface utilizes complementary excitation pulses to charge capacitors and to convert to voltage signal. This type of interface is relatively easy to implement; however, synchronizing the phase between the modulated signal and the reference is challenging. Also, AC bridge generally suffers from parasitics. Switched-cap interface is the most sophisticated circuit architecture of all, but offers extremely low noise and robust architecture against parasitics.

Like the oscillator described in the resistive readout, a capacitive version of this oscillator then implements the charging capacitor as the varying parameter. Matsumoto and coworkers developed the C–F interface circuit shown in **Figure 32** (Matsumoto and Esashi 1993). The sensor capacitor is represented as C_X. Using the input hysteresis of a Schmitt trigger inverter, the voltages across the reference capacitor (C_r) and the sense capacitor (C_X) are converted to frequency signal and the slope of the output curve represents C_r and C_X. When charging C_r, the output of Schmitt trigger forms a line, and the slope of the line is I_o/C_r, and when discharging C_X the slope becomes I_o/C_X. Assuming I_o and C_r are known, C_X is modulated to frequency signal at the output. The period and frequency can be defined as:

$$t = t_r + t_f = \frac{V_b C_r}{I_o} + \frac{V_b C_X}{I_o} = \frac{V_b}{I_o}(C_r + C_X) \qquad [31]$$

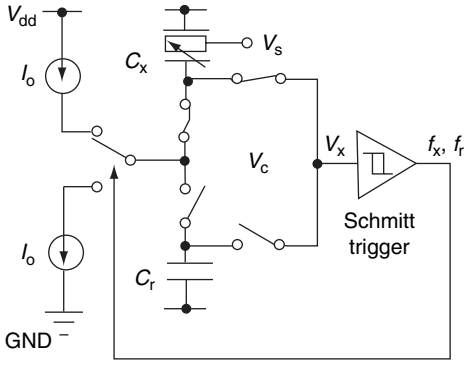


Figure 32 Schematic of a capacitance-to-frequency converter circuit implemented by Matsumoto *et al.* (Source: Matsumoto Y, Esashi M 1993 Integrated silicon capacitive accelerometer with PLL servo technique. *Sens. Actuators A Phys.* **A39**, 209–17.)

and:

$$f = \frac{I_o}{V_b(C_r + C_X)} \qquad [32]$$

where V_b is the difference in the trip points for the Schmitt trigger.

2.04.4.3.2.(i) AC bridge The AC bridge configuration has been widely used since the early 1990s, especially for microaccelerometers. Again, almost the same architecture can be used for capacitive pressure sensors. The AC bridge configuration utilizes complementary AC carriers, preferably exactly out-of-phase carriers. **Figure 33** shows a simplified system block diagram of AC bridge interface circuits (Yun *et al.* 1992).

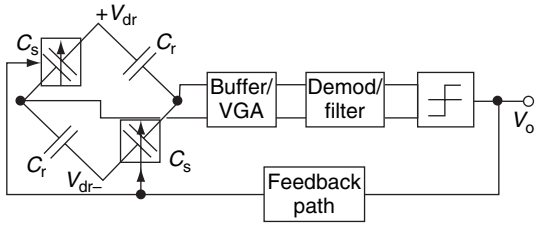


Figure 33 System block diagram for AC bridge interface circuit developed by Yun *et al.* in 1992. (Source: Yun W, Howe R T, Gray P R 1992 Surface micromachined digitally force-balanced accelerometer with integrated CMOS detection circuitry. *Solid-State Sensor and Actuator Workshop*, Hilton Head, SC, pp. 126–31.)

The sense capacitors (C_s) and reference capacitors (C_r) form the capacitive bridge. The bridge is excited by a high-frequency (in this case, 1 MHz) complementary square wave ($+V_{dr}$ and $-V_{dr}$). The output of the bridge is a linear function of capacitance change, which is amplified, demodulated, filtered, and quantized to produce a pulse-width-modulated (PWM) digital output (V_o) (Chau *et al.* 1996). Due to the vulnerable architecture for parasitics and input capacitance of electronics, the output of the bridge is fed to a unity gain buffer with low input capacitance to bootstrap the ground plane. This is necessary if the sense capacitors are small so that they are not overwhelmed by stray capacitances. The buffered signal goes to a variable gain amplifier (VGA) to amplify the signal controlled externally. The amplified signal is demodulated and filtered before being quantized. This functionality is expressed analytically as follows:

$$[A(t)\cos(\omega t)][B\cos(\omega t+\theta)] = \frac{A(t)B}{2}[\cos(\theta)+\cos(2\omega t+\theta)] \quad [33]$$

Here, $A(t)\cos(\omega t)$ is the output from the VGA ($A(t) \propto \Delta C$) and $B\cos(\omega t+\theta)$ is the reference (B is known, and θ is the phase difference between the reference and the output from VGA).

After the filter, the demodulated signal can be approximated by $A(t)\mathrm{B}\cos(\theta)$. Since B is known and $A(t)$ is proportional to capacitance change, the incident pressure that is deflecting the diaphragm can be extracted. The demodulated and filtered signal is quantized by a quantizer (in this case 1-bit quantizer) to produce a PWM signal. Zero or one of PWM corresponds to polarity of pressure input, and the width indicates the magnitude of the input. The demodulated and filtered signal is proportional to ΔC as well as $\cos(\theta)$. This indicates the phase difference between the reference and the output from the bridge significantly affect the sensing capability. This inherent sensitivity of the phase makes the AC bridge interface less attractive than the SC configuration despite the relatively simple circuit implementation architecture.

SC interfaces are the most frequently used interface configurations for capacitive detection. One example of an SC interface is shown in **Figure 34** as implemented by Chavan and Wise (2002). This SC interface circuits consist of three stages: a front-end integrator, a VGA, and sample and hold stages. The front-end integrator is the heart of the SC interface, the fully differential charge integrator with a self-biased folded-cascode amplifier. The charge integrator implements a correlated double-sampling (CDS) technique to reduce $1/f$ noise and amplifier offset. The user-configurable parameters, including feedback capacitor (C_f), reference capacitor (C_r), and on-chip gain capacitors (C_{g2}), allows adjusting output (V_{DC}), is shown as:

$$V_{DC} = \frac{C_{g2}}{C_{f2}}\left(\frac{C_s}{C_f}V_{in+} - \frac{C_R}{C_f}V_{in-}\right) + 2.5 \quad [34]$$

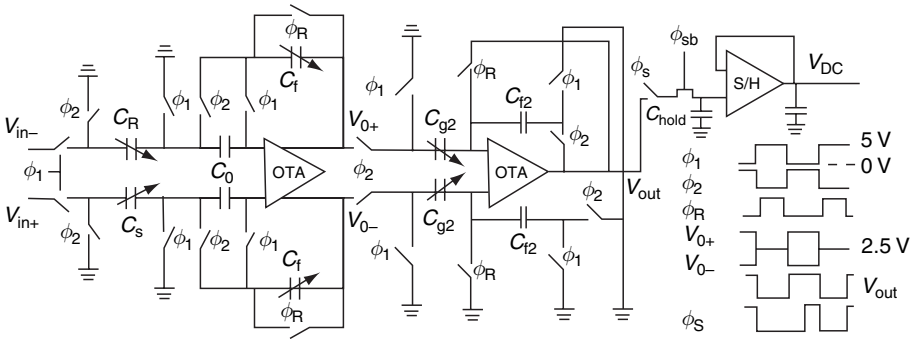


Figure 34 Switched-capacitor (SC) interface implemented for capacitive readout by Chavan and Wise. (Source: Chavan A V, Wise K D 2002 A monolithic fully-integrated vacuum-sealed CMOS pressure sensor. *IEEE Trans. Electron Devices* **49**, 164–9.)

The SC interface circuitry does not need a demodulation stage where the exact phase match is required to produce a precise output. Also, the output is independent of parasitics to the first order. These robust features make an SC interface very attractive because they allow for off-chip electronics when needed. On the other hand, SC interfaces are more complicated than other circuit implementation architectures.

2.04.4.3.3 Resonant sensor interface

Interface circuitry for micromachined resonators is based on a transimpedance amplifier (TIA). A simplified resonant interface implemented by Nguyen is shown in **Figure 35** (Nguyen 1995). Excitation is provided at the input as a voltage source ($V_I + v_i$) and the resonant device is biased by bias voltage (V_P). Due to electromechanical transduction, the output of the resonator is current (i_o), which is fed to TIA to produce voltage output. The amplitude of the output is scanned over the frequency and generates a Q factor versus pressure plot. As pressure changes, the detected Q factor changes due to the sensitivity of the resonator to air damping.

2.04.4.3.4 Pirani gauge thermal interface

A Pirani gauge measures pressure by detecting the heat flux change from a heater to a heat sink as discussed in Section 2.04.3.4. The interface for this device is straightforward yet offers high resolution and accuracy. **Figure 36** shows a benchtop measurement diagram for the device (Stark *et al.* 2003).

A current source forces a constant current through the gauge, and the voltage drop across the gauge is measured using a standard four-point probe configuration. Temperature can then be determined from the measured resistance of the gauge with the given temperature coefficient of resistance (TCR) while power is measured from the current and voltage product. Current is increased until the gauge reaches a preset temperature, typically ~30°C, and then a linear curve fit is applied to the power versus the temperature data to extract thermal impedance that is a slope of the linear curve. As pressure decreases, the slope of the curve, which corresponds to the thermal impedance, then increases as shown in **Figure 37** (Chae *et al.* 2005b).

According to Chae *et al.*, micro Pirani gauge has an uncertainty of 50 μtorr and can be used for hermeticity testing to detect a leak rate of 3.1×10^{-16} $cm^3\,s^{-1}$, assuming a common micropackage volume of 1.6×10^{-5} cm^3. This exceeds the resolution by at least four orders of magnitude over traditional leak testing.

2.04.4.4 Calibration and Temperature Compensation

Assessment of the true performance of a pressure sensor system is done by analyzing the accuracy that can be achieved when the sensing device and its interface circuitry are combined. For a given measurement, the resolution of the sensor system will then be set by device and interface circuitry noise. Spanning the dynamic range of the sensor, the system-level measurements will also be impacted by the sensor's nonlinearity. In addition, the device also needs to work over a range of temperatures. These issues can be addressed at either the device or the system level. Device-to-device tolerance must also

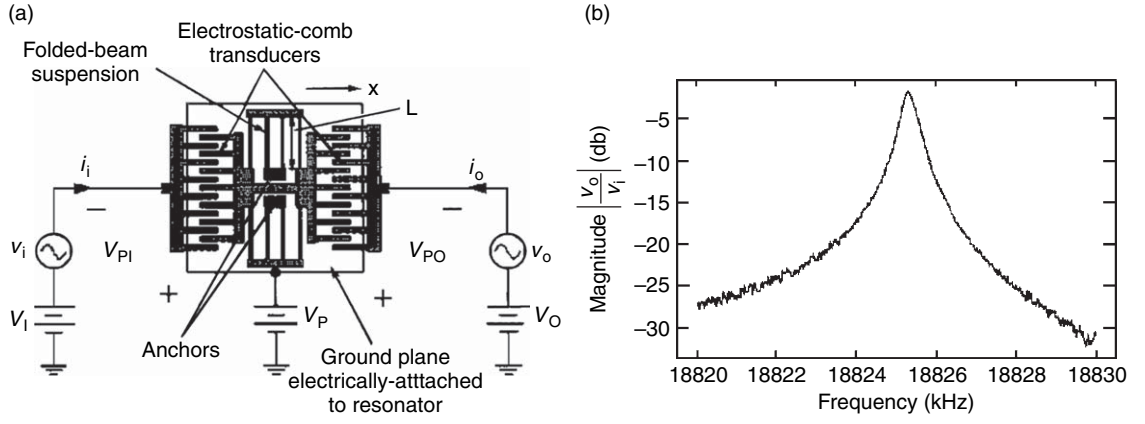


Figure 35 Resonant device interface technique. (a) The measurement configuration for a microresonator; and (b) the output amplitude plot over frequency. (Source: Nguyen C 1995 Micromechanical resonators for oscillators and filters. IEEE Ultrasonics Symposium, pp. 489–99.)

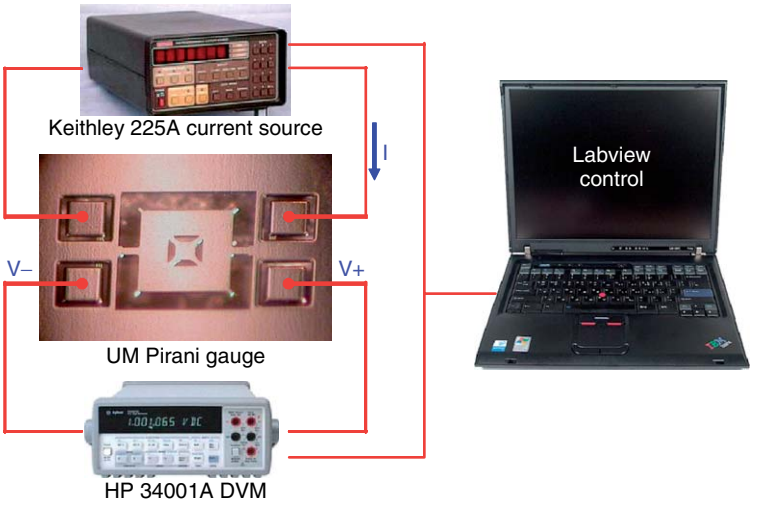


Figure 36 Simplified measurement setup for a micro Pirani gauge reproduced from Chae *et al.* (Source: Stark B, Yuhai M, Zhang C, Najafi K 2003 A doubly anchored surface micromachined Pirani gauge for vacuum package characterization. *IEEE 16th Annu. Int. Conf. Micro Electro Mechanical Systems*, Kyoto, Japan, pp. 506–9.)

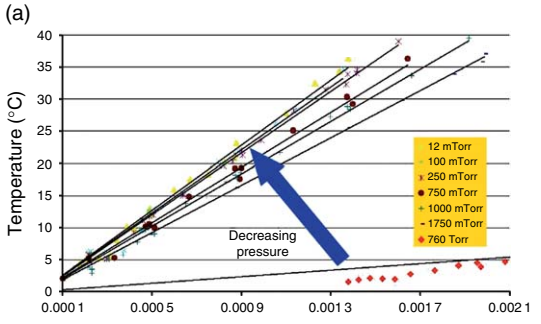


Figure 37 Micro Pirani gauge characterization showing the thermal impedance (slope) of a Pirani device versus ambient pressure. (Source: Chae J, Stark B, Najafi K 2005 A micromachined Pirani gauge with dual heat sinks. *IEEE-CPMT Adv. Packaging Micro/Nano-Scale Syst.* **28**, 619–25.)

be addressed in a manufacturing environment. To facilitate this, some level of individual device modifications can be utilized to compensate for variations. Lee *et al.* (1999) implemented a laser trimming scheme that enables calibration for offset voltage, full-scale span, and TCO voltage and full-scale span. This design used ion-implanted resistors in the interface circuit that are implanted in the same step that the diaphragm piezoresistors are implanted. This technique was able to produce devices with a temperature coefficient of 46 ppm $°C^{-1}$.

These issues can be addressed at the system level by using smart sensor systems that are microcontroller based and allow for device customization within the firmware of the device. The two predominant ways of implementing calibration at this stage are to use a lookup table or to perform a polynomial approximation (Crary *et al.* 1990). A lookup table has the potential to provide a high level of accuracy in calibration. However, there is a high demand made on device memory and there is an inherent trade-off between the level of complexity built into the lookup table and the amount of memory available to the device. Polynomial approximation has also shown potential to enable accurate calibration showing a fivefold improvement in the linearization of the sensor's data (Crary *et al.* 1990). Although the polynomial approximation method does relax the memory requirements, it does utilize more processing power. This then presents an inherent trade-off between the bandwidth of the device and the optimization of device accuracy through compensation. Calibration of the device is also a practical matter since the resolution of the sensor then needs to be maintained over the lifetime of the device. This then needs a high level of characterization of the reliability of the seal for the reference cavity.

2.04.4.5 Wireless Devices

There are two levels of merging wireless technology and wireless sensors: at the system level and at the device level. Merging these at the system level involves microcontroller-based systems, which provide the sensor interface and can be configured to transmit and receive. However, merging these technologies at the device level opens up MEMS-based pressure sensors to a wide variety of applications that

need a moderate amount of telemetry distance combined with a small device footprint. The merging of wireless technology with micromachined pressure sensors at the device level has enabled highly scalable devices that can be designed for applications spanning biomedical, environmental, and industrial monitoring. The most straightforward implementation of merging wireless technology and pressure sensors is the integration of an antenna with a capacitive sensor. When connected to form a parallel inductor and capacitor (LC) tank, this device can be queried remotely through a loosely coupled passive telemetry system. A nearby coupled inductor can then detect the resonant frequency of the LC tank by a dip in the phase of its own measured impedance. This type of system was demonstrated by Collins who, in 1967, implemented a system for the remote monitoring of a resonant LC tank to provide long-term ocular pressure measurements (Collins 1967). Readout of a capacitive sensor in this configuration defines the resonant frequency of the LC tank as:

$$\omega_o = \frac{1}{\sqrt{L_s C_s}} \quad [35]$$

At the resonant frequency of this secondary LC tank, the received peak phase dip signal can be defined by:

$$\Delta\phi_{peak} = \tan^{-1}\left(k^2 Q_{tank}^{series}\right) \quad [36]$$

where the quality factor of the secondary LC tank is defined by:

$$Q_{tank}^{series} = \frac{1}{R_s}\sqrt{\frac{L_s}{C_s}} \quad [37]$$

and k is the coupling coefficient of the loosely coupled inductors. Although this type of wireless interface provides a relatively straightforward device implementation, attention must be paid to any parasitic capacitances and resistances in the device as they directly impact the sensitivity, telemetry range, and resolution of the coupled system.

Recent work by Akar and Akin presents a monolithic version of a passive telemetry-based pressure sensor (Akar *et al.* 2001). **Figure 38** shows a cross section of the device as well as its equivalent circuit model. This device measures only 2.6 mm × 1.6 mm. This type of device is designed for an intracranial or intraocular application. The integrated inductor measures 1.2 μH and enables a wireless pressure

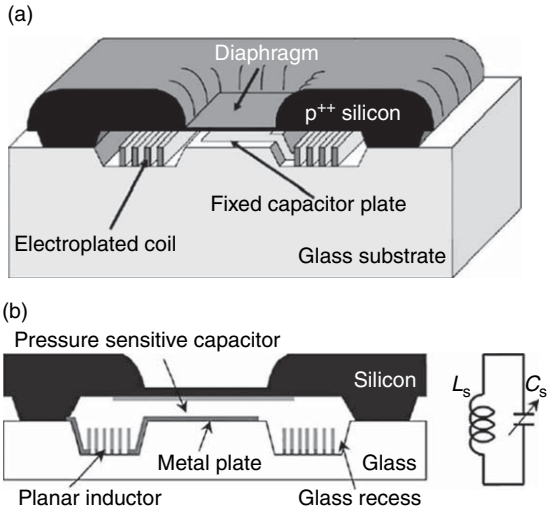


Figure 38 (a) Device structure and (b) cross section and equivalent circuit for an LC tank sensor enabling a passive, monolithic wireless pressure sensor. (Source: Akar O, Akin T, Najafi K 2001 A wireless batch sealed absolute capacitive pressure sensor. *Sens. Actuators A Phys.* **95**, 29–38.)

sensitivity of 1553 ppm $mmHg^{-1}$ with a self-resonant frequency between 95 and 103 MHz when operating between 0 and 50 mmHg.

Moving to a higher level of integration to include monolithically fabricated active circuitry then leads to the implementation of multiple sensors within the device that can be read out using backscatter telemetry. This level of integration was used to implement a dual-site pressure sensor for use in characterizing the pressure differential and flow rates within the arterial system (DeHennis and Wise 2006). The schematics for this system and device photograph are shown in **Figure 39**.

This device was developed for use within the carotid artery and has a volume of 2 mm^3 integrating a 3 μm BiCMOS process with a sputtered antenna and two bulk micromachined capacitive pressure sensors. The entire device is released using a series of dicing steps along with dissolving both the excess glass and the silicon substrates using wet etching techniques. The wireless sensor interface consumes 340 μW and uses capacitance-to-frequency conversion for readout of the vacuum-sealed pressure transducers. The integrated device has a 200 μm profile and a volume of 2 mm^3. The system can sense a reduction in flow of 13%, which corresponds to a differential pressure of 3 mmHg (**Figure 40**).

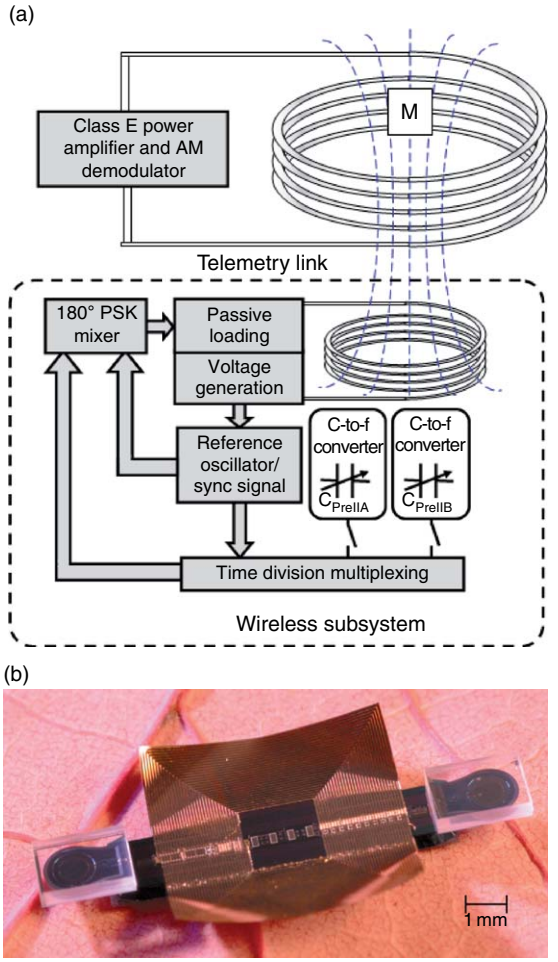


Figure 39 (a) Wireless system schematic and (b) fabricated device photograph of a two-site, wireless pressure sensor that enables intraarterial pressure and flow characterization. (Source: DeHennis A D, Wise K D 2006 A fully integrated multisite pressure sensor for wireless arterial flow characterization. *J. Microelectromech. Syst.* **15**, 678–85.)

Integration of the pressure transducer and circuitry with a hybrid attached antenna enabled the development by Stangel *et al.* (2001) of an integrated, surface micromachined sensing device that is targeted for intraocular application. This device integrates a 1.2 μm complementary metal oxide semiconductor (CMOS) interface circuit with an array of polysilicon diaphragm transducers. The on-chip temperature sensor, sensor readout, and calibration electronics are then interfaced through a 13.56 MHz backscattered modulation telemetry scheme. This device needs only the attachment of a foldable planar coil to enable remote querying of ocular pressure over long periods of time. A photograph of the integrated die is shown in **Figure 41**.

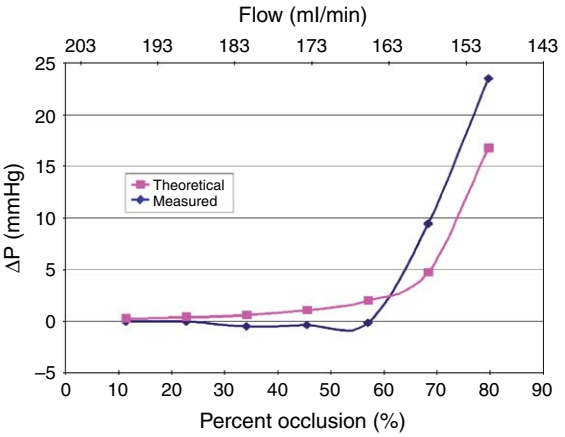


Figure 40 Pressure versus flow rate characterization of the system as the narrowing of the tube progresses. (Source: DeHennis A D, Wise K D 2006 A fully integrated multisite pressure sensor for wireless arterial flow characterization. *J. Microelectromech. Syst.* **15**, 678–85.)

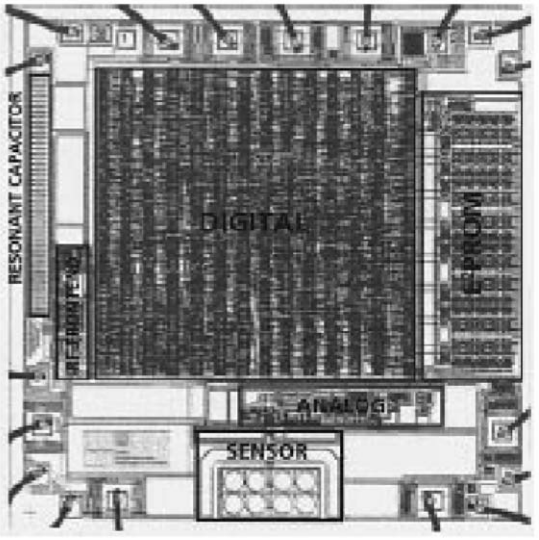


Figure 41 Photograph of the 2.6 mm × 2.6 mm square die integrating a 1.2 μM complementary metal oxide semiconductor (CMOS) process and surface micromachined pressure sensor to enable wireless intraocular pressure measurement. (Source: Stangel K, Kolnsberg S, Hammerschmidt D, Hosticka B J, Trieu H K, Mokwa W 2001 A programmable intraocular CMOS pressure sensor system implant. *IEEE J. Solid-State Circuits* **36**, 1094–100.)

2.04.5 Conclusions

This chapter has presented an overview of MEMS-based pressure sensors spanning their device-level design and their system-level application. The material presented in this chapter provides a basis for understanding the basic transducer functionality as well as

the high-level intricacies that are needed to realize a functional sensing system. Process development has enabled the ability to form the transducer on a monolithic substrate in a variety of forms. This chapter has also covered various transduction mechanisms that enable measurement of pressures from vacuum to ambient, even in applications with elevated temperatures. The presentation of interface electronics, which provides the second level of transduction, then builds to a system-level perspective of MEMS-based sensors. To date, pressure sensors have been the most successfully commercialized MEMS device. The continued technological improvements in device performance and system level of integration provide platforms for silicon-based pressure sensors to continue to reach the forefront of industrial, biomedical, and environmental applications.

References

Akar O, Akin T, Najafi K 2001 A wireless batch sealed absolute capacitive pressure sensor. *Sens. Actuators A Phys.* **95**, 29–38

Borky J, Wise K D 1979 Integrated signal conditioning for silicon pressure sensors. *IEEE Trans. Electron Devices* **ED-26**, 1906–10

Bryzek J, Roundy S, Bircumshaw B, Chung C, Castellino K, Stetter J R, Vestel M 2006 Marvelous MEMS. *IEEE Circuits Devices Mag.* **22**, 8–28

Burns D W, Zook J D, Horning R D, Herb W R, Guckel H 1995 Sealed-cavity resonant microbeam pressure sensor. *Sens. Actuators A Phys.* **48**, 179–86

Chae J, Giachino J M, Najafi K 2005a Wafer-level vacuum package with vertical feedthroughs. *18th IEEE Int. Conf. Micro Electro Mechanical Systems, MEMS 2005,* Miami, FL, USA, January 30–February 3, 2005, pp. 548–51

Chae J, Stark B H, Najafi K 2005b A micromachined Pirani gauge with dual heat sinks. *IEEE-CPMT Special Issue for Advanced Packaging for Micro/Nano-Scale Systems* Vol. **28**, 619–25

Chang-Chien P 2002 Wafer-level packaging and frequency trimming by localized mass deposition. Ph.D. dissertation, University of Michigan

Chang-Chien P P L, Wise K D 2001 Wafer-level packaging using localized mass deposition. *Proc. 11th Int. Conf. Solid State Sensors and Actuators Transducers '01/Eurosensors XV. Digest of Technical Papers*, June 10–14, 2001. Springer-Verlag, Munich, Germany, **Vol. 1**, pp. 182–5 ,BN – 3 540 42150 5

Chau H-L, Wise K D 1987 Scaling limits in batch-fabricated silicon pressure sensors. *IEEE Trans. Electron Devices* **ED-34**, 850–8

Chau H-L, Wise K D 1988 An ultraminiature solid-state pressure sensor for a cardiovascular catheter. *IEEE Trans. Electron Devices* **35**, 2355–62

Chau K, Lewis S, Zhao Y, Howe R, Bart S, Marcheselli R 1996 An integrated force-balanced capacitive accelerometer for low-*g* applications. *Sens. Actuators A Phys.* **54**, 472–6

Chavan A V, Wise K D 1997 A batch-processed vacuum-sealed capacitive pressure sensors. *International Conference on Solid-State Sensors and Actuators. Digest of Technical Papers, Transducers 97',* **Vol. 2**, pp. 1449–52

Chavan A V, Wise K D 2001b Batch-processed vacuum-sealed capacitive pressure sensors. *IEEE J. Microelectromech. Syst.* **10**, 580–8

Chavan A V, Wise K D 2002 A monolithic fully-integrated vacuum-sealed CMOS pressure sensor. *IEEE Trans. Electron Devices* **49**, 164–9

Chen J, Liu L, Li Z, Tan Z, Xu Y, Ma J 2003 On the single-chip condenser miniature microphone using DRIE and backside etching techniques. *Sens. Actuators A Phys.* **103**, 42–7

Cho S T, Najafi K, Wise K D 1992 Internal stress compensation and scaling in ultrasensitive silicon pressure sensors. *IEEE Trans. Electron Devices* **39**, 836–42

Clark S K, Wise K D 1979 Pressure sensitivity in anisotropically etched thin-diaphragm pressure sensors. *IEEE Trans. Electron Devices* **ED-26**, 1887–96

Collins C C 1967 Miniature passive pressure transensor for implanting in the eye. *IEEE Trans. Biomed. Eng.* **BME-14**, 74–83

Crary S B, Baer W G, Cowles J C, Wise K D 1990 Digital compensation of high-performance silicon pressure transducers. *5th Int. Conf. Solid-State Sensors and Actuators and Eurosensors III, Sensors and Actuators A (Physical),* Montreux, Switzerland, June 25–30, 1989, pp. 70–2

Crescini D, Ferraro V, Marioli D, Taroni A 1997 A thick-film capacitive pressure sensor with improved linearity due to electrode-shaping and frequency conversion. *Meas. Sci. Technol.* **8**, 71–7

DeHennis A D, Wise K D 2006 A fully integrated multisite pressure sensor for wireless arterial flow characterization. *J. Microelectromech. Syst.* **15**, 678–85

Eaton W P, Smith J H 1997 Micromachined pressure sensors: Review and recent developments. *Smart Mater. Struct.* **6**, 530–9

El-Bahar A, Nemirovsky Y, Soustiel J F, Feinsod M 1998 Micromachined CMOS in vivo pressure sensor. *Electrotechnical Conference, MELECON 98, 9th Mediterranean, 1998,* Tel Aviv, Israel, pp. 306–10

Elwenspoek M 1993 On the mechanism of anisotropic etching of silicon. *J. Electrochem. Soc.* **140**, 2075–80

Esashi M, Matsumoto Y, Shuichi S 1990 Absolute pressure sensors by air-tight electrical feedthrough structure. *Sens. Actuators A* **A21–3**, 1048–52

Foundries X S 2005a Surface micromachining technology for discrete absolute pressure sensors. XFAB datasheet

Foundries X S 2005a Bulk micromachining technology for discrete relative pressure sensors. XFAB datasheet

Freescale 2001 Data sheet, integrated silicon pressure sensor for manifold absolute pressure, altimeter or barometer applications on-chip signal conditioned, temperature compensated, calibrated

Freescale 2005 Micromachined pressure sensors. Freescale Semiconductor Device Data DL 200 Rev. 6.1

Freescale 2007 Data sheet, MPXA6115A, High temperature accuracy integrated silicon pressure sensor for measuring absolute pressure, on-chip signal conditioned, temperature compensated, calibrated

French P J, Muro H, Shinohara T, Nojiri H, Kaneko H 1992 SOI pressure sensor. *Sens. Actuators A Phys.* **A35**, 17–22

Frobenius W D, Sanderson A C, Nathanson H C 1973 A microminiature solid-state capacitive blood pressure transducer with improved sensitivity. *IEEE Trans. Biomed. Eng.* **BME20**, 312–14

Fryer T B 1975 Capacitance pressure transducers. In: *Indwelling and Implantable Pressure Transducers, book based on the workshop held in Cleveland, OH, on December 4 and 5, 1975. , CRC Press, Cleveland, OH,* pp. 167–72.

Gad-el-Hak M 2002 *The MEMS Handbook*. CRC Press, Boca Raton, FL

Geyling F T, Forst J J 1960 Semiconductor strain transducers. *The Bell Syst. Tech. J.* **39**, 705–31

Gieles A C M 1969 Subminiature silicon pressure transducer. *International Solid-State Circuits Conference (ISSCC),* Philadelphia, PA, USA, pp. 108–9

Giovanni M 1982 *Flat Corrugated Diaphragm Design Handbook Marcel Dekker, New York*

Greenwood J C 1984 Etched silicon vibrating sensor. *J. Phys. E Sci. Instrum.* **17**, 650–2

Guo S, Guo J, Ko W H 2000 A monolithically integrated surface micromachined touch mode capacitive pressure sensor. *Sens. Actuators A Phys.* **A80**, 224–32

Henmi H, Shoji S, Shoji Y, Yoshmi K, Esashi M 1994 Vacuum packaging for microsensors by glass–silicon anodic bonding. *Sens. Actuators A* **43**, 243–8

Ikeda K, Kuwayama H, Kobayashi T, Watanabe T, Nishikawa T, Yoshida T, Harada K 1990 Silicon pressure sensor integrates resonant strain gauge on diaphragm. *5th Int. Conf. Solid-State Sensors and Actuators and Eurosensors III*, Montreux, Switzerland, Vol. **A21***Sens. Actuators A, Phys. (Switzerland)* ,, pp. 146–50

Kakinaga T, Hatai A, Tabata O, Isono Y 2005 Silicon anisotropic wet etching simulation using molecular dynamics. *Transducers '05. 13th Int. Conf. Solid-State Sensors, Actuators and Microsystems. Digest of Technical Papers,* Seoul, South Korea, June 5–9, 2005, pp. 816–19 ,BN – 0 7803 8994 8

Kalvesten J E, Smith L, Tenerz L, Stemme G 1998 First surface micromachined pressure sensor for cardiovascular pressure measurements. *Proc. 1998 IEEE 11th Annu. Int. Workshop Micro Electro Mechanical Systems,* January 25–29, 1998, Heidelberg, Germany, pp. 574–9

Kim B, Candler R N, Hopcroft M, Agarwal M, Park W-T, Kenny T W 2005 Frequency stability of wafer-scale encapsulated MEMS resonators. *Transducers '05. 13th Int. Conf. Solid-State Sensors, Actuators and Microsystems. Digest of Technical Papers,* Seoul, South Korea, June 5–9, 2005, pp. 1965–8 ,BN – 0 7803 8994 8

Kloeck B, Collins S D, de Rooij N F, Smith R L 1989 Study of electrochemical etch-stop for high-precision thickness control of silicon membranes. *IEEE Trans. Electron Devices* **36**, 663–9

Ko W H, Wang Q 1999 Touch mode capacitive pressure sensors. *Sens. Actuators A Phys.* **A75**, 242–51

Lee B-N, Kim K-N, Park H-D, Shin S-M 1999 Calibration and temperature compensation of silicon pressure sensors using ion-implanted trimming resistors. *Sens. Actuators A Phys.* **A72**, 148–52

Liu C 2005 *Foundations of MEMS*. Upper Saddle River, NJ: Pearson Prentice Hall, 2006

Mastrangelo C H 1991 *Thermal Applications of Microbridges* University of California, Berkeley, CA

Matsumoto Y, Esashi M 1993 Integrated silicon capacitive accelerometer with PLL servo technique. *Sens. Actuators A Phys.* **A39**, 209–17

Meng G, Ko W H 1999 Modeling of circular diaphragm and spreadsheet solution programming for touch mode capacitive sensors. *Sens. Actuators A Phys.* **A75**, 45–52

Monk D 2002 MEMS Physical Sensors for Automotive Applications. *Sensor/Dielectric Science and Technology/ Electronics,* Philadelphia, PA, USA

Najafi K, Ji J, Wise K D 1990 Scaling limitations of silicon multichannel recording probes. *IEEE Trans. Biomed. Eng.* **37**, 1–10

Newell W 1968 Miniaturization of tuning forks. *Science* **161**, 1320–6

Nguyen C 1995 Micromechanical resonators for oscillators and filters. *IEEE Ultrasonics Symposium,* 7–10 Nov., Seattle, WA

Park J-S, Gianchandani Y B 2003 A servo-controlled capacitive pressure sensor using a capped-cylinder structure microfabricated by a three-mask process. *J. Microelectromech. Syst.* **12(2)**, 209–20

Park J-S, Wilson C, Gianchandani Y B 2002 *MEMS Handbook* CRC Press, UK

Petersen K, Pourahmadi F, Brown J, Parsons P, Skinner M, Tudor J 1991 Resonant beam pressure sensor fabricated with silicon fusion bonding. *Int. Conf. Solid-State Sensors and Actuators, Transducers '91. . . Digest of Technical Papers (Cat. No. 91CH2817-5).* San Francisco, CA, USA. IEEE, San Francisco, CA, pp. 664–7,BN – 0 87942 585 7

Pirani M 1906 *Verh. Dtsch. Phys. Ges.* **8**, 686

Pramanik C, Saha H 2006 Piezoresistive pressure sensing by porous silicon membrane. *IEEE Sens. J.* **6**, 301–9

Puers B, Peters E, Bossche A V D, Sansen W 1990 A capacitive pressure sensor with low impedance output and active suppression of parasitic effects. *Sens. Actuators A* **21–3**, 108–14

Pulliam W, Russler P, Fielder R 2002 High-temperature, high-bandwidth, fiber optic, MEMS pressure sensor technology for turbine engine component testing. *Proc. SPIE – The International Society for Optical Engineering: Fiber Optic Sensor Technology and Applications 2001,* Newton, MA, USA, October 30–November 1, 2001, pp. 229–38

Samaun S, Wise K D, Nielsen E, Angell J 1971 IC piezoresistive pressure sensor for biomedical instrumentation. *IEEE International Solid-State Circuits Conference,* Philadelphia, PA, USA, pp. 104–5

Sander C S, Knutti J W, Meindl J D 1980 A monolithic capacitive pressure sensor with pulse period output. *IEEE Trans. Electron Devices* **27**, 927–30

Schafer H, Graeger V, Kobs R 1989 Temperature-independent pressure sensors using polycrystalline silicon strain gauges. *Sens. Actuators A* **17**, 521–7

Schmidt M A 1998 Wafer-to-wafer bonding for microstructure formation. *Proc. IEEE* **86**, 1575–85

Seidel H, Csepregi L, Heuberger A, Baumgartel H 1990 Anisotropic etching of crystalline silicon in alkaline solutions. I. Orientation dependence and behavior of passivation layers. *J. Electrochem. Soc.* **137**, 3612–26

Smirnov I N 1977 Coefficient representing the deformation of the silicon lattice as a result of diffusion of boron and antimony. *Sov. Phys. Solid State* **19**, 859–60

Spencer R R, Angell J B 1990 A voltage-controlled duty-cycle oscillator. *IEEE J. Solid-State Circuits* **25**, 274–81

Stangel K, Kolnsberg S, Hammerschmidt D, Hosticka B J, Trieu H K, Mokwa W 2001 A programmable intraocular CMOS pressure sensor system implant. *IEEE J. Solid-State Circuits* **36**, 1094–100

Stark B, Yuhai M, Zhang C, Najafi K 2003 A doubly anchored surface micromachined Pirani gauge for vacuum package characterization. *IEEE 16th Annu. Int. Conf. Micro Electro Mechanical Systems,* Kyoto, Japan, pp. 506–9

Stemme E, Stemme G 1990 A balanced resonant pressure sensor. *5th Int. Conf. Solid-State Sensors and Actuators and Eurosensors III*, Montreux, Switzerland, Vol. **A21,** *Sens. Actuators A Phys.*, pp. 336–41

Tellier C R, Brahim-Bounab A 1994 Anisotropic etching of silicon crystals in KOH solution. *J. Mater. Sci.* **29**, 5953–71

Timoshenko S 1959 *Theory of Plates and Shells*, 2nd edn. McGraw-Hill Co, New York, NY

Tufte O, Chapman P, Long D 1962 Silicon diffused-element piezoresistive diaphragms. *J. Appl. Phys.* **33**, 3322–7

Wolffenbuttel R F, Wise K D 1994 Low-temperature silicon wafer-to-wafer bonding using gold at eutectic temperature. *Sens. Actuators A* **43**, 223–9

Wolthuis R, Mitchell G, Hartl J, Saaski E 1993 Development of a dual function sensor system for measuring pressure and temperature at the tip of a single optical fiber. *IEEE Trans. Biomed. Eng.* **40**, 298–302

Wu C, Zorman C, Mehregany M 2006 Fabrication and testing of bulk micromachined silicon carbide piezoresistive pressure sensors for high temperature applications. *IEEE Sens. J.* **6**, 316–24

Yun W, Howe R T, Gray P R 1992 Surface micromachined digitally force-balanced accelerometer with integrated CMOS detection circuitry. *Solid-State Sensor and Actuator Workshop,* Hilton Head, SC, USA, pp. 126–31

Biographies

Andrew D. DeHennis received the BS degree in physics from Millersville University in 1997 and the MS and Ph.D. degrees in electrical engineering from the University of Michigan in 2001 and 2004, respectively. From 1997 to 1999 he worked at Metrologic Instruments, Inc where his effort focused on the development of highly automated dimensioning and tracking systems. His research interests are in remotely powered wireless microsystems for biomedical and environmental monitoring applications. After graduating with his doctorate, he joined Sensors for Medicine and Science, Inc. and is currently working on the development of a fully-implantable glucose monitoring system.

Junseok Chae received the B.S. degree in metallurgical engineering from the Korea University, Seoul, Korea, in 1998, and the M.S. and Ph.D. degrees in EECS (Electrical Engineering and Computer Science) from the University of Michigan, Ann Arbor, in 2000 and 2003, respectively. After a couple of years of being research fellow at Michigan, he is now at Arizona State University as an assistant professor in electrical engineering. His areas of interests are MEMS sensors/actuators, MEMS integration with electronics, micro-packaging, micro/nano-fluidics, and bio-MEMS.

He has published over 20 journal and conference articles and a book chapter, "Monolithically Integrated Inertial Sensors" in the 2nd volume of Advanced Micro and Nanosystems (AMN), CMOS-based MEMS and NEMS, Wiley-VCH series. He holds two US patents.

2.05 Accelerometers

Huikai Xie[1], Gary K Fedder[2], and Robert E Sulouff[3], [1]University of Florida, Gainesville, FL, USA, [2]Carnegie Mellon University, Pittsburgh, PA, USA, [3]Analog Devices Inc., Cambridge, MA, USA

Glossary

DETF Double-Ended Tuning Fork
DIP Dual Inline Package
DRIE Deep Reactive Ion Etch
DSB Doubly Sustained Beam
EMI Electromagnetic Interference
FBG Fiber Bragg Grating
FIB Focused Ion Beam
FPI Fabry–Perot Interferometer
LCC Leadless Chip Carrier
MZI Mach–Zehnder Interferometer
SCS Single-Crystal Silicon
SEM Scanning Electron Microscopy
STM Scanning Tunneling Microscopy
VCSEL Vertical Cavity Surface-Emitting Laser

2.05.1 Introduction

The first micromachined accelerometer was demonstrated in the late 1970s by using the piezoresistive effect based on a bulk micromachining process (Roylance and Angell 1978, 1979). It was not until the late 1980s that interest in micromachined accelerometers became more intense. Various sensing mechanisms, including piezoresistive, piezoelectric, capacitive, and tunneling, and thermal and optical sensors were extensively investigated using both bulk and surface micromachining techniques. The main driving force behind micromachined accelerometers was the automotive airbag deployment application due to the demand for lower cost and improved reliability and performance over traditional mechanical accelerometers (MacDonald 1990, Sulouff 1991). In the mid-1990s, many commercial micromachined accelerometers were brought into the market mainly for automotive uses. Starting from the beginning of this new century, the use of micromachined accelerometers has rapidly expanded in different fields such as portable electronics, video games, and medical applications. The following discussion attempts to give a historical perspective of the accelerometer development over the past few decades.

Before micromachined accelerometers, mechanisms consisting of pendulums and strain gauges were typically used to measure acceleration (IEEE-Std-337 1972, Sawyer and Ishii 1969). For example, one of the early Bosch accelerometers used a beryllium copper spring and a strain gauge with the mass suspended in oil for damping. This accelerometer, shown in **Figure 1**, was used in early automotive applications to measure negative acceleration and judge the impact for airbag safety systems.

Strain gauges with high gauge factors (GFs) that occur with doped single-crystal silicon (SCS) are a transduction mechanism that has found its way into various accelerometer designs in the past 30 years. Because SCS is a readily available material for the integrated circuit (IC) electronics industry, it has been exploited as the building block for accelerometers. Kulite used this approach in some of the earliest piezoresistive accelerometers (Kurtz 1976). The cantilever beam mechanical shape with additional mass added on the free end and a strain gauge at the mount or clamped end have been used in different sizes and shapes to build accelerometers. The SensoNor design as shown in **Figure 2** created a diffused strain gauge on the top and bottom of a silicon beam and achieved an electrical connection with a gold–tin alloy. The beam was clamped in the

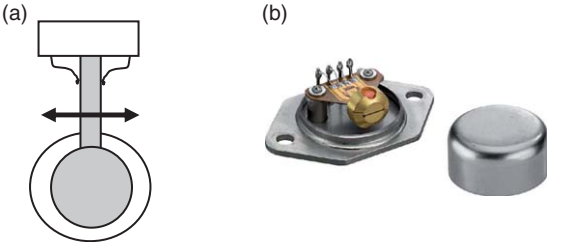


Figure 1 Schematic and photo of Bosch pendulum accelerometer. (Photo courtesy: Wilhelm Frey.)

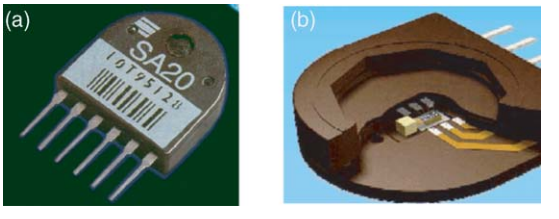


Figure 2 Photos of SensoNor SA 20 Crash Sensor.

housing base and additional mass was added to the end of the cantilever. Improvements to the design and packaging resulted in an etched structure and a plastic package with oil fill that went into large volume production in the early 1990s.

The cantilever beam accelerometer reported by Roylance and Angell (1979) used the mass of silicon created by anisotropic etching from the underside of a wafer along with boron diffusion and gold metallization from the front side to create a monolithic wafer approach. The monolithic microelectromechanical systems (MEMS) approach was further developed into a commercial product by Insouth Microsystems, Auburn, AL, in 1979, using aluminum metallization and a full piezoresistive bridge that was known as the G-Chip. The silicon accelerometer, double side processed and anisotropically etched, was mounted in a leadless chip carrier (LCC) package and combined with electrical amplification and compensation, as shown in **Figure 3**.

Anisotropically etched piezoresistive accelerometers by Delphi, IC Sensors, Nova Sensors, Denso, Bosch, and Endevco, among others, were also developed for instrumentation and automotive applications in the early 1990s. Due to the demand for automotive applications, especially airbag safety systems, many organizations entered the accelerometer field with innovative designs.

In the late 1980s and early 1990s, the use of polysilicon as an accelerometer mechanical spring and mass structure was investigated by numerous university groups, such as the University of California at Berkeley (UC-Berkeley) (Boser and Howe 1995, Lu *et al.* 1995), the University of Wisconsin (Guckel *et al.* 1992), and the University of Berlin (Fricke and Obermeier 1993). Polysilicon is a common material used in silicon gate complementary metal oxide semiconductor (CMOS) transistors and therefore is compatible with IC processing. Early work on this approach by Analog Devices Inc. (ADI) in 1989 resulted in the ADXL50, which was the first integrated surface micromachined accelerometer. An ADXL50 is shown in **Figure 4**. The ADI accelerometer used interdigitated comb fingers to sense a differential capacitance change when the mass supported by the polysilicon springs was accelerated and displaced. By integrating the electronics with the mechanical structure, a single small die was possible. This approach made it possible to add additional circuit functions and multiple axes of sensing by changing the layout while maintaining a common process.

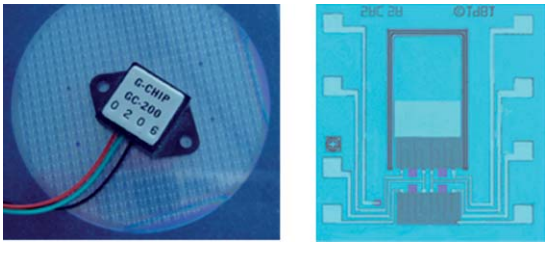
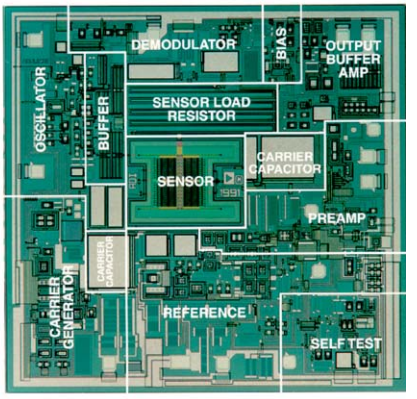


Figure 3 Photo of Insouth Microsystems G-Chip.

Capacitive sensing of displacement of a polysilicon structure was developed and introduced by Motorola in the early 1990s (Ristic *et al.* 1992). The Motorola structure used a multilayer polysilicon and a silicon-capped G-Cell (**Figure 5**) that was then packaged with electronics. Bulk micromachined capacitive accelerometers were also developed in the early 1990s by Centre Suisse d'Electronique et de Microtechnique SA (CSEM, Rudolf and Bergqvist 1991), Litton (Warren 1996), Applied MEMS in late 1990s, and the University of Michigan (Najafi *et al.* 2003), among many others.

The metal-plated accelerometer design created by Silicon Designs Inc. in the early 1980s used an off-center mechanical mass and a torsion spring to sense differential capacitance and therefore acceleration (Cole 1991). This design was then converted to a polysilicon mechanical structure by Ford and manufactured for a limited time in automotive airbag applications (Kemp and Spangler 1995). Capacitive designs have been a popular approach to accelerometers due to the low-temperature coefficient of the measurement as compared to piezoresistive designs and because of the opportunity to use an electrostatic voltage on the capacitive plates to perform a self-test.

Automotive airbag applications have fueled the demand for very large volumes of low-cost and high-reliability accelerometers. The initial application replaced a mechanical switch that could detect a negative acceleration and inflate an airbag. The accelerometer provided more information at a significantly reduced price and could be used in all systems and vehicles by a software change in the electronics. With the addition of airbags throughout the car, such as variable inflate passenger and side impact airbags, additional accelerometers per vehicle have been required. This has continued to fuel the demand for accelerometers with growth rates of over 20% per year for over a decade. In recent years, the use of accelerometers to control vehicle stability and



Analog Devices' ADXL-50, the industry's first surface micromachined accelerometer, includes on-chip signal conditioning.

Figure 4 Photo of ADXL50.

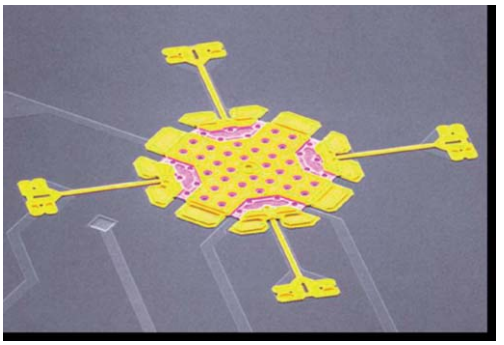


Figure 5 Die photo of a Motorola G-Cell.

rollover has added to the demand, especially in the low-gee range of 1–2 gees (also *g*). Companies such as VTI, LETI, Bosch, Denso, and ADI have supplied low-range accelerometers in adaptive braking and vehicle stability applications.

With the introduction of low-gee accelerometers for automotive applications, a wide range of applications in the consumer field have emerged. Accelerometers that can measure tilt and motion have been used in wireless mouse applications for interface to computers and video games. Sports applications such as golfing, tennis, or baseball have used accelerometers to analyze and coach players. A popular application for laptop computers has been to protect the data and disk drive due to mechanical shock. With the availability of three-axis accelerometers and prices less than $2, the feasibility of adding motion sensing to a wide range of hand-held devices has occurred. Cell phones, with the ability to recognize gestures and perform situation analysis for improved display orientation and battery life, promise to provide a market demand in excess of 700 million devices per year. When combined with automotive and other applications, the world market for accelerometers is expected to grow to over 1 billion devices.

In the remaining part of this chapter, we will introduce basic accelerometer concepts in Section 2.05.2 and examine various acceleration sensing mechanisms in Section 2.05.3. Some representative examples of accelerometer fabrication are presented in Section 2.05.4. Then the design considerations, packaging issues, and application considerations are discussed in Section 2.05.5, 2.05.6, and 2.05.7, respectively. Summary and future trends are given in Section 2.05.8.

2.05.2 Basic Accelerometer Concepts

Acceleration is the change in velocity for a given time period. If we confine our discussions to rectilinear motion, then a mass moving in a straight line has increasing acceleration with increasing velocity, described as scalar with a positive sign, and if the velocity is decreasing, the acceleration value is negative. Acceleration can be positive in either direction along the straight line of travel as it is dependent on the change in velocity. The most common error in acceleration measurement is sign. To ensure a good transferable convention, the earth's gravitational field is used and direction is then applied to give a positive acceleration when a reference arrow on the accelerometer or outline drawing is noted in relation to the

earth and the electrical output is referenced to ground.

To understand accelerometer operation, we first start with some of the basic building blocks of mechanics. Newton's second law of motion describes a relationship between force, mass, and acceleration. Specifically, this is $F = ma$, where F is the force, m is the mass, and a is the acceleration. This makes it possible to determine acceleration by measurement of the force on a given mass. A mass at the earth's surface has an acceleration of 9.8 m s^{-2} or 1 g if left to free fall. The change in velocity as a function of time of a mass is acceleration and as per Newton's second law would require a force to create such a velocity change. MEMS accelerometers are devices that determine the force required to create a velocity change. By keeping the mass constant and accelerating the frame, as illustrated in **Figure 6**, the spring, which connects the frame to the mass, is acted upon by the force due to acceleration and changes length. The mass is often called proof mass or seismic mass. The spring can either extend or contract as a result of positive or negative acceleration.

To measure force, it is common to use a spring that has a linear displacement with the force, or $F = kx$, where k is the ratio of force to displacement and is called the spring constant. The measurement of the proof mass displacement in relation to the frame thus becomes a direct indication of acceleration. Because we are referring to changes as a function of time, the dynamic or time-dependent aspects of the force measurement are also important considerations. When there is no drag force to oppose the acceleration force, displacement occurs in direct proportion to the acceleration. When the velocity and the direction are reversed, inertia creates an overtravel in the displacement and adds force to the spring in opposition to the force due to acceleration. If this motion continues in a harmonic fashion with an oscillation frequency that adds force in opposition to the acceleration and no damping or drag forces occur, displacement in excess of the acceleration force will result around the natural resonant frequency of the spring–mass system. The displacement at the natural frequency is maximized 90° out of phase with the frame acceleration. However, spring–mass systems in most accelerometer configurations have drag and viscous forces that act upon the moving mass and therefore dampen the oscillations. These damping forces can create nonlinear displacements with frequency and make it imperative that bandwidth and usable frequency ranges be considered.

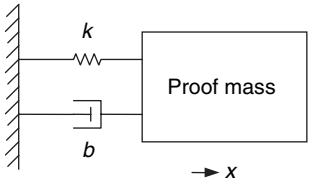


Figure 6 Spring–mass system.

The governing equation for the spring–mass system is readily given by

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = ma \qquad [1]$$

which results in the transfer function of the system,

$$H(s) = \frac{X(s)}{A(s)} = \frac{1}{s^2 + (b/m)s + (k/m)} = \frac{1}{s^2 + (\omega_0/Q)s + \omega_0^2} \qquad [2]$$

where b is the damping coefficient of the proof mass, ω_0 is the resonant frequency defined by $\omega_0 = \sqrt{k/m}$, and Q is the quality factor defined by $Q = (m\omega_0/b)$. For accelerometers working at a frequency much lower than the resonant frequency, the mechanical sensitivity can be obtained by setting $s = 0$ in the denominator in eqn [2], which yields

$$\frac{x}{a} = \frac{m}{k} = \frac{1}{\omega_0^2} \qquad [3]$$

It is interesting to note that the mechanical sensitivity is inversely proportional to the square of the resonant frequency. This represents a fundamental design trade-off between sensitivity and bandwidth. Thermomechanical (Brownian) noise limits the ultimate achievable resolution. The Brownian noise spectral density, referred to the input acceleration, is given by

$$\sqrt{\frac{a_n^2}{\Delta f}} = \frac{\sqrt{4k_B Tb}}{m} \qquad [4]$$

where a_n is the Brownian equivalent acceleration noise, Δf is the bandwidth of the system, which is usually set by the interface circuit, k_B is the Boltzmann's constant, and T is the absolute temperature in Kelvin. The unit is m s^{-2} Hz$^{-1/2}$ or g Hz$^{-1/2}$.

The natural resonances of MEMS accelerometers are in the range of 100 Hz to 20 kHz, and can vary from underdamped with Qs of 100 or more to overdamped with Qs below 1. **Figure 7** shows the schematic of a system for testing accelerometers. The accelerometer housing is connected through a

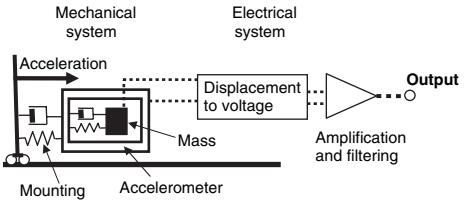


Figure 7 Schematic of an acceleration measuring system.

mounting structure to the module or assembly. As this mounting can create a spring–mass system with additional resonance and damping characteristics, the mechanical input and the accelerometer's indicated acceleration can be different. The accelerometer has a mechanical spring–mass and damper configuration, which also establishes a transfer function from the case or package to the moving mass. The displacement of the mass in relation to the accelerometer frame or housing is detected and converted to an electrical signal in several different ways, the most common being a piezoresistive bridge, a capacitive measurement, or a piezoelectric charge across a capacitor. There are advantages and error terms associated with each method, as will be discussed in Section 2.05.3. The voltage derived from the displacement transduction is then signal conditioned through amplification and filtering and perhaps digital conversion. The electrical signal transfer function is influenced by temperature, power supply changes, electrical noise, frequency, and load characteristics.

2.05.3 Acceleration Sensing Mechanisms

2.05.3.1 Piezoresistive Accelerometers

The resistance R of certain kinds of resistors changes with an applied mechanical load. This behavior is called the piezoresistive effect and can be expressed as

$$\frac{\mathrm{d}R}{R} = (1 + 2\nu)\varepsilon + \frac{\mathrm{d}\rho}{\rho} \qquad [5]$$

where ν is the Poisson's ratio, ε is the strain, and ρ is the resistivity. Note that there are two terms on the right-hand side of eqn [5]. The first term represents the resistance change due to geometry, while the second term is due to stress-induced resistivity change. The stress-induced resistivity is negligible for metals but is very large in semiconductors such as silicon. For a more convenient comparison of different materials, normalized resistance change, or GF, is normally used and it is readily obtained from eqn [5] as

$$\mathrm{GF} = \frac{\mathrm{d}R/R}{\varepsilon} = (1 + 2\upsilon) + \frac{\mathrm{d}\rho/\rho}{\varepsilon} \qquad [6]$$

GF is about 2 for metals as the Poisson's ratio of most metals is about 0.35. Semiconductors have much higher GF as the second term in eqn [6] for semiconductors is on the order of 100. Silicon is the main material for making piezoresistive accelerometers. The GF of silicon is strongly dependent on the doping level. Lower doping concentration yields higher sensitivity. However, devices with higher doping concentration are less sensitive to temperature change. Thus, the doping concentration must be carefully chosen. Detailed discussion can be found in Madou (2002).

In its simplest form, a piezoresistive accelerometer comprises a cantilever beam with a proof mass attached to it, as shown in **Figure 8(a)**. Because the resistance change of a piezoresistor increases with increasing stress, the piezoresistor is typically placed at the anchor edge of the cantilever beam where the stress is maximum. Both in-plane (lateral-axis) and out-of-plane (z-axis) acceleration sensing can be realized. Differential bridges can be formed for lateral-axis sensing either by using a T-shaped cantilever beam (**Figure 8(b)**) (Chen *et al.* 1997) or by forming piezoresistors on the sidewalls of a cantilever beam using a sidewall ion implantation (**Figure 8(c)**) (Partridge *et al.* 2000). When the beam bends laterally, the two piezoresistors on the sidewalls will experience stresses with opposite signs, and thus one piezoresistor's resistance decreases while the other's increases. Lateral acceleration sensing was also demonstrated by using a (110) wafer for obtaining wet-etched vertical sidewalls (Suminto 1991). z-Axis acceleration sensing typically requires a reference resistor fixed on the substrate because it is difficult to form a symmetric diffusion layer on the underside of cantilever beams.

Due to the vector nature of acceleration, measuring acceleration in all the three axes is often required. To assemble three single-axis accelerometers will drastically increase the packaging size and cost. Therefore, many efforts have been devoted to developing monolithic three-axis (or triaxial) piezoresistive accelerometers (Amarasinghe *et al.* 2007, Chaehoi *et al.* 2005, Kruglick *et al.* 1998, Kwon and Park 1997, Kyung Il *et al.* 2003, Plaza *et al.*

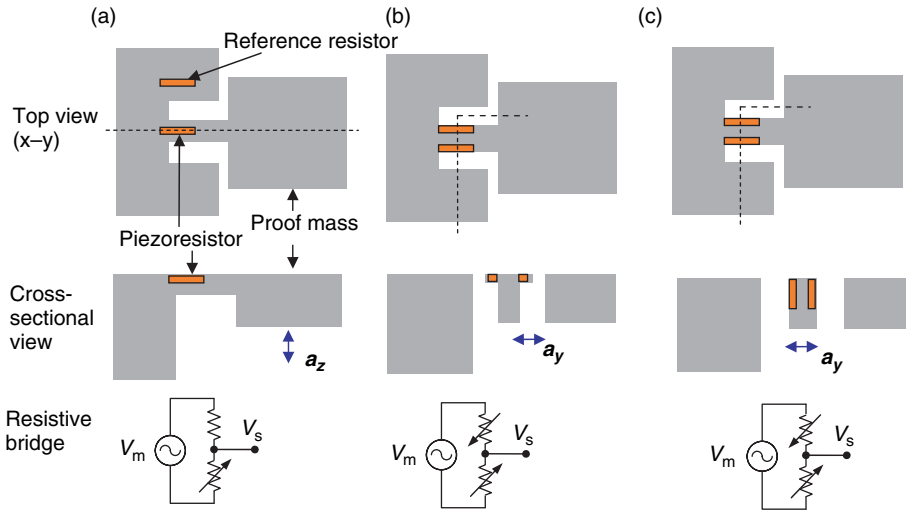


Figure 8 Schematics of piezoresistive sensing. (a) Out-of-plane (*z*-axis) sensing with a quarter bridge. (b) In-plane (lateral-axis) sensing with a half-bridge where the piezoresistors are formed from the frontside diffusion. (c) In-plane (lateral-axis) sensing with a half-bridge where the piezoresistors are formed on the sidewalls by inclined ion implantation.

1998, Rdjegrd *et al.* 2005, Takao *et al.* 1995). The challenge to fabricate three-axis piezoresistive accelerometers has been the lateral-axis sensing. A commonly used three-axis piezoresistive accelerometer design employs a frame mass suspended by four symmetric beams, which are connected to the center anchor, as shown in **Figure 9**. The five piezoresistors are connected in series and driven by a constant current source. The center piezoresistor (R5) is used as a reference resistor to eliminate the thermal drift of the piezoresistors. For an *x*-axis acceleration, R1 and R3 change their values oppositely while R2 and R4 change with the same value and sign. Thus, the *x*-axis acceleration can be detected by the differential output of R1 and R3. Similarly the *y*-axis acceleration can be obtained by the differential of R2 and R4. The sum of all four piezoresistors results in the *z*-axis acceleration. This design employs a simple structure and a straightforward fabrication process. However, the lateral-axis sensitivity is much lower than the *z*-axis sensitivity, and DC offsets are also large.

Another solution to the lateral-axis sensing is to make piezoresistive structures aligned vertically with respect to the substrate. Kruglick *et al.* (1998) reported a three-axis piezoresistive accelerometer with thin-film structures that were bent up out of plane. As shown in **Figure 10**, the out-of-plane structures are used to detect *x*- and *y*-axis acceleration. The piezoresistors are located at the hinges of the bent-up

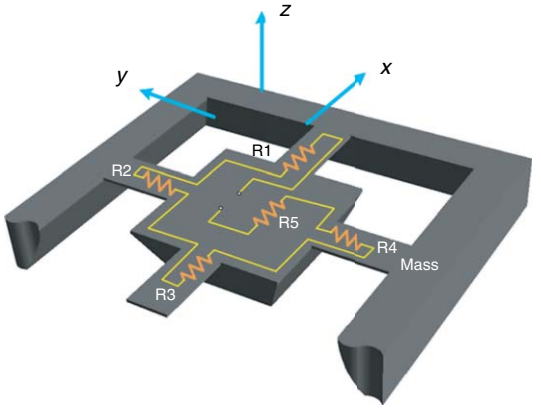


Figure 9 Schematic view of a bulk micromachined piezoresistive three-axis accelerometer.

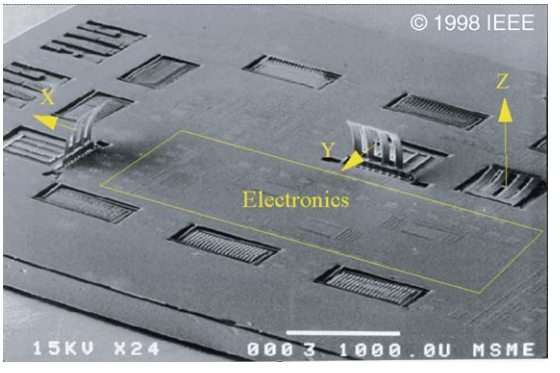


Figure 10 Scanning electron micrograph (SEM) of a surface-micromachined piezoresistive three-axis complementary metal oxide semiconductor (CMOS) accelerometer.

structures. This accelerometer was integrated with CMOS circuits. However, the extra assembly step, low GF of polysilicon, robustness, and temperature stability are still concerns.

The advantages of piezoresistive accelerometers include the simplicity of structural design and readout circuitry due to the low impedance at the sensor output. Their fabrication processes are also relatively simple. A bulk micromachined accelerometer can be made in two steps: one backside aqueous KOH etch to form thin sensing beams and then a frontside KOH etch to release the proof mass. After that, a top cover and a bottom cover are bonded to seal the accelerometer. The covers provide overload stops and damping control. The packaging process can also be simplified by using an integrated overload stop (Barth *et al.* 1988).

One of the key process issues for bulk micromachined piezoresistive accelerometers is the etch stop that controls the thickness of the sensing beam. A simple timed etch often yields poor thickness control. Electrochemical stop and p+ doping provide automatic etch stop but require extra steps. Another possible etch stop is a buried oxide (BOX) layer. Plaza *et al.* (2002) used silicon-on-insulator (SOI) wafers to achieve automatic etch stop and overrange protection as well. The cross-sectional view of the device is shown in **Figure 11**.

Because the main process step for making bulk piezoresistive accelerometers is wet silicon etching, the major obstacle to integrate piezoresistive accelerometers is the protection of ICs during this step. Many researchers have demonstrated integrated piezoresistive accelerometers, and most of them are foundry CMOS compatible (Chaehoi *et al.* 2003, 2005, Kruglick *et al.* 1998, Riethmuller *et al.* 1991, Seidel *et al.* 1995, Shi *et al.* 2004). Furthermore, deep reactive ion etch (DRIE) of silicon has also been used to fabricate piezoresistive accelerometers (Huang *et al.* 2003). This dry etch technique is more expensive than wet etching, but it greatly simplifies the integration process. A desirable attribute of piezoresistive accelerometers is that the output impedance of a piezoresistive sensor node is low ($\sim$kΩ) compared to that in capacitive sensing approaches. Thus, chip-to-chip wire bonding does not attenuate the sensing signal significantly. Therefore, two-chip or hybrid packaging is a more viable approach for micromachined piezoresistive accelerometers.

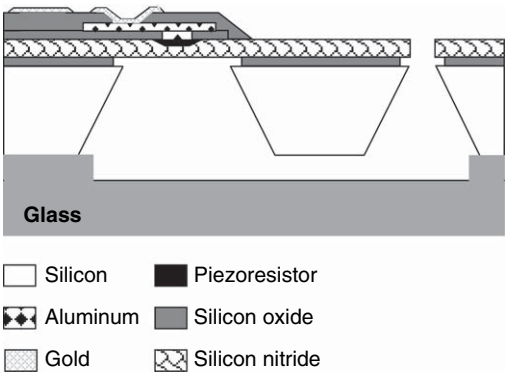


Figure 11 Schematic cross section of a silicon on insulator (SOI) piezoresistive accelerometer with integral overrange protection tabs. (Source: Plaza J A, Collado A, Cabruja E, Esteve J 2002 Piezoresistive accelerometers for MCM package. *J. Microelectromech. Syst.* **11**, 794–801.)

2.05.3.2 Piezoelectric Accelerometers

Piezoelectricity is the ability of some materials to create an internal polarization in response to strain. In other words, when a force or stress is applied to a piezoelectric material, a voltage is generated, which is a measure of the force or stress. The piezoelectric effect is reversible in that piezoelectric crystals, when subjected to an externally applied voltage, will produce a strain, i.e., can be used as actuators. This is extremely useful in the case of resonant accelerometers where both actuation and sensing are required in a single device. There are three basic classes of piezoelectric materials used in microfabrication: piezoelectric crystals such as quartz, lithium niobate, and gallium phosphate; thin-film piezoelectrics, such as zinc oxide, aluminum nitride, and lead zirconate titanate (PZT); and polymer-film piezoelectrics, such as polyvinylidene fluoride (PVDF). Note that piezoelectrics in general are not suitable for quasi-static sensing of strain because of the parasitic effects of small DC leakage currents, but they are highly effective for sensing of vibratory or resonant motions.

A basic piezoelectric cell is sketched in **Figure 12**. The piezoelectric effect can be expressed as

$$q = d_{33}F \qquad [7]$$

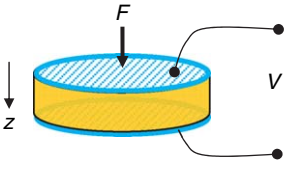


Figure 12 Piezoelectric effect.

where q is the total charge, d_{33} is the piezoelectric charge constant in units of $C\,N^{-1}$ (d_{33} refers to the case where both induced polarization and applied stress are in direction 3 or z), and F is the external force. The output voltage is then given by

$$V = \frac{q}{C} = \frac{d_{33}F}{C} \qquad [8]$$

where C is the sensing capacitance formed by the piezoelectric layer. According to this equation, a small capacitance will yield high sensitivity. However, small sensing capacitance also has a large impedance, which in turn reduces the signal in the presence of parasitic interconnect capacitance. Thus the placement of the electrodes and the piezoelectric material must be carefully considered. The design goal is to maximize C and q/C at the same time. One good practice is to place the electrodes around the maximum stress region(s).

A basic piezoelectric accelerometer design is illustrated in **Figure 13**. It is essentially a sandwiched structure with the piezoelectric layer in the middle and two conducting layers on the top and bottom, respectively. As discussed above, the upper electrode only covers the near-anchor portion of the cantilever beam where the stress is maximal. The piezoelectric layer can be either surrounded by dielectrics (**Figure 13(b)**) or directly in contact with the electrodes (**Figure 13(c)**). The dielectric isolation can significantly reduce the leakage current and improve the quasi-static response performance (Chen *et al.* 1982). However, adding the isolation layer also increases process complexity and reduces the sensitivity.

Similar to three-axis piezoresistive accelerometers, monolithic three-axis piezoelectric accelerometers can also be realized using a single microstructure (Okada 1995, Zou *et al.* 2004). One of the many possible designs is shown in **Figure 14**. The center proof mass is connected to the substrate via four symmetric beams. When there is a z-axis acceleration, the proof mass produces tensile stresses in the top halves of the bimorph beams (i.e., x_1, x_2, y_1, and y_2), and compressive stresses in z_1–z_4. Thus, the sum of z_1 to $-z_4$ measures the z-axis acceleration. The differential components x_1-x_2 and y_1-y_2 are zero. When there is an x-axis acceleration, the proof mass rotates around the y-axis (**Figure 14(b)**) and produces tensile stresses in x_1 and z_3 and compressive stresses in z_1 and x_2. It also produces shear stresses in y_1, y_2, z_2, and z_4, which can be neglected. Thus, an x-axis acceleration produces a signal from the differential of x_1 and x_2. The differential of y_1 and y_2 and the sum of z_1 to z_4 are both zero because the stresses in y_1 and y_2 have the same amplitude and sign while z_1 and z_3 have the same amplitude but opposite signs. The same principle applies to the y-direction.

Among various piezoelectric materials, quartz is preferred because of its high-temperature stability and high quality factor. High-performance, wide-bandwidth quartz piezoelectric accelerometer products are manufactured by many companies such as Kistler, Endevco, and Honeywell. Resonating beams are typically used for quartz accelerometers due to the high quality factor, which will be further discussed in Section 2.05.3.7.

ZnO and PZT are two other main materials for piezoelectric accelerometers. Because these materials can be deposited as thin films, surface micromachining as well as bulk micromachining can be used. **Figure 15** shows two surface micromachining processes that use oxide and silicon substrate,

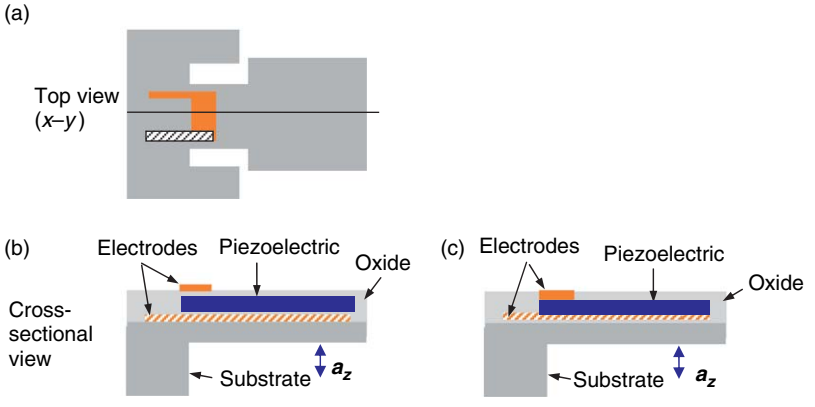


Figure 13 A typical z-axis piezoelectric accelerometer structure.

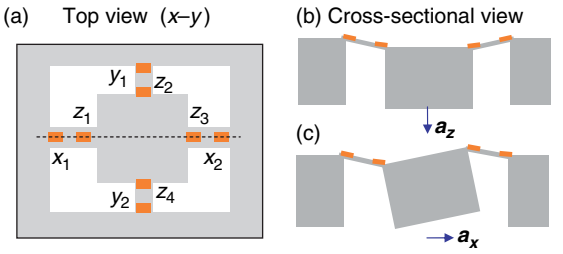


Figure 14 Schematic of three-axis piezoelectric accelerometer made with a single proof mass.

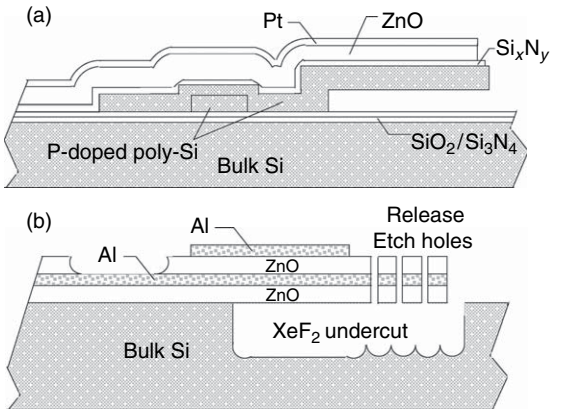


Figure 15 Cross-sections of cantilever structures formed in two processes based on (a) sacrificial oxide and (b) sacrificial silicon fabrication. (Source: Devoe D L, Pisano A P 2001 Surface micromachined piezoelectric accelerometers (PiXLs). *J. Microelectromech. Syst.* **10**, 180–6.)

respectively, as the sacrificial layer (Devoe and Pisano 2001). A sensitivity of 45 fC g^{-1} was demonstrated. Wang *et al.* (2002) used bulk micromachining to suspend a bulk silicon proof mass on a thick-film PZT annular diaphragm. The sensitivity was about 1–7 pC g^{-1}. In addition, ZnO can be deposited using magnetron sputtering at low temperature, so ZnO-based piezoelectric accelerometers can be made CMOS compatible (Chen *et al.* 1982). Furthermore, PVDF has also been used to make piezoelectric accelerometers (Benech *et al.* 1996, Marcal *et al.* 1997, Toda and Thompson 2006).

2.05.3.3 Capacitive Accelerometers

Capacitive accelerometers are dominating the automotive and consumer electronics markets due to their high sensitivity, good temperature performance, low fabrication cost, small size, and easy integration with CMOS. Capacitive displacement sensing structures are very simple. They basically are a pair of electrodes with one of them movable. The two most commonly used electrodes are parallel plates and interdigitated comb fingers. In microfabrication, the lateral (in-plane) dimensions can be easily made large (up to a few tens of millimeters), but the vertical (out-of-plane) dimensions are typically restricted to a few microns. Thus, parallel plate electrodes are normally used for vertical acceleration sensing while comb-finger electrodes are suitable for lateral acceleration sensing.

Figure 16 shows the basic configurations for capacitive displacement sensing with horizontal plates. The suspended plate can move either vertically (**Figure 16(a)** and **16(b)**) or laterally (**Figure 16(c)**). The configuration with a fully differential bridge (**Figure 16(b)**) is always preferred for DC offset and temperature variation cancellation. However, the configuration with a fixed reference capacitance is much easier to implement.

Figure 17 shows the configurations for capacitive displacement sensing with comb fingers. The configuration in **Figure 17(a)** is typically used for lateral displacement sensing. The three comb fingers form two capacitors C_1 and C_2. Each capacitance includes a sidewall parallel plate capacitance and a fringing capacitance. The middle finger is movable and often called a rotor. When the rotor moves along the x-direction, C_1 decreases while C_2 increases, resulting in a fully differential capacitive half-bridge, as shown in **Figure 17(b)**.

$$V_s = \frac{C_1 - C_2}{C_1 + C_2 + C_P} V_m \qquad [9]$$

where C_P is the parasitic capacitance. The parasitic capacitance attenuates the signal and should be minimized.

The configuration illustrated in **Figure 17(c)** is used for z-axis displacement sensing. The comb fingers consist of multiple conductor layers. When all the conductor layers are electrically connected, it becomes similar to what is shown in **Figure 17(a)**. Note that the conductor layers are separated by dielectrics and thus can be connected in many ways. **Figure 17(c)** is one of the wiring configurations, where the middle finger (rotor) has two electrodes and the stationary fingers (stators) have a single electrode each. With this wiring, similar to that shown in **Figure 17(a)**, two capacitors C_1 and C_2 are formed. The difference is that the capacitors in **Figure 17(c)** add up the capacitances on both sides of the rotor. Thus, a small gap variation in the lateral direction

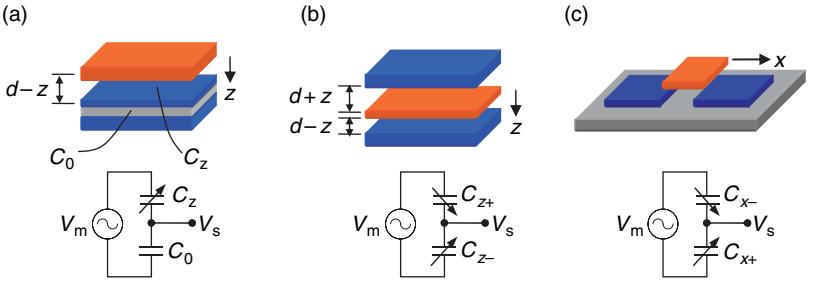


Figure 16 Basic capacitive displacement sensing configurations with horizontal plates. (a) Half differential capacitive bridge where a reference capacitor is embedded on the substrate. (b) Fully differential capacitive bridge where the movable plate is suspended in between two fixed plates. (c) A movable plate is suspended over two separate electrodes.

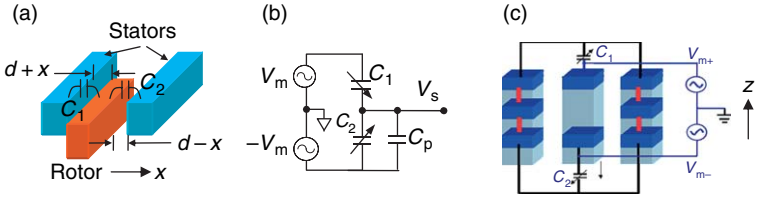


Figure 17 Basic capacitive displacement sensing configurations based on the sidewall capacitance of comb fingers.

will not change C_1 and C_2, which means that the device is insensitive to lateral acceleration. However, C_1 and C_2 are formed by splitting the sidewall capacitance into two parts vertically. When the rotor moves in the z-direction, as shown in Figure 17(c), C_1 decreases while C_2 increases, resulting in a differential output. One may notice that C_1 and C_2 are not equal at the original position. This initial capacitance difference will generate a large DC offset. This problem can be solved by using two sets of comb fingers and swapping the rotor and stator of the second set (Xie and Fedder 2002).

MEMS capacitive accelerometers can be categorized into three types: thin-film accelerometers fabricated using surface micromachining; bulk accelerometers fabricated using bulk micromachining and/or wafer bonding technology; and other accelerometers fabricated using combined surface and bulk micromachining processes. Surface micromachined accelerometers have been commercialized for more than 10 years (Cournar *et al.* 2004). Due to its good mechanical properties and mature deposition process, polysilicon has been the major structural material for surface micromachined accelerometers. Researchers from UC-Berkeley have demonstrated several generations of single- and three-axis capacitive accelerometers with noise floors on the order of 1 milli-$g\,\mathrm{Hz}^{-1/2}$ (Boser and Howe 1996, Lemkin and Boser 1999, Lemkin *et al.* 1997, Lu *et al.* 1995). The commercial accelerometers from ADI, Bosch, Motorola, Freescale, and ST Microelectronics (ST) are all based on thin-film or thick-film polysilicon. An alternative material is multilayer aluminum/oxide composite. Carnegie Mellon University developed a maskless, post-CMOS micromachining process that is foundry CMOS compatible (Fedder *et al.* 1996). This process has been used to demonstrate both lateral- and vertical-axis accelerometers (Luo *et al.* 2000, Xie and Fedder 2000, Zhang *et al.* 1999). One of the advantages of the multilayer composite is its capability to realize differential capacitive comb-finger displacement sensing in both lateral and vertical directions (Xie and Fedder 2002). The main drawbacks include the curling and poor temperature performance. Due to their thickness and size limitations, surface micromachined accelerometers normally can only achieve noise floors of 0.1–1 milli-$g\,\mathrm{Hz}^{-1/2}$. Another promising material is polycrystalline silicon germanium (poly-SiGe). Due to its low deposition temperature, poly-SiGe can be integrated with CMOS (Franke *et al.* 2003).

In contrast, most bulk micromachined capacitive accelerometers can reach micro-g or even submicro-g resolutions. For example, researchers from the University of Michigan have developed several innovative bulk micromachining processes and demonstrated many high-performance capacitive accelerometers (Chae *et al.* 2002, 2004, Selvakumar

et al. 1996, Yazdi and Najafi 1997, Yazdi *et al.* 1999). There are numerous other bulk micromachined capacitive accelerometers. Some of them are listed in **Table 1**. Most of the device fabrications employ wafer bonding and wet etching for obtaining large proof mass. SOI substrate is also becoming more popular, but the fabrication cost especially for CMOS integration is still high.

Other interesting capacitive accelerometers such as electrostatically elevated spherical accelerometers were also reported (Qu 1999, Takeda 2000, Toda *et al.* 2002). These technologies require special processes and equipment, which are currently too expensive for commercialization. To take the advantages of the large proof mass of bulk silicon structures and the readout circuitry integration of surface micromachining processes, a DRIE CMOS-MEMS process has been developed at Carnegie Mellon University (Xie *et al.* 2002). It is a maskless post-CMOS process that can create thick SCS structures using only dry etch steps. The process was further improved at the University of Florida and a three-axis capacitive accelerometer with a lateral-axis noise floor of 12 micro-$g\,\mathrm{Hz}^{-1/2}$ was demonstrated (Qu *et al.* 2006).

The resolution versus device size of capacitive accelerometers for different technological approaches is illustrated in **Figure 18**. The thin-film surface micromachined devices will continue dominating the low- to medium-end markets while the higher-resolution but more expensive bulk micromachined devices will find niche applications in high-end markets such as defense, space, security, and harsh environment. The DRIE CMOS-MEMS technology has the potential to achieve high resolutions at small form factors.

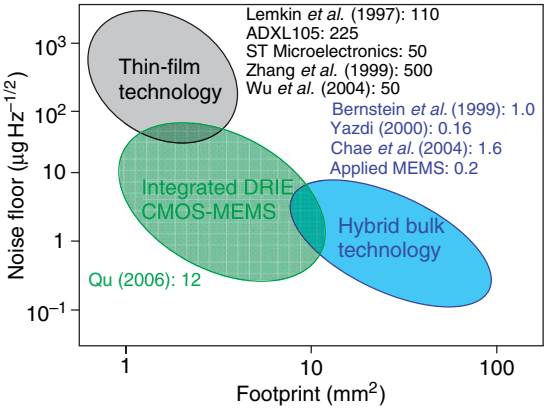


Figure 18 Capacitive accelerometers: resolution performance versus device size with different technological approaches. Noise values associated with representative references are given in units of μg $\mathrm{Hz}^{-1/2}$. (Source: Qu H 2006 Development of DRIE CMOS-MEMS process and integrated accelerometers. University of Florida.)

2.05.3.4 Thermal Accelerometers

Thermal equilibrium of a system will be disturbed if there exists a forced fluidic flow or a moving part due

Table 1 Summary of bulk micromachined capacitive accelerometers

Contributors	*Year published*	*Structure and/or process used*	*Noise floor*
Rudolf *et al.*	1990	Sandwiched silicon glass	1 micro-$g\,\mathrm{Hz}^{-1/2}$
Henrion *et al.*	1990	Sandwiched silicon glass Multistep wet etch	1 micro-$g\,\mathrm{Hz}^{-1/2}$ 260-Hz bandwidth
Warren	1994	SIMOX	N/A
Bernstein *et al*.	1999	Silicon–glass bonding Dual chips Combination of wet and dry etch Automatic wet etch stop	1 micro-$g\,\mathrm{Hz}^{-1/2}$ 1-kHz bandwidth
Kulah *et al*.	2000	Combination of bulk and surface micromachining Wet release	0.23 micro-$g\,\mathrm{Hz}^{-1/2}$
Chae *et al*.	2002	SOG structure Silicon–glass bonding Silicon thinning by CMP Dry release	100 micro-$g\,\mathrm{Hz}^{-1/2}$
Yazdi *et al*.	2003	Combination of bulk and surface micromachining Wet release	0.18 micro-$g\,\mathrm{Hz}^{-1/2}$
Applied MEMS		Bulk micromachining, wafer bonding	0.25 micro-$g\,\mathrm{Hz}^{-1/2}$

Source: Qu *et al.* (2006).
CMP, chemical–mechanical planarization; MEMS, microeectromechanical systems; SIMOX, Separation by IMplantation of OXygen; SOG, silicon on glass.

to an external acceleration. The temperature change at a fixed point of the disturbed system is a measure of the acceleration. There are basically two ways to measure acceleration thermally:

(1) To utilize a moving mass to change the heat transfer due to conduction and
(2) To use the natural convection of gas or mixed liquid/gas in a sealed chamber

We discuss each of the approaches below.

2.05.3.4.1 Thermal accelerometers based on conduction heat transfer

Hiratsuka *et al.* (1991) proposed the first micromachined thermal accelerometer and demonstrated the concept of using a thermopile and an integrated heater to measure small displacements of a proof mass. One of the advantages of thermal accelerometers is their immunity to electromagnetic interferences (EMIs). The working principle is illustrated in **Figure 19**. The device consists of a heated plate sandwiched by two heat sinks. The heat of the hot plate can be transferred through radiation, convection, and conduction. However, the gap between the hot plate and the heat sink is very small (in a few microns or less), so the convection can be neglected. In addition, as the temperature change is small, radiation is also negligible. Therefore, the conduction dominates. In this case, the heat flux transferred from the hot plate to the heat sinks is inversely proportional to the gap. In other words, the temperature change is proportional to the gap change. Hence, if one of the plates is connected to a proof mass, then the acceleration will be linearly proportional to the temperature change of the hot plate, which is typically measured by a thermopile. This type of thermal accelerometers has two possible implementations: one is to use the hot plate as the proof mass and the other is to use the sink as the proof mass.

Dauderstadt *et al.* demonstrated a thermal accelerometer using the top heat sink as the proof mass.

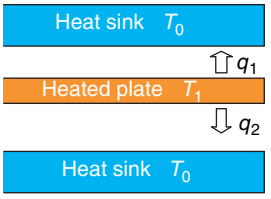


Figure 19 Cross section of a thermal accelerometer showing the concept of using conductive heat transfer to measure small displacements.

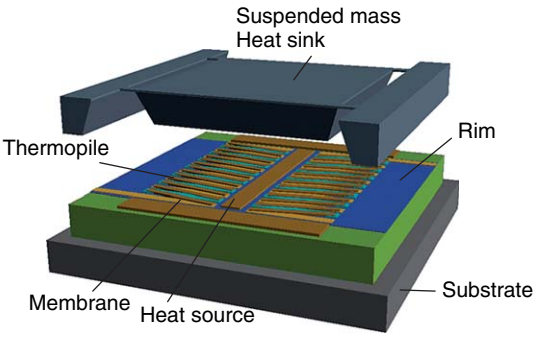


Figure 20 A conduction-based thermal accelerometer. (Source: Dauderstadt U A, de Vries P H S, Hiratsuka R, Sarro P M 1995 Silicon accelerometer based on thermopiles. *Sens. Actuators A Phys* **46**, 201–4.)

The device design is illustrated in **Figure 20**. It consists of two chips that are bonded together. The proof mass is formed in the top chip by KOH etching and p–n junction electrochemical etch stop. The suspension beams of the proof mass are etched using a reactive ion etching (RIE) step from the frontside of the wafer. Polysilicon–aluminum thermopiles are used to measure the temperature. The thermopiles are fabricated on the lower chip with a silicon nitride layer for thermal isolation. A KOH etch step is used to remove the silicon below the silicon nitride. In this way the heat conduction from the heat source to the substrate can be minimized. The polysilicon is also used to form the heater. The polysilicon–aluminum thermocouples have a relatively large Seebeck coefficient of about 0.2–0.4 mV K^{-1}. The tested sensitivity of the accelerometer is 20 mV g^{-1}. The accelerometer works at frequencies up to 300 Hz and has a temperature coefficient of -0.12% K^{-1} (Dauderstadt *et al.* 1998).

2.05.3.4.2 Thermal accelerometers based on convective heat transfer

Different from the conduction-based thermal accelerometers, convection-based thermal accelerometers do not need a solid proof mass and they have no moving mechanical structures. So convective thermal accelerometers are robust and can stand high shock. Convective thermal accelerometers come in two types: one uses only gas while the other has a liquid-filled chamber. We discuss the gas type first.

Dao *et al.* (1996) proposed an idea in which the temperature change due to convective heat transfer of gas in a sealed chamber is used to measure acceleration. Leung *et al.* (1998) demonstrated the first micromachined convective thermal accelerometer

in 1998. The device structure is illustrated in **Figure 21(a)**. It consists of three resistor strips on a silicon substrate. The silicon under the resistors is removed to provide good thermal isolation. The resistor on the center strip is used as a heater. When the heater is on, the density of the air surrounding it becomes smaller and there is a temperature gradient from the heater to either of the side strips. The resistors embedded in the two side strips are used as temperature sensors and form a half-bridge to output a differential voltage. When there is no acceleration, the heater produces a symmetrical temperature distribution (the solid lines in **Figure 21(b)**). Thus the differential output is zero. When an external acceleration is applied, the symmetry is disturbed as illustrated by the dashed lines in the same figure. The resulting temperature difference is picked up by the differential resistor bridge, which produces an output corresponding to the applied acceleration.

The device fabrication is straightforward. The heater and temperature sensors are made of a polysilicon layer, which is sandwiched by two oxide layers. The bonding pads are formed by electroplating gold. The final step is to etch the cavity using ethylene diamine pyrocatechol (EDP). After the microfabrication is finished, the device is packaged in a sealed chamber to prevent external air flow from disturbing the device operation.

The prototype reported in Leung *et al.* (1998) had a chip size of 3 mm × 5 mm. The temperature coefficient of resistivity of the polysilicon was measured as $\sim 0.2\%\ ^{\circ}C^{-1}$. The measured noise level was 0.5 milli-*g* rms at a heater power of 20 mW. It was found that the sensitivity was linearly proportional to the heater power. It was also found that the sensitivity was proportional to the square of the chamber pressure. This implies that the same device can be made 100 times more sensitive simply by packaging it into a chamber with a 10 times higher pressure. But it also puts stringent requirements on the packaging for long-term pressure control. The frequency range for the prototype device was from DC to 20 Hz.

The fabrication of convective thermal accelerometers can be made IC compatible. Milanovic *et al.* (2000) successfully fabricated such accelerometers in a 2 μm CMOS process plus one additional maskless post-CMOS micromachining process. There are some new developments in the past few years. For example, non-polysilicon heater/temperature sensing materials such as platinum have been investigated (Mailly *et al.* 2003). Porous silicon instead of a cavity has also been used to achieve good thermal isolation (Goustouridis *et al.* 2004). Furthermore, a liquid-filled thermal accelerometer with a thermally generated bubble was reported (Liao *et al.* 2006). The noise equivalent acceleration was found to be 1 mg rms. The measured response time was approximately 60 ms. This liquid-filled type has the potential to achieve faster response time but may consume more power.

The commercialization of convective thermal accelerometers is very successful. MEMSIC Inc. licensed the convective thermal accelerometer technology and has developed a series of monolithically

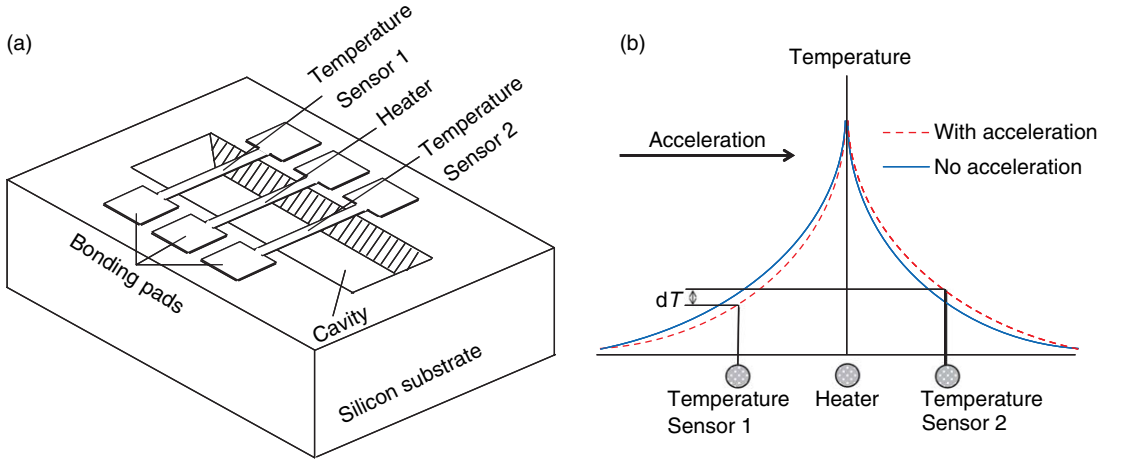


Figure 21 Device structure of the thermal accelerometer. (Source: Leung A M, Jones J, Czyzewska E, Chen J, Woods B 1998 Micromachined accelerometer based on convection heat transfer. *Proc. 11th IEEE Int. Workshop Micro Electro Mechanical Systems (MEMS '98)*.)

integrated accelerometers based on a standard IC process. **Figure 22** is a die photo of a MEMSIC's two-axis accelerometer. The high integration, small form factor, and low cost have made these products very competitive in consumer and electronics markets for low-bandwidth applications.

2.05.3.5 Optical Accelerometers

Optical sensors are known for their immunity to EMI, safety, and low cost. If optical fiber based, they also have low transmission loss, inherent multiplexing capability, and potential high resolution. Fiberoptic accelerometers have been demonstrated in the 1980s without using MEMS technology (Rines 1981, Spillman 1982). Resolutions of about 1 μg were achieved using a tilting mirror (Soref and McMahon 1984) or fiber microbending effect (Freal *et al.* 1987). However, these devices are bulky. It was not until the mid-1990s when research activities on micromachined optical accelerometers started to grow (e.g., Marty *et al.* 1995). Just like other optical sensors, optical accelerometers can be divided into two categories: one is intensity based and the other is interferometry based. We discuss micromachined optical accelerometers separately in each category.

2.05.3.5.1 Optical accelerometers with intensity modulation

Intensity modulation-based optical sensors have simple structures and good linearity, and they are also inexpensive. **Figure 23** shows various configurations

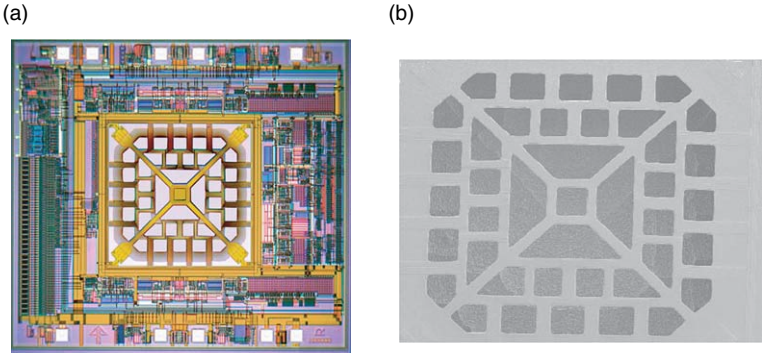


Figure 22 (a) Die photo of a MEMIC thermal accelerometer. (b) Scanning electron micrograph (SEM) of the thermal sensing element. (Photo courtesy: Y Cai of MEMSIC.)

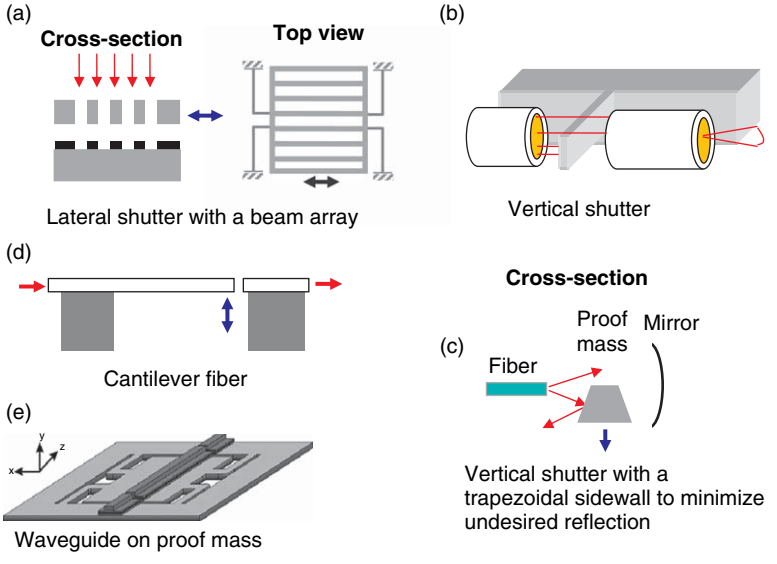


Figure 23 Schematic views of various intensity modulation optical accelerometers.

of optical intensity modulation micromachined accelerometers.

(a) This is a lateral-axis accelerometer. The shutter is placed between the light source and the photodetector. The shutter consists of alternating opaque/transparent regions. Note that the transparent regions are not necessarily through-holes. For instance, silicon is transparent to infrared light ($\lambda > 1\,\mu m$). This type of optical accelerometer has been implemented both in a bulk micromachining process using (110) silicon wafers (Abbaspour-Sani *et al.* 1995) and in a surface micromachining process (Tien 1998).
(b) This is a vertical (z-axis) optical accelerometer. The shutter can be obtained either by silicon DRIE or by anisotropic etch of (110) silicon substrate. An accelerometer with a 2-mg resolution and a 1-kHz bandwidth was demonstrated (Guldimann *et al.* 2000).
(c) This is also a z-axis optical accelerometer, but the shutter does not have vertical sidewalls. Instead, its cross section is a trapezoid. The slope reflects undesired light away from the fiber to minimize back reflection (Marty *et al.* 1995).
(d) This is a very simple structure. The fiber does it all, providing spring force, proof mass, and sensing element. However, cleaving and polishing the fiber are challenging (Kalenik and Pajak 1998).
(e) In this method, the waveguide is integrated on the proof mass (Llobera *et al.* 2004).

2.05.3.5.2 Interferometric optical accelerometers

Acceleration generates through a proof mass a displacement or stress, which can be measured by optical interferometers. The commonly used interferometers include the Michelson interferometer, the Mach–Zehnder interferometer (MZI), and the Fabry–Perot interferometer (FPI). For micromachined accelerometers, FPIs are often used due to their compactness and high resolution. The working principle of FPIs is based on multiple beam interference. A simplified Fabry–Perot (FP) etalon is sketched in **Figure 24**. Interference fringes will be generated due to multiple reflections when a light beam is incident on the etalon. The condition for a fringe maximum is given by

$$\frac{4\pi}{\lambda_0} nd\cos\theta = 2N\pi - \delta_r \qquad [10]$$

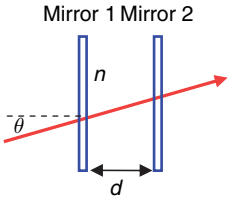


Figure 24 Schematic view of a Fabry–Perot interferometer.

where λ_0 is the vacuum wavelength, n is the refractive index of the medium between the two mirrors, d is the separation of the reflecting surfaces, θ is the incident angle, N is the order of interference, and $\delta_r/2$ is the phase change for one reflection, which is equal to 0 or π for dielectrics.

There are four variables in eqn [10], i.e., λ_0 , n, d, and θ. The transmission of a FPI can change with any of the variables. For acceleration sensing, d and θ can be used. However, $\cos\theta$ is approximately equal to 1 and is not sensitive to the change of θ when θ is small. Thus, for accelerometers, the change of the gap d is used to measure acceleration. There are two configurations for FPI accelerometers depending on the light source employed.

(1) The light source is a collimated monochromatic laser. In this case, the transmitted optical power is a measure of the gap change, which is proportional to the acceleration.
(2) The light source is broadband.

In this case, the shift of the transmission peak wavelength will be measured. This configuration is typically preferred as the measurement is independent of the absolute intensity reading. But cautions must be taken when designing an FPI accelerometer. The wavelength resolution is greatly affected by the reflectance of the mirrors, as shown in the following equation:

$$\Delta\lambda_{min} = N\pi\left(\frac{R^{1/2}}{1-R}\right)\lambda_0 \qquad [11]$$

where $R = (R_1R_2)^{1/2}$ with R_1 and R_2 being the reflectances of the two mirrors. For example, the resolving power is reduced by a factor of 2 when R is just slightly changed from 99% to 98%.

The FP cavity can be formed either vertically or horizontally. As shown in **Figure 25(a)**, a vertical FP cavity is formed between a vertical sidewall and the end face of an optical fiber. The vertical sidewall is

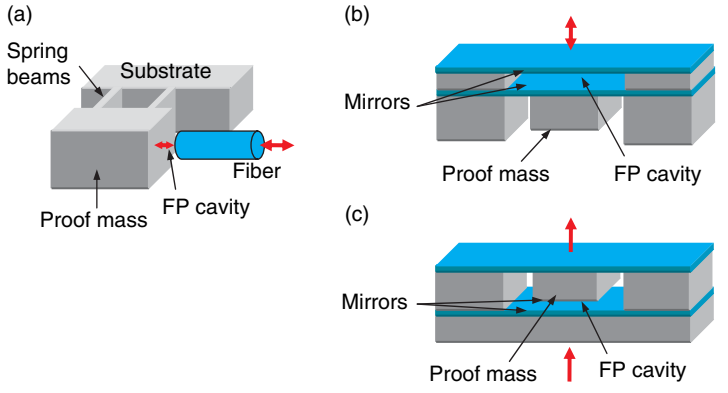


Figure 25 Schematic views of Fabry–Perot interferometer (FPI)-based accelerometers. (a) Lateral-axis sensing with a vertical FP cavity. (b) and (c) Vertical-axis sensing a horizontal cavity.

part of a proof mass. So the gap changes when there is an external acceleration. The optical fiber brings the light in and collects the reflected light back to a photodetector via a beam splitter (not shown). The wavelength shift of the reflection spectrum can be used to deduce the acceleration. Both (110) silicon substrate (Uenishi *et al.* 1994) and (100) silicon substrate (Schropfer *et al.* 1998) have been used to obtain vertical sidewalls. Smooth sidewall surfaces are always desired. Several approaches have been used to reduce the sidewall roughness such as deep etch shallow diffusion process (Juan and Pang 1998) and oxidation (Sparacin *et al.* 2005). Smooth sidewalls can also be achieved by optimizing the processing parameters of the Bosch ASE process (Donohue *et al.* 2004). Due to the low reflectance of the bare silicon sidewall (~30%), the resolution of this type of accelerometers is only on the order of 1 mg (Schropfer *et al.* 1998).

Horizontal cavities can be realized using wafer bonding as shown in **Figure 25(b)** and **25(c)**. If the proof mass is opaque to the incident light (**Figure 25(b)**), the reflected light from the FPI will be picked up by a photodetector (Baglio *et al.* 2004). If the proof mass is transparent to the source light (**Figure 25(c)**), the transmitted light will be collected to calculate the acceleration (Perez *et al.* 2005). The horizontal surfaces can be made extremely flat, and multilayer antireflection and high-reflection coatings can also be easily applied. Thus, the horizontal FPI accelerometers can achieve much higher resolutions (~1 micro-$g\text{Hz}^{-1/2}$) than their vertical FPI counterpart.

A Bragg grating is a transparent device with a periodically modulated refractive index, as shown in **Figure 26**. The wavelength that satisfies the

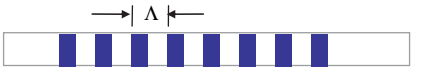


Figure 26 Schematic view of a Bragg grating.

Bragg condition is called the Bragg wavelength. The Bragg's law can be expressed as

$$\lambda_B = n\Lambda/2 \qquad [12]$$

where n is the refractive index of the grating material and Λ is the period of the grating. Only the Bragg wavelength is reflected and other wavelengths are transmitted. So a Bragg grating can be used as a filter or as a reflector. Note that a small index modulation is sufficient to achieve nearly total reflection for λ_B.

Bragg gratings can be made in a planar waveguide or in an optical fiber. Planar Bragg gratings are widely used in laser diodes and integrated optics while fiber Bragg gratings (FBGs) find extensive applications in optical communications. Bragg gratings can be created using femtosecond laser pulses or UV writing where the periodic pattern can be generated holographically or by a phase mask.

Package sizes of FBG-based accelerometers are relatively large due to the need of assembling circular optical fibers (Berkoff and Kersey 1996, Todd *et al.* 1998, Zhu *et al.* 2003). The demonstrated resolution is just around 1 micro-$g\text{Hz}^{-1/2}$, but there is a plenty of room to improve. To make optical sensors more compact, planar Bragg gratings are preferred. One such device is illustrated in **Figure 27**, where a Bragg grating is embedded in a planar waveguide (Storgaard-Larsen *et al.* 1996). A (110) silicon substrate and KOH etching can be used to obtain vertical thin beams. The Bragg grating is formed on

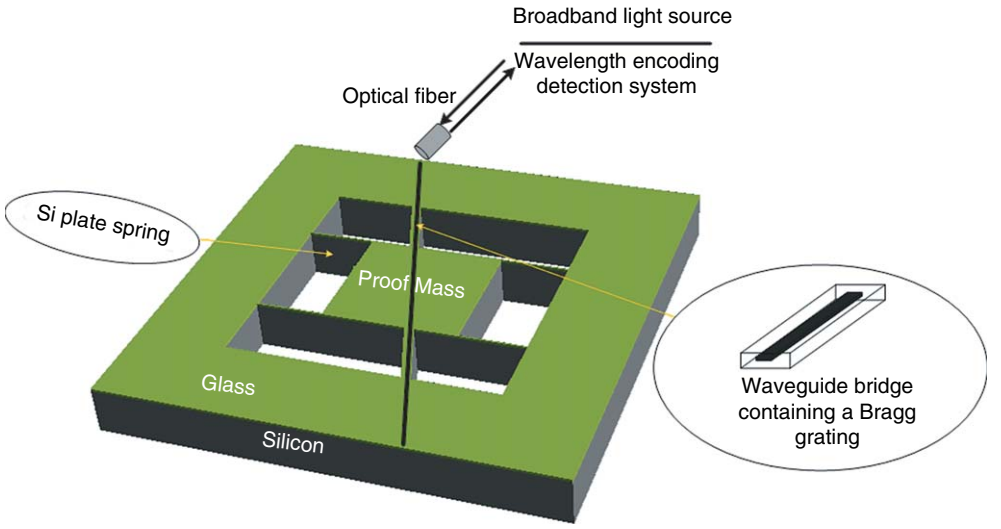


Figure 27 Schematic view of a planar Bragg grating-based lateral-axis accelerometer.

top of the thin beams using UV writing. When the proof mass moves due to external acceleration, the thin beams deform and in turn induce a stress in the Bragg grating. The strain of the Bragg grating then shifts the reflection spectrum, which is picked up by a photodetector.

There are also interferometric accelerometers based on diffraction gratings. Loh *et al.* (2002) demonstrated a high-resolution accelerometer with a bulk micromachined silicon proof mass and an interferometric position sensor. The interferometer consists of interdigitated fingers that are alternately attached to the proof mass and support substrate. Illuminating the fingers with coherent light generates a series of optical beams corresponding to different diffraction orders, as shown in **Figure 28(a)**. The intensity of a given beam depends on the out-of-plane separation between the proof mass fingers and support fingers. Proof masses with mechanical resonances ranging from 80 Hz to 1 kHz were fabricated with a two-mask process involving two DRIEs, an oxide etch stop, and a polyimide protective layer. The structures were packaged with a laser diode and photodiode into 8.6-cm^3 acrylic housings, as illustrated in **Figure 28(b)**. The 80-Hz resonant proof mass has a noise equivalent acceleration of 40 nano-$g\,Hz^{-1/2}$ and a dynamic range of 85 dB at 40 Hz.

Figure 29 shows another configuration for position sensing based on a diffraction grating, in which a sensing diaphragm rests above a rigid diffraction grating back electrode (Hall *et al.* 2006). In contrast, the design shown in **Figure 28** has all the grating fingers suspended. When illuminated from the backside with a coherent light source, such as a vertical

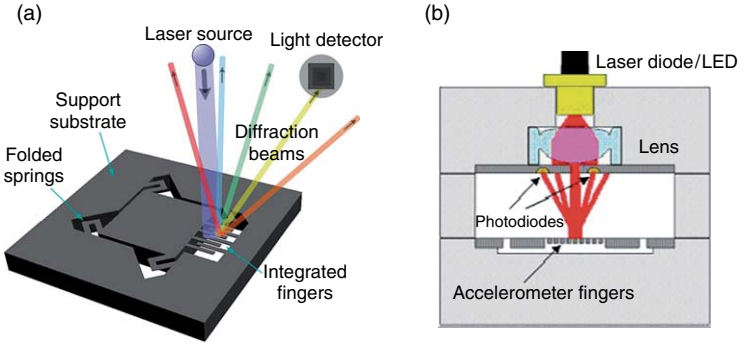


Figure 28 Diffraction grating-based *z*-axis accelerometer. (a) Schematic view. (b) Packaged device. LED, light-emitting diode. (Source: Loh N C, Schmidt M A, Manalis S R 2002 Sub-10 cm^3 Interferometric accelerometer with nano-*g* resolution. *J. Microelectromech. Syst.* **11**, 182–7; reproduced with permission from IEEE.)

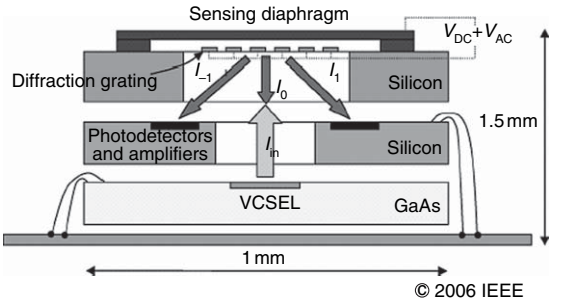


Figure 29 A position sensor with a fixed diffraction grating. VCSEL, vertical cavity surface-emitting laser. (Source: Hall N A, Okandan M, Degertekin F L 2006 Surface and bulk-silicon-micromachined optical displacement sensor fabricated with the SwIFT-Lite-8482™ Process. *J. Microelectromech. Syst.* **15**, 770–6; reproduced with permission from IEEE.)

cavity surface-emitting laser (VCSEL), as shown in Figure 29, a series of diffraction beams are generated. The angles of the diffraction orders remain fixed, and their intensities are modulated by the diaphragm deflection. The zero-order beam intensity is complementary to all higher-order beams, which enables a differential detection scheme for sensitivity doubling and laser intensity noise cancellation without any additional optical components. This design is robust, and the light source and photodiodes can be integrated. It is very suitable for microphones and pressure sensors, but additional steps or wafer bonding are needed to make the proof mass for accelerometers.

As can be seen from the various micromachined optical accelerometers discussed above, the sensitivity and resolution of optical accelerometers are normally not dependent on the gap size. This greatly eases the fabrication as compared to that for tunneling or capacitive accelerometers. However, optical accelerometers have much more stringent requirements in terms of surface flatness and roughness. In addition, optical accelerometers generally have large sizes mainly because many discrete optical components such as lenses, fibers, mirrors, beam splitters, laser diodes, and photodetectors are needed. With the progress of integrated optics, more and more optical components will be integratable and overall packaging size will be reduced.

In summary, optical accelerometers have unique features such as immunity to EMI, fire safety, low transmission loss, and multiplexing capability, so they are the better candidates for applications in harsh environments (high voltage or high temperature), underwater and underground. Although optical micromachined accelerometers are still in the early development stage, some demonstrated results such as a $40\,ng\,Hz^{-1/2}$ resolution from an interdigitated diffraction-based accelerometer are very promising. There is no doubt that we will continuously see increasing research activities in this area.

2.05.3.6 Tunneling Accelerometers

Scanning tunneling microscopy (STM) has been a powerful tool to study the atomic scale structure of surfaces since it was invented in 1981 by Binning and Rohrer. Electron tunneling occurs between two conductors separated by a sufficiently thin insulating layer or vacuum. The tunneling current I is a function of the gap $d(\text{Å})$ and has the following dependence:

$$I \propto V \exp\left(-\alpha\sqrt{\phi}d\right) \qquad [13]$$

where ϕ is the effective height of the tunneling barrier in electron-volt, V is the bias voltage in volts, and $\alpha = 2.05(\text{eV})^{-1/2}\text{Å}^{-1}$. This exponential relation makes the tunneling current extremely sensitive to even a tiny gap change. For example, with typical values of ϕ and d, the current varies by an order of magnitude for each angstrom change of the gap. This high position sensitivity creates an opportunity for making accelerometers with much higher resolutions than other types of accelerometers. A scanning tunneling probe is illustrated in Figure 30. At constant current mode, the feedback loop maintains the tunneling current constant by adjusting the gap between the tip and the sample surface. The tip height is typically adjusted by a piezoelectric actuator. Thus, the voltage applied to the actuator is a measure of the tip height. By scanning the tip over the surface and recording the actuator voltage, the surface structure of the sample can be reconstructed. STMs can reach sufficient resolution to show single atoms and are widely used in both industrial and fundamental research.

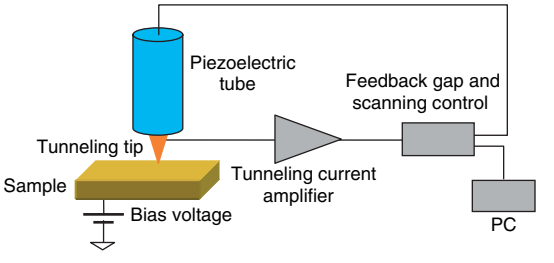


Figure 30 Schematic view of a scanning tunneling probe.

Electron tunneling was used in accelerometer designs in the late 1980s (Baski *et al.* 1988, Waltman and Kaiser 1989). In 1991, Kenny *et al.* reported the first micromachined tunneling accelerometer, and a gap change resolution of 2×10^{-4} Å $\mathrm{Hz}^{-1/2}$ and an acceleration resolution of 10 n$g\,\mathrm{Hz}^{-1/2}$ were measured at 1 kHz. Electrostatic actuation was employed in this design, eliminating the need for piezoelectric actuators. Compared with piezoelectric actuation, electrostatic actuation is much less sensitive to temperature drift and is immune to creep and hysteresis. The basic structure of a micromachined tunneling accelerometer consists of a proof mass, a tunneling tip, and an electrostatic actuator. The tunneling tip can be made either on the proof mass or on the substrate, as shown in **Figure 31(a)** and **31(b)**. According to eqn [3], the mechanical sensitivity is inversely proportional to the square of the resonance frequency of the spring–mass system. Therefore the early tunneling accelerometers have bandwidths up to only about 100 Hz. In order to increase the bandwidth, a cantilever is placed above the tunneling tip as shown in **Figure 31(c)** (Kenny *et al.* 1994). The resonance frequency of the cantilever is much higher than that of the proof mass. So, the cantilever can follow closely the motion of the proof mass at much higher frequencies than the proof mass resonance. A bandwidth of 1 kHz was reported by using this approach (Rockstad *et al.* 1996). Note that the responsivity of the device decreases rapidly after 100 Hz (the resonance frequency of the proof mass). The soft, lightweight cantilever can also reduce the crush failure of the tip. Another way to protect the tip is to use a pull-back electrode as shown in **Figure 31(d)** and **31(e)** (Yeh and Najafi 1995, Zavracky *et al.* 1996).

Tunneling accelerometers have been demonstrated using surface, bulk, and mixed micromachining processes. The microfabrication challenges include the tip fabrication and the tip gap control during wafer bonding. Numerous tip fabrication methods have been developed and some of them are illustrated in **Figure 32**. The method in **Figure 32(a)** has been widely used (Dong *et al.* 2005, Kenny *et al.* 1991, Yeh and Najafi 1997, Zavracky *et al.* 1996). In this method, the tip is fabricated by etching silicon anisotropically (e.g., EDP or KOH) or isotropically (e.g., hydrofluoric acid–nitric acid–acetic acid (HNA)) with a square oxide mask. The silicon undercutting forms a sharp silicon tip. After that the wafer is passivated (e.g., plasma-enhanced chemical vapor deposition (PECVD) oxide). Then a Ti (or Cr)/Pt/Au layer is coated on the tip by either lift-off or sputtering with a shadow mask. The method in **Figure 32(b)** follows the same steps as those in **Figure 32(a)** and then uses one backside and one front-side anisotropic wet tech (Liu *et al.* 1998). The tip is formed on a thin-film cantilever, which prevents the tip from crash damage. The method in **Figure 32(c)** employs focused ion beam (FIB) lithography to obtain small patterns (<100 nm). Followed by an ion milling step, a Ti/Pt/Au tip is formed. Using this process, a surface micromachined tunneling accelerometer was demonstrated with a noise floor of 85 µ$g\,\mathrm{Hz}^{-1/2}$ (Kubena *et al.* 1996). The method in **Figure 32(d)** uses hot embossing to create a polymethylmethacrylate (PMMA) tip from a silicon pyramid etched by KOH (Cui and Wang 2005). The last method is quite different from other methods. In this method (see **Figure 32(e)**), the tip moves laterally, while all other methods use vertically moving tips. The lateral tip can be fabricated using the single-crystal reactive etching and metallization (SCREAM) process (Shaw *et al.* 1993).

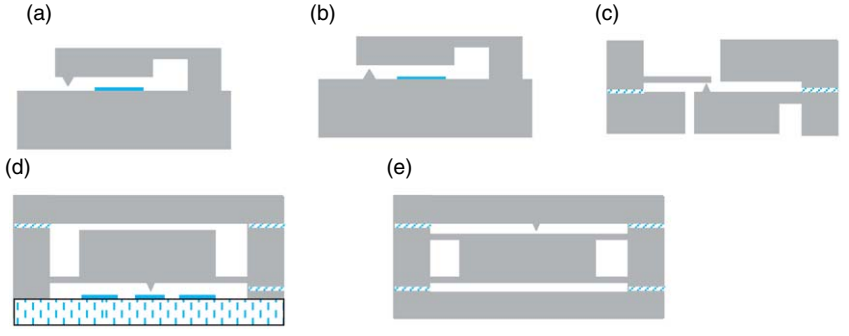


Figure 31 Schematic cross-sectional views of micromachined tunneling accelerometers. (a) Tunneling tip is on the proof mass. (b) Tunneling tip is on the substrate. (c) Wide-bandwidth design. (d) and (e) Dual-plate designs with tip protection and overrange protection.

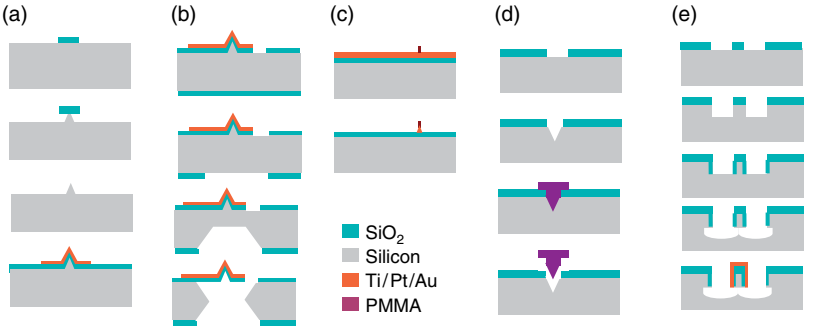


Figure 32 Schematic views of various tip fabrication methods. (a) A tip fabricated on a proof mass or substrate using a square oxide mask with anisotropic or isotropic wet etching. (b) A tip on a thin-film cantilever. (c) A small tip by focused ion beam (FIB) lithography plus ion milling. (d) A polymer tip by hot embossing. (e) A lateral tip fabricated by the SCREAM process.

A lateral tunneling accelerometer with a resolution of $20\,\mu g\mathrm{Hz}^{-1/2}$ was demonstrated using this process (Hartwell *et al.* 1998).

2.05.3.7 Resonant Accelerometers

Resonant accelerometers generate direct acceleration-to-frequency outputs. This provides not only the immunity to any amplitude measurement errors but also quasi-digital outputs to simplify the interfacing with digital systems. With the high quality factor, resonant accelerometers have high resolution and wide dynamic range. Quartz, single crystal silicon (SCS), and polysilicon are the common materials for resonators. Quartz resonant accelerometers have good performance and are commercially available especially for inertial-grade applications (Albert 1994, Kourepenis *et al.* 1991, Le Traon *et al.* 1998, Norling 1988). The high performance of quartz accelerometers is attributed to its high quality factor and inherently piezoelectric property for both actuation and sensing. SCS also has excellent mechanical properties and is the major substrate material for IC products. Various SCS resonant accelerometers have been demonstrated (Aikele *et al.* 2001, Burrer and Esteve 1995, Ferrari *et al.* 2005, Helsel *et al.* 1994, Ohlckers *et al.* 1998, Roszhart *et al.* 1995). Although polysilicon resonators show an inferior quality, the process maturity, the extensive use in IC fabrication, and the ease of IC integration make them very competitive. This is similar to the case for capacitive accelerometers. Some early micromachined resonant accelerometers are based on polysilicon (Burns *et al.* 1996, Chang *et al.* 1990, Guckel *et al.* 1992, Roessig *et al.* 1997, Seshia *et al.* 2002). Polysilicon resonators are not only used for accelerometers but also for radio frequency (RF) filters, clock generators, and biochemical sensors.

A resonant accelerometer is a device that has an element vibrating at its resonance. The resonance frequency shifts in the presence of an external acceleration. The vibrating element can be a beam, bridge, or tuning fork (Stemme 1991). The resonance frequency can be varied by a change in stress, mass, temperature, etc. For accelerometers, the most commonly used effect is the stress-induced stiffness change, which in turn changes the resonance frequency of the resonator. To excite the resonance, piezoelectric, electrostatic, and electrothermal actuations are often used. The vibration can be sensed piezoelectrically, capacitively, or piezoresistively. So there are several excitation and sensing combinations. The most commonly used pairs include piezoelectric/piezoelectric, electrostatic/capacitive, electrothermal/piezoresistive, and electrostatic/piezoresistive.

High mechanical quality factor (Q-factor) is needed for high sensitivity and high resolution. There are a few damping mechanisms, which can be expressed as a specific Q-factor for each: viscous and acoustic damping, Q_a, damping due to imbalances, Q_s, and damping resulting from internal material-related losses, Q_i. The overall Q is given by

$$\frac{1}{Q}=\frac{1}{Q_a}+\frac{1}{Q_s}+\frac{1}{Q_i} \qquad [14]$$

The first term ($1/Q_a$) dominates for any resonators in air. Air damping is negligible at a sufficiently low pressure (~0.1 mtorr). So vacuum packaging is needed for micromachined resonant accelerometers. The second term ($1/Q_s$) accounts for the

energy loss to the anchor. This energy loss can be minimized by mechanically isolating the resonator from its support(s). Thus a resonator design that minimizes the forces and torques acting on the support(s) should be used. A well-known example is the double-ended tuning fork (DETF) (Helsel *et al.* 1994). As shown in **Figure 33**, the beams oscillate 180° out of phase to cancel reaction forces at the ends of the beams. Thus, the dynamically balanced DETF resonator has a high Q. The third term $(1/Q_i)$ concerns the internal energy losses and it is dependent on the purity, dislocations, and thermoelastic losses of the resonator material. The thermoelastic losses are due to irreversible heat conduction resulting from the flexural motion. Polycrystalline materials have lower Q than single-crystal materials. If an SCS or quartz is used, the third term can be neglected (Stemme 1991).

There are mainly two types of micromachined resonant accelerometers in terms of the vibrating element design: DETF and doubly sustained beam (DSB). The basic structure of DETF is already shown in **Figure 33**. Using this DETF design and a bulk micromachining process, Helsel *et al.* (1994) reported an SCS-based resonant accelerometer with a resolution of 50 μg. In order to increase the scale factor, Roessig *et al.* (1997) proposed a novel lever structure. As shown in **Figure 34(a)**, the leverage mechanism can easily provide a force amplification of over an order of magnitude. Using the topology sketched in **Figure 34(b)**, a surface micromachined resonant accelerometer with a 2-μm-thick polysilicon structural layer demonstrated a scale factor of 2.4 Hz g^{-1} and a resolution of 1.5 mg $Hz^{-1/2}$ (Roessig *et al.* 1997). Using the same process and DETF design but a different topology as shown in **Figure 34(c)**, Seshia *et al.* (2002) improved the scale factor to 17 Hz g^{-1} and the resolution to 40 $\mu g\,Hz^{-1/2}$ with a working frequency range up to 1 kHz. Jia *et al.* (2004) implemented a similar design using an SOI substrate and achieved a scale factor of 27 Hz g^{-1} and a resolution of 160 $\mu g\,Hz^{-1/2}$.

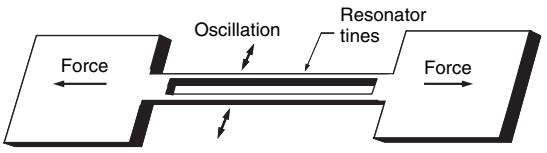


Figure 33 Schematic view of a double-ended tuning fork (DETF) resonator. The beams vibrate 180° out of phase to provide dynamic moment cancellation.

Recently, Su *et al.* (2005) reported a two-stage leverage mechanism to further increase the scale factor. A resonant accelerometer design with a two-stage leverage mechanism and a pair of DETF is illustrated in **Figure 35(a)**. The two-stage leverage was measured to have a force amplification factor of 80, and the overall scale factor was improved to 153 Hz g^{-1}.

There are a few variations for the DSB design. Three of them are illustrated in **Figure 36**. In all cases, there are a proof mass and a suspension flexure. The vibrating element is a resonant microbeam with one end anchored on the substrate and the other end attached to the proof mass or the flexure. The resonant frequency of the microbeam is much higher than that of the proof mass. The microbeam resonance is in the range of 100 kHz to 1 MHz, while the proof mass resonance is on the order of 100 Hz. An external acceleration deflects the flexure through the proof mass. The flexure deflection then stretches or compresses the microbeam, resulting in a respective decrease or increase of the resonance frequency of the microbeam. This frequency shift is a measure of the external acceleration.

In the case in **Figure 36(a)**, the doubly sustained microbeam is formed on top of the flexure with a small gap. To maximize the sensitivity, the microbeam is placed close to the substrate where the stress is maximal. Burns *et al.* (1996) developed a combined bulk and surface micromachining process to demonstrate this type of resonant accelerometer. The microbeam was fabricated from thin films of mechanical-grade polysilicon with integral polysilicon vacuum encapsulation. The microbeam was electrostatically excited and the vibration was detected piezoresistively. Resonant accelerometers with scale factors greater than 700 Hz g^{-1} and a base microbeam frequency of 524 kHz were fabricated. The Q was about 20,000.

In the case in **Figure 36(b)**, the resonant microbeam is clamped to the substrate on one end and connected to the proof mass on the other end. The proof mass is supported by some thin hinge beams. Note that the microbeam is typically thinner and longer than the hinges. Roszhart *et al.* (1995) demonstrated a 2-μg resolution resonant accelerometer using a bulk micromachining process plus wafer bonding. A scanning electron micrograph (SEM) of a fabricated device chip is shown in **Figure 37(a)**. The Q was higher than 60,000. Burrer

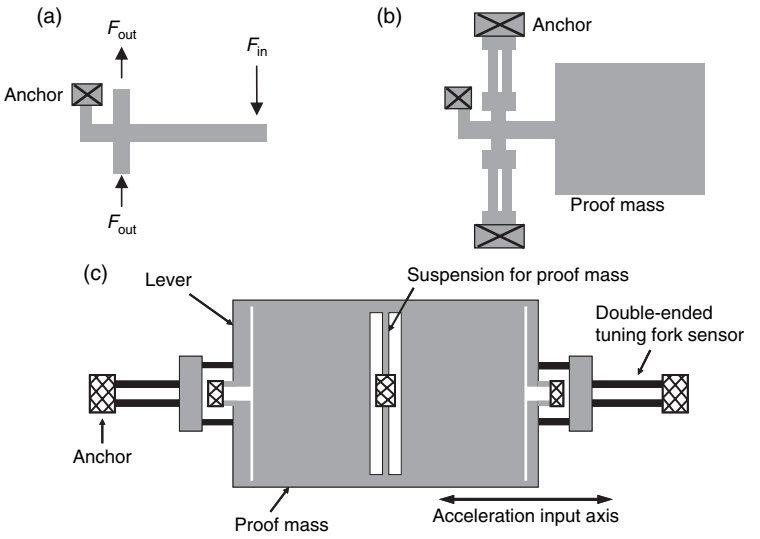


Figure 34 (a) Basic leverage mechanism. (b) A design with a proof mass on one side. (c) A symmetric design.

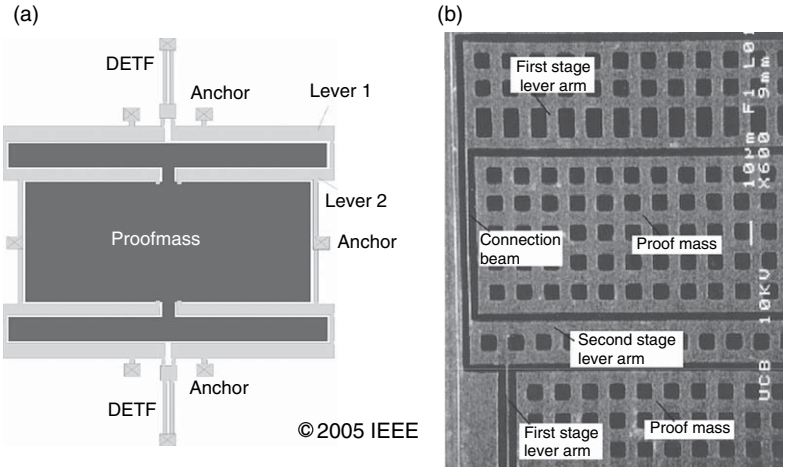


Figure 35 (a) Schematic of a resonant accelerometer with a two-stage leverage mechanism. (b) A scanning electron micrograph (SEM) of the two-stage microleverage mechanism. DETF, double-ended tuning fork. (Source: Su S X P, Yang H S, Agogino A M 2005 A resonant accelerometer with two-stage microleverage mechanisms fabricated by SOI-MEMS technology. *IEEE Sensors J.* **5**, 1214–23; reproduced with permission from IEEE.)

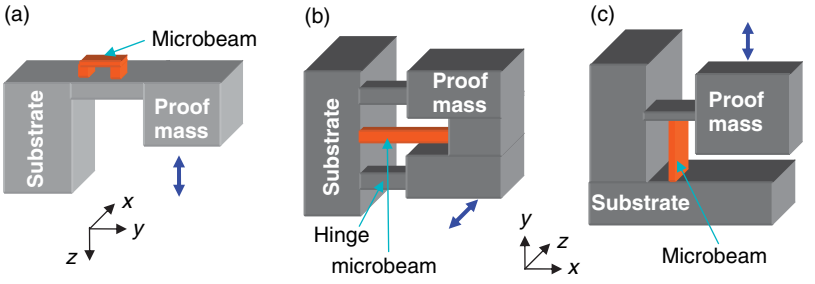


Figure 36 (a) Resonant microbeam on top of the hinge. (b) Resonant microbeam in between hinges. Note that the microbeam is thinner and longer than the hinge beams. (c) Resonant microbeam for lateral strain.

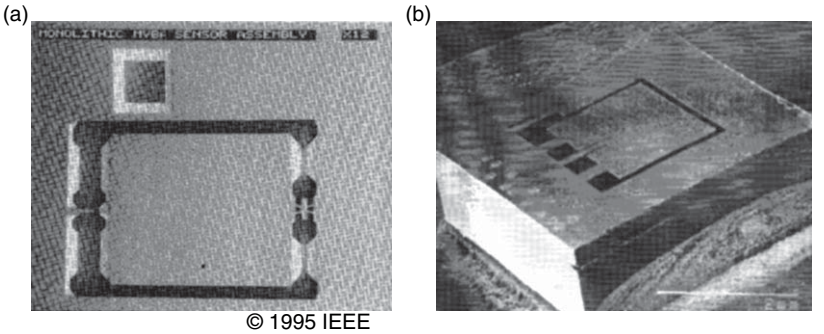


Figure 37 (a) Scanning electron micrograph (SEM) of a chip from Roszhart's design (b) SEM of a chip from Burrer's design.

and coworkers also made similar devices (Burrer and Esteve 1995, Burrer *et al.* 1996). **Figure 37(b)** shows a SEM of a fabricated device. Electrothermal drive and piezoresistive sense were used. The measured scale factor was 260 Hz g^{-1}. It is also interesting to note that Ohlckers *et al.* (1997) demonstrated a vibrating beam accelerometer without using a microbeam. Instead, a high-order mode (~400 kHz) of the cantilever proof mass assembly was used. The Q of that mode is around 1000 at atmospheric pressure. So no vacuum packaging is needed. The measured scale factor is 6–7 Hz g^{-1}.

In the case in **Figure 36(c)**, the microbeam is anchored to the substrate on one end and perpendicularly connected to the suspension beam of the proof mass. Both the suspension beam and the microbeam are narrow and thick, so that the lateral strain or acceleration can be detected. **Figure 38** shows such an accelerometer developed by Aikele *et al.* (2001) using a bulk micromachining process. The thickness of the sensor structure is 30 μm. The microbeam is thermally excited to a lateral vibration by an implanted resistor. The vibration is sensed by an implanted piezoresistor. Both resistors are located on a U-shaped structure. The resonant frequency of the microbeam is 400 kHz, which can be easily excited by the U-shaped structure. The Q is about 60,000 at 0.01 mbar. The measured scale factor is 70 Hz g^{-1}. The resonant frequency of the proof mass is also as high as 17 kHz, making the sensor insensitive to any vehicle vibrations.

All the resonant accelerometers discussed so far are based on stress-induced resonance shift. There also exist other effects, which can be used to detect acceleration by a resonator. For example, Lee *et al.* (2000) proposed to use a gap-sensitive electrostatic stiffness changing effect to detect external acceleration. In this case, electrostatic actuation is used to excite the resonance. The idea is based on the fact that the effective stiffness of an electrostatic actuator changes with a varying gap. An imbalance proof mass with a torsional spring was used to create a differential output in a differential z-axis accelerometer (Seok *et al.* 2002). The same group extended the design to lateral-axis accelerometers and achieved a 5.2-μg resolution and 128 Hz g^{-1} scale factor (Seok and Chun 2006).

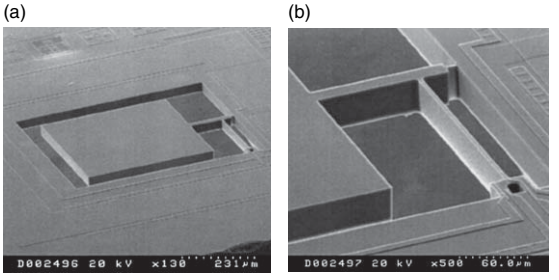


Figure 38 (a) Scanning electron micrograph (SEM) of a sensor chip from Aikele's design. (b) SEM showing the microbeam, the hinge and the U-structure for excitation and detection.

2.05.4 Accelerometer Processing

There are a large number of different micromachining processes. Here we only select a few representative processes to address fabrication issues such as temperature, stress, size, cost, stiction, circuit integration, and packaging. Because of their dominance in the commercial products market, piezoresistive and capacitive accelerometers have been given more attention.

2.05.4.1 First Bulk Micromachined Piezoresistive Accelerometer

The first micromachined accelerometer was developed at Stanford University in 1978 (Roylance and Angell 1978). It was a bulk silicon micromachining process plus silicon-to-glass bonding. The process flow is shown in **Figure 39**. It starts with n-type (100) silicon. The first step was to etch some holes completely through the wafer. These through-holes were used as two-side alignment markers. Next, thermal oxide and p^+ diffusion were used to form p-type piezoresistors (**Figure 39(a)**). After that, a KOH etch was performed on the backside (**Figure 39(b)**) and stopped at twice the desired silicon thickness. Next, the frontside oxide on the trench region was removed, followed by another KOH etch (**Figure 39(c)**). This second KOH etch was stopped when the trench region was etched through, yielding the desired thickness for the piezoresistive sensing region (**Figure 39(d)**). Next, the remaining oxide was stripped (**Figure 39(e)**). Finally, the silicon wafer was anodically bonded to a top and bottom glass plate for hermetic sealing (**Figure 39(f)**). Both plates had wells for microstructure clearance. The top glass plate also had sputtered aluminum for contacts. A SEM of a fabricated device before glass bonding is shown in **Figure 39(g)**. This device was capable of measuring accelerations from 1 mg to 50 *g* over a 100-Hz bandwidth.

2.05.4.2 Motorola's Surface Micromachined *z*-Axis Capacitive Accelerometer (G-Cell)

The basic idea of this process is to create a fully differential capacitive output by sandwiching a movable plate in between a fixed top and bottom plate. This process was developed in the early 1990s (Ristic *et al.* 1992, Shemansky *et al.* 1995), and surface micromachining was used. The process flow is shown in **Figure 40**. It starts with a 2.4-μm-thick thermal oxide for isolation and a thin silicon nitride layer for protection during the sacrificial etching step (**Figure 40(a)**). Then the first polysilicon layer (poly-1) is deposited and patterned, followed by the deposition and patterning of the first polysilicate glass (PSG) spacer layer (PSG-1) (**Figure 40(b)**). After that, the second polysilicon layer (poly-2) is deposited and patterned to define the proof mass (**Figure 40(c)**). Then a second PSG spacer layer (PSG-2) is deposited and patterned (**Figure 40(d)**), followed by the deposition and patterning of a third polysilicon layer (poly-3) (**Figure 40(e)**). Finally the PSG layers are etched to release the microstructure (**Figure 40(f)**). A released accelerometer is shown in **Figure 40(g)**, where the middle electrode (poly-2) is suspended via four long beams. The top electrode (poly-3) is anchored rigidly to the substrate, and the bottom electrode (poly-1) is directly deposited on the substrate and isolated by a thick oxide layer. Thus, a

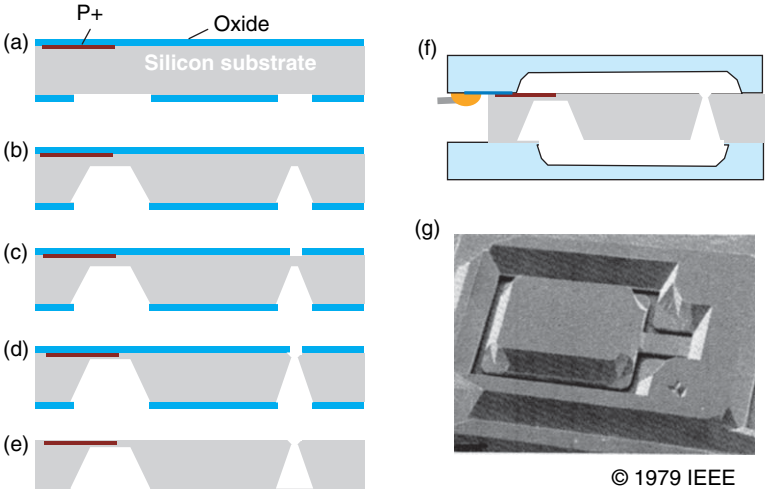


Figure 39 (a)–(f) Cross-sectional process flow of the first micromachined accelerometer. (g) Scanning electron micrograph (SEM) of a fabricated device shown from the backside. (Source: Roylance L M, Angell J B 1979 A batch-fabricated silicon accelerometer. *IEEE Trans. Electron Devices* **26**, 1911–17; reproduced with permission from IEEE.)

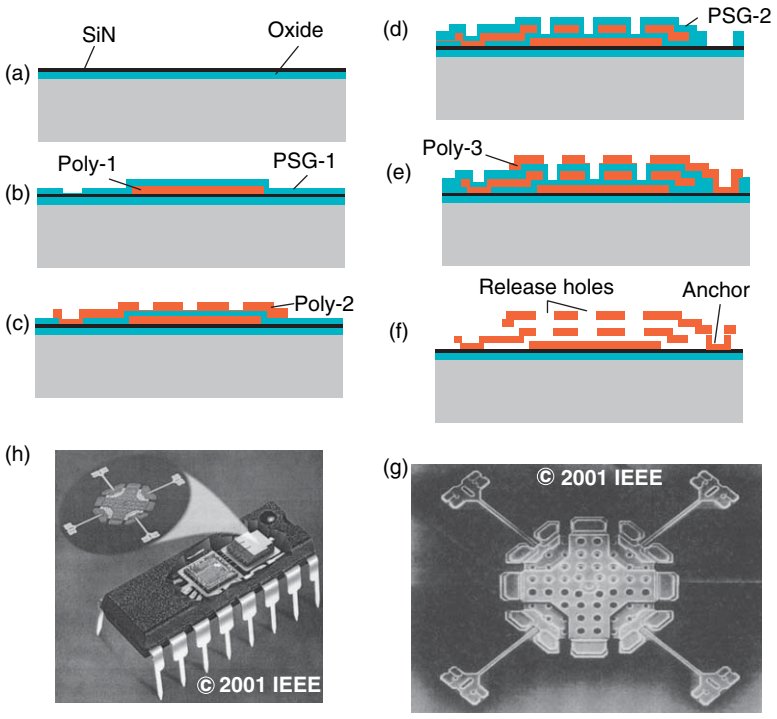


Figure 40 (a)–(f) Cross-sectional process flow of Motorola's G-cell accelerometer. (g) Scanning electron micrograph (SEM) of a released device. (h) Two-chip packaging. (Source: Li G, Tseng A A 2001 Low stress packaging of a micromachined accelerometer. *IEEE Trans. Electron. Packaging Manuf.* **24**, 18–25; reproduced with permission from IEEE.)

z-axis accelerometer with differential capacitive sensing is obtained. This configuration has several advantages including low cross-axis sensitivity, overrange protection, and self-test capability. A two-chip package (a sensor chip and an application-specific integrated circuit (ASIC) chip are wire bonded in a dual inline package (DIP)) is shown in **Figure 40(h)** (Li and Tseng 2001). These accelerometers are targeted at airbag applications. They have a 50-g full-scale output and can withstand over 5000-g impact.

2.05.4.3 ADI's iMEMS Process

ADI, teaming up with UC-Berkeley, started to develop integrated MEMS process in the late 1980s and brought up its first production of accelerometers in 1993 (Lewis *et al.* 2003, Payne *et al.* 1995, Sherman *et al.* 1992). The process flow is shown in **Figure 41**. This is an intermediate CMOS micromachining process, i.e., the MEMS structural layer is formed in between the CMOS process steps. In this process, standard IC process steps are performed until a gate is formed (**Figure 41(a)**). The remaining steps for circuits are just passivation and metallization. Polysilicon and oxide, respectively, are used as the structural and sacrificial layer (**Figure 41(b)**). After oxide deposition, low-pressure chemical vapor deposition (LPCVD) polysilicon is deposited at about 650 °C. In order to minimize residual stresses, a high-temperature annealing is followed. Due to the high deposition and annealing temperature, the aluminum layer must be deposited after the structural polysilicon layer is formed (**Figure 41(c)**). Then the metal layer and the circuit area are covered by an oxide/Si_3N_4 layer. For the final release step, instead of using vapor HF or critical point drying, a very inexpensive buffered HF etch is employed (**Figure 41(d)**). To prevent stiction, photoresist ties are patterned to hold the MEMS structures in place during buffered HF etching. After that, photoresist is stripped in oxygen plasma. Single-axis, dual-axis, and three-axis accelerometers with resolutions on the order of 100 $\mu g\,Hz^{-1/2}$ are commercially available.

2.05.4.4 Bosch Epipoly Capacitive Accelerometer

Thicker structural layers yields larger sensing capacitance and smaller thermomechanical noise with a given size. Bosch developed a thick structural layer process based on epitaxial polysilicon and used it for

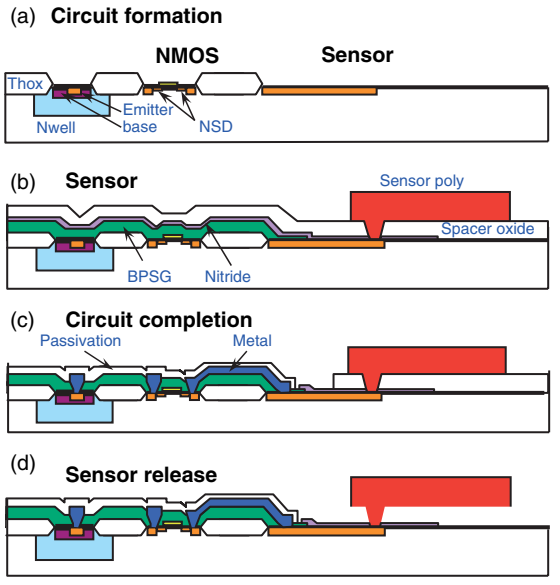


Figure 41 Cross-sectional view of Analog Devices Inc. (ADI's) iMEMS process (NMOS: N-type MOSFET; NSD; N-type source drain; and BPSG: borophosphosilicate glass).

fabricating inertial sensors (Lutz *et al.* 1997, Offenberg *et al.* 1995, Reichenbach *et al.* 2003). Several processes have been developed by Bosch. The process presented here is modified to enable three-axis sensing by using an interconnect polysilicon layer to form *z*-axis displacement sensing (Reichenbach *et al.* 2003).

The cross-sectional view of the process is shown in **Figure 42**. First, a 2.5-μm-thick passivation oxide is thermally grown on a silicon substrate, followed by a 0.45-μm-thick LPCVD polysilicon (**Figure 42(a)**). This thin polysilicon layer serves as an interconnection layer and the bottom electrode of *z*-axis sensing capacitors as well. This approach also provides better isolation and smaller parasitic capacitance compared to the diffusion wiring and p–n junction isolation in an early process (Offenberg *et al.* 1995). Then a 1.6-μm-thick sacrificial oxide layer is deposited and patterned (**Figure 42(b)**). The most critical step of the process is the epitaxial growth of an 11-μm-thick polysilicon layer (epipoly). After that, the epipoly is planarized, doped, and annealed to minimize residual stresses (**Figure 42(c)**). Next, aluminum bond pads are formed and covered by patterned photoresist, and then DRIE is performed to anisotropically etch the epipoly (**Figure 42(d)**). Finally, a vapor HF technique is used to remove the sacrificial oxide and release the microstructure (**Figure 42(e)**). A release structure is shown in **Figure 42(f)**. After the fabrication, each sensor chip is encapsulated by a silicon cap and wire bonded to an ASIC chip in a package, as shown in **Figure 42(g)**. This process can be used to make low-gee, airbag, and three-axis accelerometers. The resolution of the low-gee accelerometers is on the order of 1 mg.

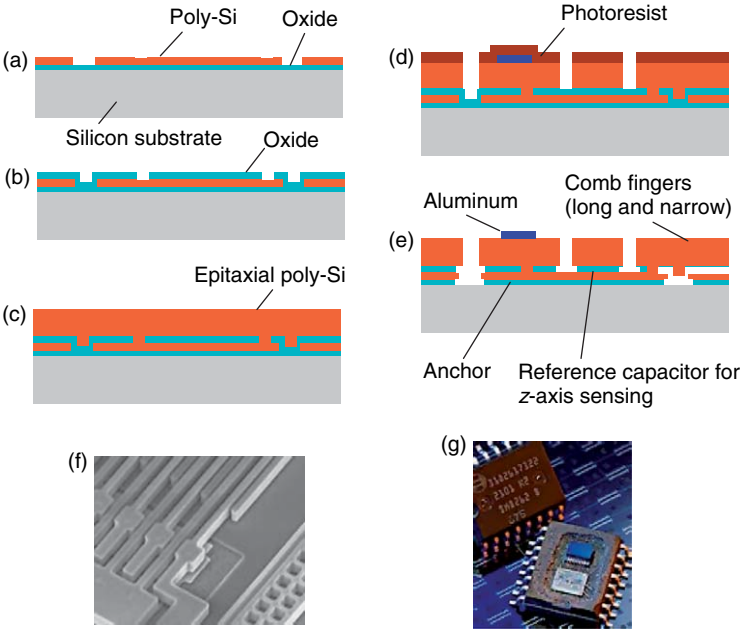


Figure 42 (a)–(e) Cross-sectional process flow of Bosch capacitive accelerometers. (f) Scanning electron micrograph (SEM) of a released microstructure. (g) Two-chip packaging.

2.05.4.5 ST's Capacitive Accelerometer

Similar to Bosch's approach, ST also uses low-stress, large-thickness epitaxially grown polysilicon as structural layers. The process developed by ST is called THELMA (thick epipoly layer for microactuators and accelerometers). ST has been developing MEMS accelerometers since 1996. The process flow is shown in **Figure 43(a)–(e)**. Deep silicon etching is used to obtain high aspect ratio vertical sidewalls. The stiction problem is overcome by using vapor HF dry etching. The sensor microstructure is capped by using wafer-to-wafer bonding. A SEM of a sensor structure and a packaged accelerometer are shown in **Figure 43(f)** and **43(h)**, respectively.

2.05.4.6 ADI's SOI iMEMS Process

This newly developed iMEMS process based on an SOI substrate is CMOS compatible and provides trench isolation and robust SCS microstructures (Chen *et al.* 2005). The top silicon layer (about 10 μm thick) is used as the MEMS structural layer, and the BOX layer (about 1 μm thick) is used as the sacrificial layer. The process flow is shown in **Figure 44**. It consists of three stages:

(1) Pre-CMOS-MEMS definition. In this stage, electrically isolated silicon islands are formed using isolation trenches. The isolation trenches are obtained by DRIE trench etch, silicon nitride passivation of trench sidewalls, and trench refilling by polysilicon; and the wafer is planarized using chemical–mechanical planarization (CMP) (**Figure 44(a)**). An oxide layer protects the trenches and MEMS sensor area during CMOS processing.

(2) Standard CMOS fabrication (**Figure 44(b)**). The planarized wafer can be directly sent to a CMOS foundry. The only extra effort is to align the first CMOS photomask to the prefabricated markers on the wafer. Because no high-temperature processing is necessary for MEMS after stage (1), metallization is much

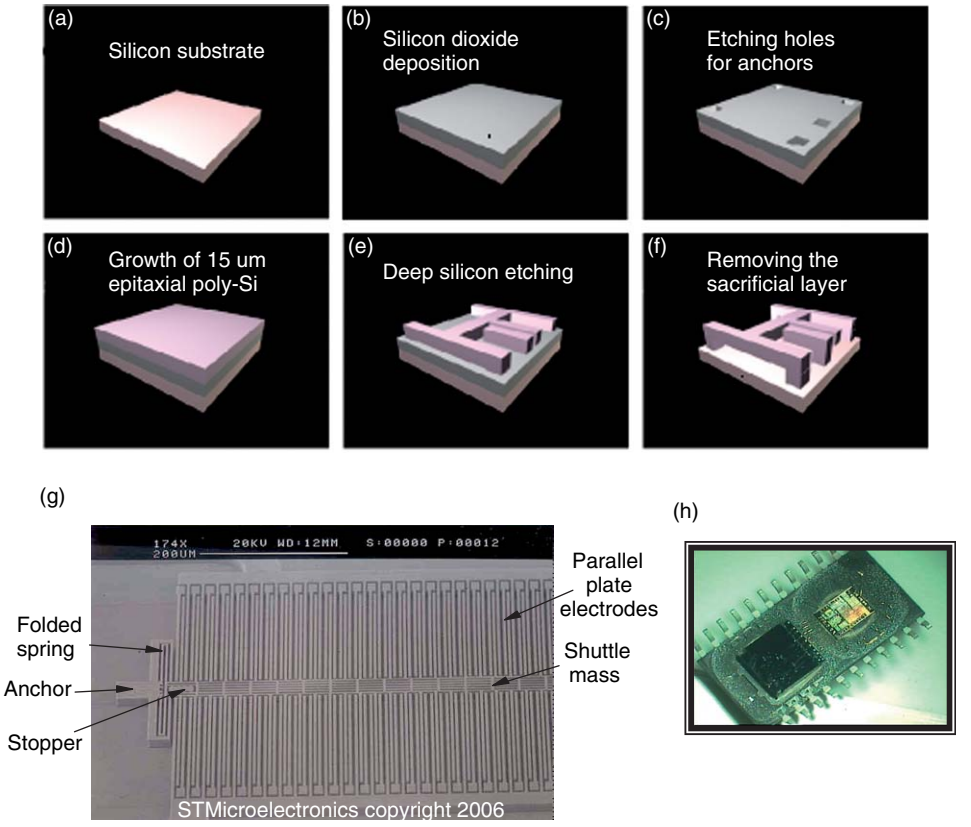


Figure 43 (a)–(f) Cross-sectional process flow of ST Microelectronics (ST) capacitive accelerometers. (g) Scanning electron micrograph (SEM) of a released microstructure. (h) Two-chip packaging. CMOS, complementary metal oxide semiconductor; DRIE, deep reactive ion etch; BOX, buried oxide.

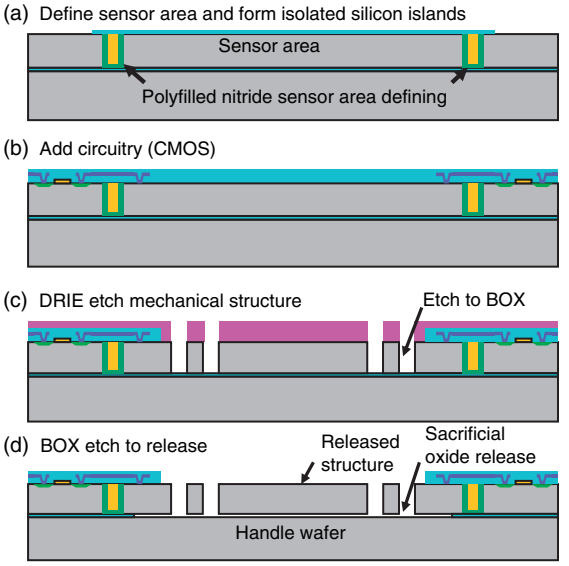


Figure 44 Cross-sectional view of Analog Devices Inc. (ADI's) silicon on insulator (SOI) iMEMS process.

more flexible and multiple interconnect metal layers can be used to decrease the circuit area. The CMOS metal layers also interconnect the silicon islands to form capacitive displacement sensors.

(3) Post-CMOS-MEMS release. There are two steps involved in this stage. First, silicon DRIE is used to form narrow-gap, interdigitated comb fingers, and define the suspension flexure and the proof mass (**Figure 44(c)**). Photoresist can be used as the etching mask. For the final release step, instead of using vapor HF or critical point drying, a very inexpensive buffered HF etch is employed. To prevent stiction, photoresist ties are patterned to hold the MEMS structures in place during buffered HF etching. The photoresist is later stripped in oxygen plasma to release the moving structure (**Figure 44(d)**). The working of a high-*g* accelerometer has been demonstrated with this process (Davis *et al.* 2004).

2.05.4.7 SensoNor's Resonant Piezoresistive Accelerometer – SA30

SensoNor's crash sensor SA30 is a resonant accelerometer with electrothermal excitation and piezoresistive sensing (Jakobsen 1996, Kvisteroy and Jakobsen 1998, Ohlckers *et al.* 1998). A die photograph of the sensor is shown in **Figure 45(g)**, where a center proof mass is suspended in two flexible beams to form a resonator structure. The resonator shifts its frequency corresponding to the external acceleration.

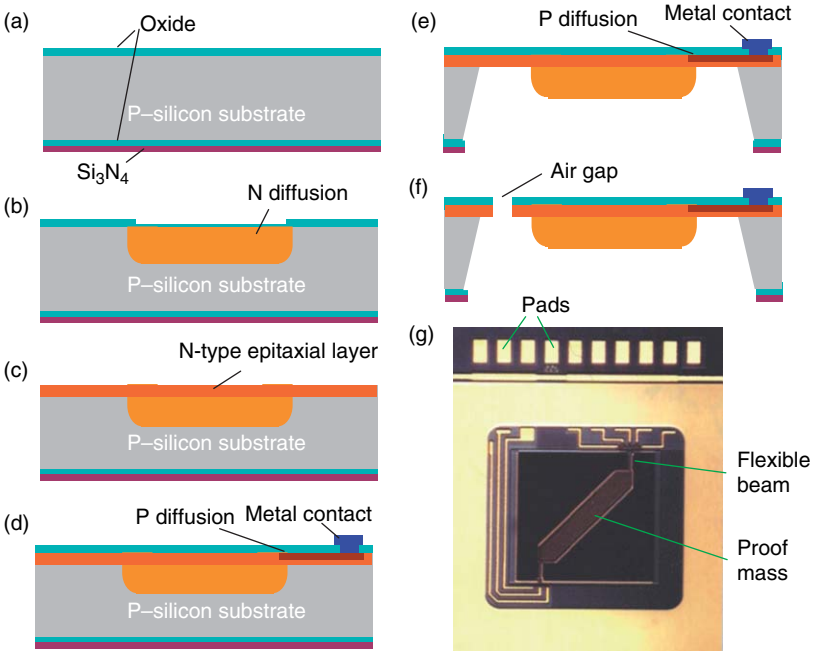


Figure 45 (a)–(f) Cross-sectional process flow of SensoNor's SA30 crash sensor. (g) Photograph of the sensor chip of a SA30 sensor.

The thickness of the proof mass is 20 μm and the thickness of the flexible beams is 3 μm. A basic fabrication process flow is shown in **Figure 45(a)–(f)**. The actual process might be different. The process starts with a double-side polished p-type (100) silicon substrate. First a ~1-μm-thick oxide is thermally grown on both sides of the wafer, followed by a 0.3-μm-thick LPCVD Si_3N_4 (**Figure 45(a)**). Then, phosphorus ions are implanted and driven 17 μm deep into the substrate (**Figure 45(b)**). Next, a 3-μm-thick n-type epitaxial layer is grown on the frontside (**Figure 45(c)**). After that, boron ions are implanted to form thermal resistors and p-type piezoresistors, followed by the formation of metal contacts to the diffusion resistors (**Figure 45(d)**). Then the process is turned to the backside. An aqueous KOH etch is applied and the Si_3N_4 layer is the etching mask. An electrochemical technique can be used to automatically stop the etching at the p–n junction. Because there are different p–n junction depths in the wafer, different silicon thicknesses are obtained (**Figure 45(e)**). Finally, the flexible beams and proof mass are released by RIE etching (**Figure 45(f)**).

Anodic bonding is used to hermetically seal the silicon die between two glass plates. The structure thickness can be controlled within 1%. A Q-factor higher than 2000 is obtained at atmospheric pressure. The resonance frequency for the micromachined device is approximately 700 kHz. The sensitivity is in the range 60–70 Hz g^{-1} (Ohlckers *et al.* 1998).

2.05.4.8 MEMSIC Thermal Accelerometer – Surface Micromachining

The fabrication process is foundry CMOS compatible (Zhao and Hua 2004). The process flow is shown in **Figure 46**. It starts with a CMOS wafer. The first step is to define the MEMS structure and protect on-chip CMOS circuits. The mask layer can be photoresist, polyimide, or metals (**Figure 46(a)**). Then an anisotropic oxide etch is performed to expose the silicon substrate (**Figure 46(b)**), followed by a DRIE silicon etch (**Figure 46(c)**). Finally an isotropic silicon etch is used to undercut the silicon beneath the microstructure (**Figure 46(d)**). A microstructure created by this process is shown in **Figure 46(e)**, and a SEM of a release sensor die in **Figure 46(f)**. This process involves very low cost due to its standard CMOS compatibility and simple process steps. The resultant accelerometers are robust because they have no moving parts. The resolution and bandwidth of these accelerometers are about 1 mg and 30 Hz.

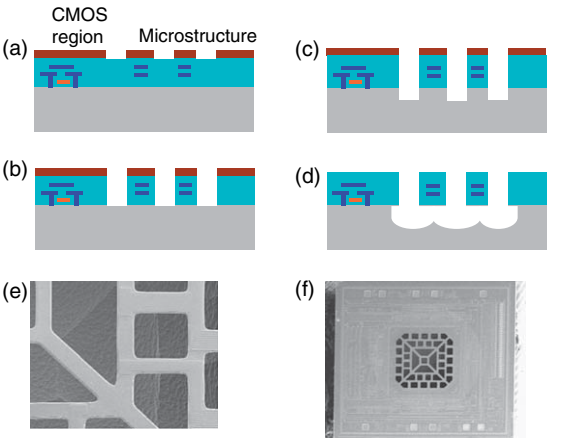


Figure 46 (a)–(d) Cross-sectional process flow of MEMSIC's complementary metal oxide semiconductor (CMOS) microelectromechanical systems (MEMS) process. (e) A released microstructure showing the undercut silicon pit and suspended oxide/metal beams. (f) Scanning electron micrograph (SEM) of a whole sensor die.

2.05.4.9 A DRIE CMOS-MEMS Three-axis Accelerometer

DRIE CMOS-MEMS processes provide large proof mass and integrated interface circuits. One of the key challenges is to obtain reliable electrical isolation of silicon. Because electrical isolation structures often have poor thermal conductivity, overheating problems occur during the DRIE silicon etch. Overheating in DRIE results in large silicon undercut, which increases the gaps of capacitive sensing comb fingers. The DRIE CMOS-MEMS process introduced here uses two silicon etch steps to obtain electrical isolation and narrow comb-finger gaps, and solves the overheating problem by using a temporary thermal path.

The process flow is illustrated in **Figure 47**. It starts with a backside etch that produces a silicon membrane and defines the structure thickness (**Figure 47(a)**). Then an anisotropic oxide etch is performed, followed by applying a thick sacrificial polymer layer (~50 μm) to the back of the silicon membrane (**Figure 47(b)**). This polymer layer may simply be the photoresist and will be used as a temporary heat path. Next, the top Al layer is etched (**Figure 47(c)**), and then a deep silicon etch and undercut is used to form electrical isolation structures for silicon (**Figure 47(d)**). After that, a second anisotropic SiO_2 etch and DRIE silicon etch is performed to define comb fingers and mechanical springs (**Figure 47(e)**). Finally the photoresist is removed in oxygen plasma for final device release.

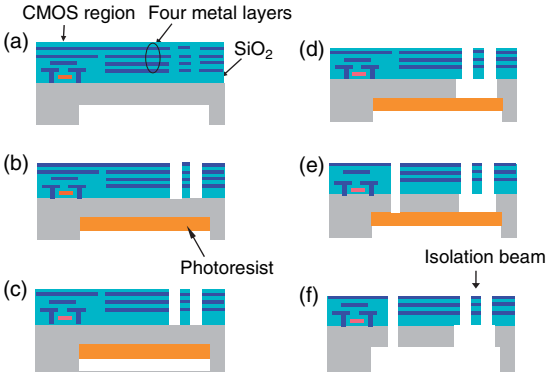


Figure 47 A deep reactive ion etch (DRIE) complementary metal oxide semiconductor (CMOS)-microelectromechanical systems (MEMS) process flow. (a) Backside etch. (b) Backside photoresist coating followed by front side anisotropic SiO_2 etch. (c) Top Al etch. (d) Deep Si etch and undercut to form isolation structures. (e) Anisotropic SiO2 etch and DRIE Si etch for comb fingers and mechanical springs. (f) Photoresist ashing for final release.

Figure 48 shows a SEM of a three-axis accelerometer fabricated using this process. One inset in the figure shows the backside of a group comb fingers with trench isolation. The thickness of the structural silicon layer is about 50 μm. The measured noise floor of the accelerometer is about 12 $\mu g\ Hz^{-1/2}$ (Qu *et al.* 2006).

2.05.5 Design Considerations

As we have seen in the above sections, there are various acceleration sensing mechanisms and each of them can then be applied to accelerometers by using different fabrication technologies. Which way to go may often largely depend on the available facilities and expertise. The following provides some general guidelines for decision making.

2.05.5.1 Single Axis or Three Axis

Acceleration is a vector. In many cases, acceleration in all three axes must be monitored. Assembling two or three devices to form three-axis sensing will drastically increase the packaging size compared with a monolithic three-axis accelerometer. Although several companies including ADI, Bosch, and ST have already launched monolithic three-axis accelerometers in the market, the fabrication complexity and compromised performance in one or two of the axes are still concerns. Therefore, the size constraint must be carefully evaluated to determine whether three-axis accelerometers should be used. Table 2 lists the three-axis readiness of all the sensing mechanisms.

As shown in Table 2, each sensing mechanism has its favorable sensing axis. Capacitive sensing (surface) is suitable for lateral-axis sensing using comb fingers. Its *z*-axis implementation is more challenging due to the substrate parasitic capacitance and/or the difficulty for forming fully differential capacitive bridges. Both capacitive sensing (bulk) and piezoresistive sensing (bulk) are typically used for *z*-axis sensing, but the twisting modes of a proof mass can be used for lateral-axis sensing. Tunneling sensing in a lateral axis is not easy to implement because of the difficulty to fabricate a sharp tip oriented in the wafer plane, while thermal sensing is suitable for lateral axes. Resonant accelerometers can depend either way on

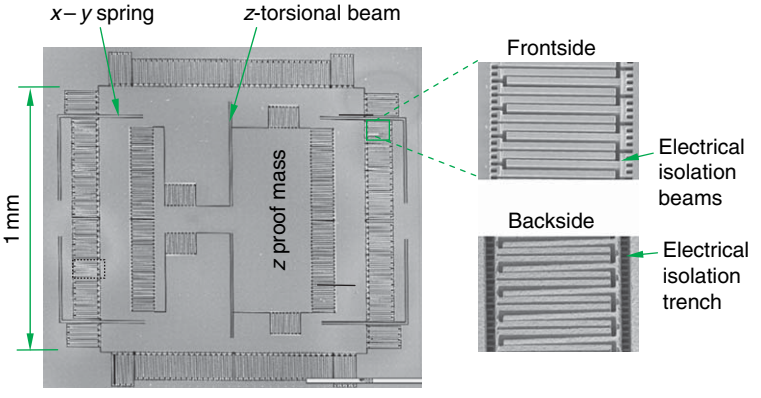


Figure 48 Scanning electron micrograph (SEM) of a deep reactive ion etch (DRIE) complementary metal oxide semiconductor CMOS-microelectromechanical systems (MEMS) three-axis accelerometer.

Table 2 Comparison of the readiness of various acceleration sensing mechanisms

Type	*Primary*	*Secondary*	*Triaxial readiness*[a]
Piezoresistive	Vertical	Lateral	2
Piezoelectric	Vertical	Lateral	2
Capacitive			
Surface (poly-Si or poly-SiGe)	Lateral	Vertical	2
Surface (multiconductor layer)	Lateral	Vertical	3
Bulk micromachining	Vertical	Lateral	2
Thermal	Lateral	Vertical	1
Tunneling	Vertical	Lateral	1
Optical	Vertical	Lateral	1
Resonant	Either	Either	2

[a]1 – possible but difficult, 2 – easy but with performance degradation, 3 – good.

the actual actuation and sensing mechanism employed. There are many three-axis accelerometer designs reported from both academia and industry. These designs are mainly based on piezoresistive, capacitive, and piezoelectric sensing.

2.05.5.1.1 Three-axis piezoresistive accelerometers

Takao *et al.* (1995) fabricated a three-axis piezoresistive accelerometer for high-temperature applications using an SOI substrate. Andersson (1995) fabricated inclined beams using anisotropic etching of silicon to achieve three-axis sensing. Kwon and Park (1997) reported a three-axis piezoresistive accelerometer using polysilicon on top of SOI substrate, but the cross-axis sensitivities were about 10%. Plaza *et al.* (1998) reduced the cross-axis sensitivities of a three-axis piezoresistive accelerometer to less than 1.6% using a BESOI substrate and a combined bulk and surface micromachining process. Takao *et al.* (1997, 2001) utilized piezoresistive effects of p-MOSFETs (metal oxide semiconductor field effect transistors) and made a CMOS-integrated three-axis piezoresistive accelerometer. Kruglick *et al.* (1998) demonstrated a CMOS three-axis piezoresistive accelerometer by using assembled vertical plates for lateral-axis sensing. Chaehoi *et al.* (2005) combined piezoresistive *z*-axis sensing and heat-transfer lateral-axis sensing on a single chip to make a CMOS three-axis accelerometer using anisotropic wet silicon etching.

2.05.5.1.2 Three-axis capacitive accelerometers

Mineta *et al.* (1996) used a proof mass whose center of gravity is raised above its suspension beams so that vertical and lateral accelerations can be detected by a parallel shift and tilt of the proof mass, respectively. Lemkin *et al.* (1997) demonstrated a force-balanced CMOS three-axis accelerometer using a surface micromachining process. Lotters *et al.* 1998 used a highly symmetric cubic structure and a polydimethylsiloxane (PDMS) spring flexure. Matsumoto *et al.* (1999) achieved three-axis acceleration sensing using an SOI substrate. Li *et al.* (2001) used a SiO_2/Si_3N_4 flexure to suspend a bulk micromachined proof mass. Reichenbach *et al.* (2003) demonstrated a three-axis accelerometer using a single proof mass. Szaniawski *et al.* (2004) designed a three-axis accelerometer using multiuser MEMS processes (MUMPs). Rodjegard *et al.* (2005) created unique slanted beams for both in-plane and out-of-plane acceleration sensing. Chae *et al.* (2005) achieved about 1 $\mu g Hz^{-1/2}$ noise floors for both in-plane and out-of-plane acceleration by using a combined surface and bulk micromachining process. Qu *et al.* (2006) demonstrated an integrated three-axis accelerometer with a noise floor of 12 $\mu g Hz^{-1/2}$ using a DRIE CMOS-MEMS process.

There are also three-axis piezoelectric accelerometers based on either ZnO or PZT (Kunz *et al.* 2001, Okada 1995, de Reus *et al.* 1999, Scheeper *et al.* 1996). All the above-discussed three-axis accelerometers either use a single proof mass, one for in-plane sensing and one for out-of-plane sensing, or use one for each axis. Sharing a proof mass means a small chip size, but at the same time the coupling and cross-axis sensitivities will be increased.

On the commercial side, three-axis capacitive accelerometers have been released by a few companies in the past two years (Analog Devices Inc. 2005, Bosch Sensortec 2005, Colibrys Inc. 2006, ST Microelectronics 2004, Freescale 2006). Some three-axis piezoresistive accelerometers are also available as volume products (Fujitsu 2005, Hitachi Metals 2005). **Table 3** lists some of the three-axis accelerometers that currently are commercially available. Very recently, MEMSIC added *z*-axis acceleration capability to their dual lateral-axis thermal accelerometers (Zhao and Cai 2005).

Table 3 List of some commercially available three-axis accelerometers

	Model	*Size (mm)*	*Range (g)*	*Noise ($\mu g\ Hz^{-1/2}$)*	*Bandwidth (kHz)*	*Current (mA)*	*Packaging*
ADI	ADXL330	$4 \times 4 \times 1.45$	±3.6	280	1.6	0.2	Single-chip
Bosch	SMB363	$4 \times 4 \times 1.2$	±2	200	1.0	0.6	Two-chip
Colibrys	SF300L	$14 \times 14 \times 2.3$	±3	0.25	1.0	<30	Two-chip
Freescale	MMA7260Q	$6 \times 6 \times 1.45$	1.5–6	350	3.4 (*z*) 6.0 (*xy*)	0.5	Two-chip
Hitachi	H30CD	$2.9 \times 2.9 \times 0.9$	±2–4	–	–	0.35	Two-chip
Hitachi	H34CD	$3.4 \times 3.4 \times 0.9$	±3			0.4	
Kionix	KXPS5	$3 \times 5 \times 0.9$	±3	200	1.0	1.0	Two-chip
Kionix	KXP84	$5 \times 5 \times 1.2$	±2	175	3.3 (*xy*) 1.7 (*z*)	1.0	Two-chip
Fujitsu	FAR-S2AB	$5 \times 5 \times 2.3$	2–4	–	0.3	3.0	Two-chip
ST	LIS302DL	$3 \times 5 \times 0.9$	2–8	–	0.2	0.3	Two-chip
ST	LIS3L02AL	$5 \times 5 \times 1.52$	±2	50	1.5	0.85	Two-chip

ADI, Analog Devices Inc.; ST, ST Microelectronics.

2.05.5.2 Sensing Method

Which sensing method to choose will be determined by several factors: resolution, bandwidth, cost, available facilities, etc. **Table 4** compares various sensing methods.

2.05.5.3 Fabrication Method

For each sensing method, there are various fabrication processes and structural materials to choose from. Anisotropic wet bulk micromachining is the most cost-effective process, but it requires protection and etch stop, which greatly restricts its application range. Surface micromachining is quite mature now. Polysilicon layers with thickness greater than 10 μm have been demonstrated (Furtsch *et al.* 1997). Low-temperature deposition materials such as electroplated metals (Cole 1991) and polycrystalline silicon germanium (Franke *et al.* 2003, Rusu *et al.* 2003) have also been developed. In addition, thin-film materials such as Al/oxide composite from CMOS processing can be used for structural materials (Fedder *et al.* 1996, Xie and Fedder 2000). DRIE silicon etch is a newly added powerful tool and has been used to achieve bulk silicon equivalent performance with a foundry CMOS-compatible process (Qu *et al.* 2004, Xie *et al.* 2002).

2.05.5.4 Integrated vs. Hybrid

Integration is always desired, but the cost for the integration may be high. The majority of three-axis accelerometers in the market are hybrid, where single- or dual-axis sensing elements and signal conditioning circuits are assembled together and packaged in the same enclosure. There are some trade-offs between monolithic integration and hybrid packaging. The advantages of monolithic integration include small packaging sizes, less signal attenuation, and accurate alignment of the three orthogonal sensing axes. But fabrication processes are more complicated and most likely the yield will not be high. Readers who are interested in the integrated approach may further refer to Fedder *et al.* (2005),

Table 4 Performance comparison of various sensing methods

Type	*Resolution ($\mu g\ Hz^{-1/2}$)*	*Bandwidth (kHz)*	*Acceleration range*	*Readout circuits*	*Fabrication cost*	*Temperature performance*
Piezoresistive	100–1000	0.1–5	Low to medium	Easy	Low	Poor
Piezoelectric	1–1000	~15	Wide	Easy	Medium	Good
Capacitive	0.01–100	1–20	Wide	Difficult	Low	Very good
Thermal	~1000	<0.1	Wide	Medium	Low	Good
Tunneling	1–0.01	~0.1	Low	Medium	High	Good
Optical	0.01–1000	0.1–5	Wide	Easy	Varies	Good
Resonant	1–1000	0.1–1	Low to medium	Medium	Varies	Good

which provides a comprehensive overview of integrated accelerometers and gyroscopes. The hybrid solution, on the other hand, allows mechanical structures and readout circuits both being optimized and fabricated separately by dedicated and well-tuned processes. The higher yield and the always improving packaging techniques may make up the relatively large package sizes.

2.05.5.5 Readout Circuits

The readout circuits for piezoresistive accelerometers are simple because of the low impedance at the sensing nodes. Optical accelerometers also have a fairly simple interface circuit, which basically is a photodiode plus a transimpedance amplifier. Resonant accelerometers need more complicated circuits as both resonance excitation and strain detection are involved. Self-oscillation circuits and force feedback closed-loop control circuits are discussed in detail in Khammash *et al.* (2005), Yeh and Najafi (1998), and Partridge *et al.* (1999). The interface circuitry for capacitive accelerometers is the most challenging due to the high impedance of the capacitive sensing nodes and the parasitic capacitance. Interested readers can refer to Yun *et al.* (1992), Kemp and Spangler (1995), Boser (1997), Kulah *et al.* (2000), Kajita *et al.* (2002), Wu *et al.* (2002), Amini *et al.* (2004), Yazdi *et al.* (2004), and Fang *et al.* (2006).

2.05.5.6 Accelerometers for Applications in Harsh Environments

There are many situations where the sensors must operate in extremely high temperature, shock, vibration, or radiation. For example, for munition launching, the acceleration can easily exceed 10 kg. For underground drilling, the temperature could be as high as 500 °C or more. The strong radiation must be considered in all space missions (Brown 2003, Cass 2001). The piezoresistive method is often used for high-*g* applications. Piezoresistive accelerometers that can operate up to 20–100 kg have been reported (Ning *et al.* 1995, Wang *et al.* 2003). But piezoresistive accelerometers have poor temperature performance. Capacitive accelerometers can be used for high-*g* applications in the 1–20 kg range with better temperature stability (Beliveau *et al.* 1999, Davis *et al.* 2004, Stauffer 2006, Stewart *et al.* 2004). In order to further improve the temperature performance, more stable materials such as SiC have been used (Atwell *et al.* 2003). The radiation effects of some commercial accelerometers have been studied (Cournar *et al.* 2004, Knudson *et al.* 1996, Lee *et al.* 1996). Boyadzhyan and Choma (1998) found that tunneling accelerometers can survive strong direct gamma radiation (Boyadzhyan and Choma 1998).

2.05.6 Packaging

The structure placed around the mechanical spring–mass system of an accelerometer creates a package that mechanically interfaces the accelerometer sensing mechanism with the mechanical stimulus that is being measured. Accelerometers are one of the few MEMS devices that can be sealed in an IC package. Most other sensors, for example, pressure sensors, must allow media to enter the package. The accelerometer package performs several different functions. The primary purpose is to electrically interconnect and mechanically mount the sensor. The package also isolates the internal atmosphere from the environment and therefore protects it from contamination. For resonant and tunneling accelerometers, the package must also provide a low-pressure environment to achieve high quality factor and reduce noise (Lee *et al.* 2000, Wang *et al.* 1997). For thermal accelerometers, the package should be hermetic and maintain a constant chamber pressure. Electrical signals due to leakage or coupling to electromagnetic fields must be isolated from the sensor to minimize error.

Packaging can occur after IC dicing within a preformed container or by molded encapsulation of the sensing element. Alternatively, packaging can occur at the silicon die or wafer level in a procedure commonly referred to as capping. There are two types of capping: bond capping and thin-film encapsulation.

2.05.6.1 Capping

The attachment of a structure over the mechanical MEMS features can be accomplished at the die level and therefore makes it possible to assemble a MEMS die with the same tools and procedures as typical ICs.

The cap is made from a silicon or glass wafer with an etched cavity and is aligned and bonded to the MEMS wafer (**Figure 49(a)**). The most common bonding material is a glass frit paste that is screen printed onto the cover or cap in a shape that outlines the bonding area (Felton *et al.* 2004, Li and Tseng 2001, Lutz *et al.* 1997, Yun *et al.* 2005). The

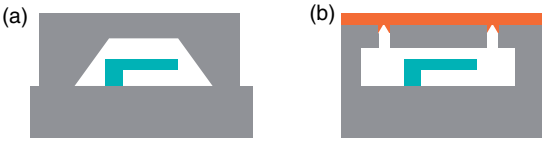


Figure 49 Cross-sections of capping. (a) Bond capping. (b) Thin-film encapsulation.

paste is dried of solvents and sintered, then aligned with the MEMS wafer. Sufficient force and temperature is applied to ensure contact across the wafer and wetting of the interfaces. The atmosphere used during the bonding operation should remain oxidizing to ensure adequate glass wetting and avoid reduction of the oxide to metals such as lead, as lead oxide is a common constituent in the frit. A low-moisture sealing atmosphere and gettering with borates are commonly used to ensure that the atmosphere remains dry inside the cap. Glass frit can create a seal over topology including circuitry and therefore preserve the die area, as space need not be reserved for sealing. Sealing is also possible with solders or thermocompression of gold-to-gold surfaces. In addition to using hot plates, the thermal energy for bonding can also be provided by localized heating of on-chip polysilicon heaters (Lin 2000). When the cap is etched for the cavity over the sensor, a channel is also created around the perimeter of the seal area. A diamond saw that is used for wafer dicing makes a cut through the cap wafer and into the perimeter channel. To avoid loose fragments of silicon, this perimeter is filled with wax from the edges of the wafer. After the cap wafer is cross-cut, the wax and fragments are removed. The final step in the capped and sealed sensor wafer is sawing through the wafer to create an individual die.

Protecting the MEMS structure by depositing additional layers of material over the MEMS structure to create a cover is referred to as thin-film encapsulation or in situ capping (Hochst *et al.* 2004, Liu and Tai 1999, Metzger *et al.* 2007, Park *et al.* 2006). The use of vapor etchants such as xenon difluoride or vapor HF makes it possible to remove silicon or silicon dioxide after the cap or top layer is in place. To complete the sealing of the MEMS structure, the access areas for vapor etch are plugged, as illustrated in **Figure 49(b)**. Bosch has used the high-speed deposition rates achieved in epitaxial silicon reactors to deposit thick polysilicon layers over oxide. This results in a polysilicon structure that can be layered to form sensor elements and a cap. The limitation with this approach is that the temperatures are over 850 °C.

With capped die, it is possible to create wafer scale packaging. This is accomplished by creating solder bumps on the capped die and directly mounting the die with the solder bumps. Vias through the cap or the substrate can provide interconnection to the sensor.

2.05.6.2 Hermetic and Plastic Packaging

Although dust and protective seals can occur with polymers, they allow moisture to enter the part. Glass or metal sealing creates a moisture barrier and is referred to as hermetic. Accelerometers, along with temperature sensors and magnetic sensors, can function inside of a hermetic package while most pressure sensors and flow and chemical sensors must interact with the environment and are open. A large base of experience has been developed for IC hermetic packages and is directly applicable to MEMS accelerometers. A sealed package, assuming a proper atmosphere exists inside, helps to ensure a stable and reproducible device. Moisture and condensable vapors can change the mass or pressure or create leakage paths for accelerometers and thus change the indicated acceleration. Reliability issues with corrosion, stress corrosion, and migrated resistive shorts can occur as well as electrical drift.

It is common for MEMS devices to have special coatings on the structures to minimize stiction (Ashurst *et al.* 2001, Srinivasan *et al.* 1998). These coatings can migrate and agglomerate with time, temperature, and moisture and reduce the effectiveness of the antistiction coatings. Hermetic packages ensure that the internal atmosphere remains dry and in equilibrium with the antistiction coatings. In the case of structures to be operated in resonance, such as resonant accelerometers, a vacuum environment is used and therefore a sealed package will maintain pressures less than 100 mTorr and, with careful package construction and gettering, this pressure can be reduced to 1 mTorr.

Plastic packages, although not hermetic due to the nature of moisture diffusion through polymers, are still widely used for MEMS accelerometers. Through the use of capping or sealing at the die level, a plastic package with a hermetic cavity around the MEMS structure is created. This makes it possible to

assemble the accelerometer on a conventional IC packaging line. An uncapped die must be handled in a class 100 clean room, which is unconventional for die attach and wire bonding operations, and special tools and processes must be used to protect the fragile MEMS structure from liquids and particles as well as mechanical damage.

Plastic packages are a cost-effective and high-volume method of packaging, which can be found to occur in two types of configurations, postmolded and premolded. The premolded package is made by injecting plastic around a lead frame to form an outline with a cavity for future die placement. The postmolded plastic package process begins with the die attachment to the lead frame followed by gold thermocompression bonds connecting the pads on the die to the leads on the frame. The part is then molded over with plastic.

2.05.6.3 Package Outline and Mounting

Although MEMS accelerometers have no standard outline or mounting conventions, the main features have been fashioned from IC packages and therefore have common outlines with single and dual-in-line packages and surface-mounted parts. The very high volume use of hermetic LCCs for cell phones and crystal oscillators is a popular package for accelerometers as well. With the strong trend to smaller size parts both in footprint and in thickness, the use of plastic packages has been the only economic solution. To achieve thin packages, the back grinding of the die and often the cap is performed to achieve a total device thickness of less than 1 mm (Lacsamana *et al.* 2005), as shown in **Figure 50**.

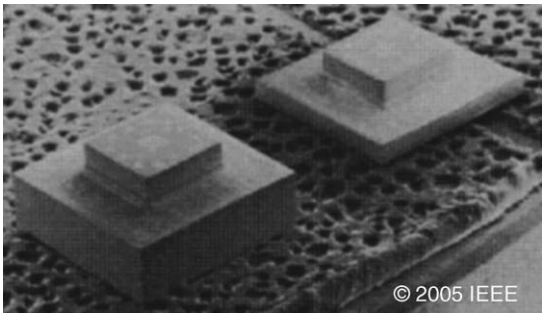


Figure 50 Scanning electron micrograph (SEM) showing a standard thickness capped die and 250-μm backgrinded capped die. The die size is 2.3 mm × 2.3 mm.

2.05.7 Application Considerations

The applications of accelerometers in different areas have different considerations. High-volume and low-cost applications such as in automotive, consumer, and health care value pricing and reliability, while high-performance and low-volume applications in defense and aerospace value accuracy and measurement dependability. **Figure 51** gives an overall view of accelerometer applications in various markets (Jean-Michel 2004). In the following, we discuss accelerometer applications in automotive, aerospace, measurement and instrumentation, and consumer markets.

2.05.7.1 Automotive Markets

MEMS accelerometers are most widely known for their applications in safety systems or airbags (Sulouff 1991). The deceleration of a vehicle can be sensed throughout the car, but due to the deformation that occurs during an impact, it is possible to have a large deceleration in the initial impact zone and very little change in velocity at the passenger location. Initial MEMS acceleration applications were at the center of the vehicle in the region referred to as the transmission hump or firewall. This made it possible to determine the amplitude and frequency spectrum of the deceleration and trigger the airbag after a threshold velocity change had occurred. The accelerometers in safety applications measure from 35–50 *g* of amplitude and 400 Hz of bandwidth (filtered). In measuring crash results the transfer function that includes the mounting of the accelerometer in the module is important as many printed circuit board configurations can have *Q*s of over 30 at

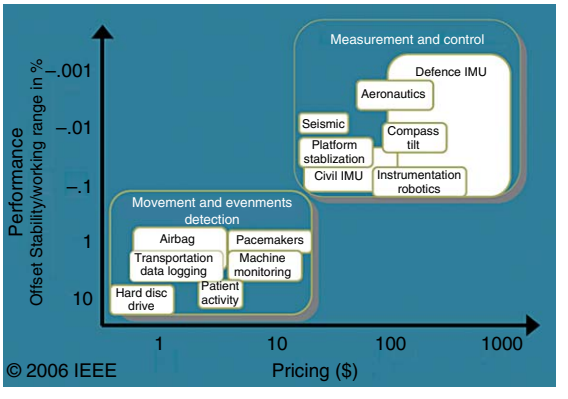


Figure 51 Representation of accelerometer market.

frequencies above 400 Hz and can create signals that appear as decelerations in the range of interest. Clipping of signals beyond the power supply limits also can confuse some algorithms when integration of acceleration to determine velocity is performed. In these algorithms, a measure of the velocity is derived from integration of the deceleration but when a nonsymmetrical response occurs the results are not indicative of actual low-frequency motion. The mechanical dampening as well as the electronic filtering of accelerometers in safety systems and for that matter in all applications must be determined to avoid errors. It is not uncommon to have an inappropriate system response to rocks, curbs, or road features when acceleration outside of the 400-Hz bandwidth of the electronic circuit excites large and often unsymmetrical signals.

The use of acceleration sensors in the forward or crush zone of a vehicle as well as in the side panels for better crash discrimination and side impact measurement has become common. The side impact event occurs in a much shorter time frame, which puts more of a burden on the sensor to discern the level of impact and the appropriate time to fire the side airbags. Frontal crashes have from 30 to several hundred milliseconds to measure and signal the start of an airbag inflation while side impact events must measure and initiate the inflation signal from 5 to 50 ms. In some systems, a pressure sensor along with the accelerometer is used to detect the crushing of the door panel to better discriminate the impact event. The structure of the vehicle is also part of the application consideration, as soft features such as a door panel will crush while support features such as a pillar will remain rigid and transmit the acceleration signal. The accelerometers mounted in the front areas of a vehicle give additional information for the inflation of airbags, especially for passenger side bags, which tend to be larger. Tragic results have occurred when individuals, especially children, have been out of the normal sitting position and a passenger airbag was inflated. This can occur when speeds are as low as 14 miles per hour. The variable inflation rate bags control the rate of inflation, which helps to match the impact event with the safety needs. Data recorded from the sensors mounted at the front crush region especially with an easily crushed vehicle give important information about the crash event.

Low-gee range devices for automotives are used for vehicle stability and vehicle dynamic control. Low-gee accelerometers have bandwidth of up to several hundred hertz and amplitudes from ± 2 to $\pm 10\,g$. The ability to control braking and suspension stiffness has made it possible to add additional handling response to a vehicle. When turning, especially on slippery surfaces such as wet pavement or snow, individual wheels can be braked and therefore tracking according to the driver intentions can be achieved. The low-gee accelerometer is used to measure the lateral acceleration that would create a spin as well as to measure the longitudinal acceleration that describes the increase or decrease in acceleration along the driving axis. These accelerations as well as the yaw rate and steering wheel angle give information to control the vehicle. In addition the vertical acceleration component or z-axis is measured to give a complete understanding of the acceleration. In recent years, information on vertical acceleration has been used to determine the onset of rollover, which is both a vehicle stability and a safety system input requirement. Rollover most often occurs during turning so lateral and longitudinal accelerations and yaw rate are beneficial information as is a rate sensor that measures roll. During rollover, airbags and seat belts can be activated in the most appropriate sequence to minimize injury. This points out the merging of safety systems and vehicle control systems in automotive applications that use accelerometers to sense the increased possibility of a safety event such as rollover while increasing the stability of a vehicle with electronic sensing and controls.

In addition to airbag and vehicle control system applications, accelerometers are finding increasing use in security systems in cars. Early security or alarm sensors were not sensitive to very low frequency so a vehicle could be jacked up slowly and the high-value wheels removed as well towing could occur. With accelerometers that can measure tilt, the alarm system can determine the onset of possible theft. This has become a feature built into some vehicles at the manufacturing stage rather than as an add-on.

2.05.7.2 Aerospace Markets

Accelerometers for aerospace applications have a long history in both small and large missile applications and more recently in projectiles that are used for delivering explosives. Because accelerometers along with gyroscopes can determine the position of a missile, ship, or airplane without additional information, accelerometers for navigational applications have been used for decades. The major source of error in determining position by the use of an accelerometer is the drift in the zero value. With the

double integration of acceleration any DC bias or error would contribute to progressively larger position error with time. Navigational accelerometers with nautical mile per hour of measurement errors have bias stability less than 25 micro-g and are very well characterized for response to vibration, shock, and temperature, which are modeled and compensated for to achieve the nautical mile in an hour's performance. The alignment errors between the different axes are corrected and compensated in an inertial measurement unit or IMU, which has three gyroscopes typically ring laser-based and three accelerometers. To achieve the overall accuracy or performance, trade-offs in accelerometer and gyro performance occur within the IMU. Tactical-grade accelerometers are required to perform for shorter periods of time such as minutes, in harsh environments of shock, vibration, and temperature extremes. They are several orders of magnitude less accurate in terms of positional measurement with null or bias stability from 200 to 1000 μg and scale factor accuracy of 300 ppm. Noise in the range of 50 $\mu g\,\mathrm{Hz}^{-1/2}$ is also typical of tactical-grade accelerometers.

The higher performance of military and aerospace accelerometers coupled with their lower volumes of usage result in higher pricing. **Figure 52** shows a view of an accelerometer used in aerospace applications such as instrument readout. The capacitive accelerometer and its associated electronics are mounted inside of a metal package and the silicon accelerometer has larger mass than automotive accelerometers. The larger mass creates a greater signal and thus makes it possible to achieve a better signal-to-noise ratio and higher resolution.

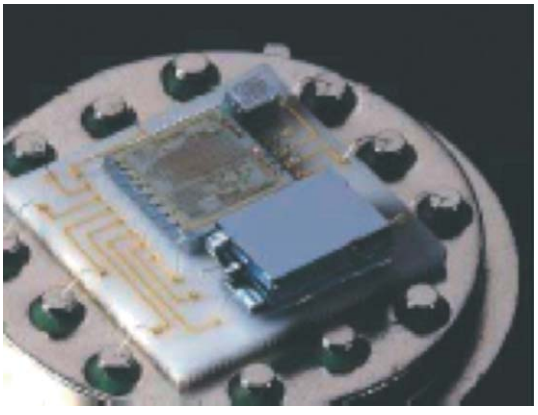


Figure 52 Photo of a Colibrys microelectromechanical systems (MEMS) accelerometer for aerospace applications.

2.05.7.3 Measurement and Instrumentation Markets

The application of accelerometers to measurement and instrumentation is a wide field with many different areas. The measurement of acceleration for shock and vibration has been a major engineering design activity for the last 50 years or more. The piezoelectric properties of quartz and ceramic materials such as PZT have been used to determine the vibration characteristics of mechanical structures from rotating machinery to missile frames and wing sections among others. Accelerometers are mounted over the structure of interest, which can be full size or a scale model and the structure is stressed, tested, or operated in extreme conditions to determine the mechanical response. Small size is crucial in this case to ensure little change of response caused by adding the accelerometer(s).

One example is the controlled crash test to certify that automotive safety systems perform to the specified standards. Accelerometers are used to instrument the vehicle initially to characterize the structure for the proper calibration and operation of an airbag system and later as a confirmation that the proper level of impact and response has occurred. Accelerometers are also used in rotating machinery to monitor bearing noise and predict time to failure. Increases in noise due to metal-bearing surface changes can be sensed with accelerometers. An increase in signal at selected frequencies can be used to determine bearing life. Dynamic balancing of wheels and other mechanisms can be aided by the measurement of acceleration during rotation. The measurement of tilt or deviation from a vertical reference is a common application of accelerometers. Drilling into or under structures and the profiling of structures as referenced to a vertical datum can be helpful in construction, crane operation, and robotic applications.

Another example is dishwasher and washing machines, the vibrations can degrade the machine performance and reduce their lifetime. MEMS accelerometers can be used to measure the vibrations and eventually to identify potential machine problems before equipment failure and allow preventive maintenance to be implemented.

2.05.7.4 Consumer Markets

Accelerometers in consumer products have found some interesting applications in such areas as disk drive protection and gesture recognition. A major change is occurring due to the increase in

functionality along with a significant reduction in pricing. The use of accelerometers to measure motion, when applied as input signals to computers, has been well documented for many years (Cheok *et al.* 2002). The accelerometer-enabled 3D wireless mouse makes it possible to move an object in space and have a corresponding object or cursor follow in a computer-generated visual model. Computer or video games can exploit this technique and make it possible to play the games or do virtual activities such as swinging a tennis racket or drive a vehicle by moving a hand-held controller. Starting from 2005, the availability of three-axis accelerometers with prices less than $2 has opened up new markets for accelerometers in video game controllers and cell phones. With the use of accelerometers a cell phone can conserve power needs by darkening the display when not needed or make selections by scrolling down a list by tilting. It can also convert the display to a horizontal or vertical format. Moreover, the measurement of activity or steps by an accelerometer creates a pedometer function that can encourage a healthy lifestyle.

Accelerometers when synchronized with a row of light-emitting diodes (LEDs) can spell out a message when the LEDs are waved across a room. A common consumer application of accelerometers now increasingly used in cell phones is the correction for tilt for a magnetic sensor. With the use of global positioning system (GPS) and a magnetic sensor, location-based services are enabled, making it possible to identify special sales or lunch menus by just pointing a cell phone at a building. There are of course many ways to analyze a golf swing or other types of motion to improve performance or rehabilitate in physical therapy and the wide range of applications to activate toys and robotic objects are just starting to be discussed.

2.05.8 Summary and Future Prospects

In this section, various acceleration sensing mechanisms are discussed. Among them, piezoresistive sensing needs standard IC processing and materials only, can be batch fabricated using inexpensive bulk micromachining, and does not require small gaps or precise gap control. Its major limitation is the large temperature sensitivity. Capacitive sensing has good temperature stability, high resolution, and low fabrication cost, but its performance is greatly affected by parasitic capacitance and it is also susceptible to EMI. Optical sensing is immune to EMI and no connecting wires are needed especially in an electrically noisy environment. Optical sensing also inherently has the capabilities of remote access and multiplexing. The challenge is the packaging as discrete optical components are involved. Piezoelectric sensing has large bandwidth and high resolution. It is also suitable for building resonant accelerometers because of its dual nature of actuation and force sensing. Its main limitation is poor DC response due to leakage. The use of non-IC materials also increases the fabrication cost. Tunneling accelerometers have the potential to obtain ultrahigh nano-g resolution, but the fabrication process is complicated and the tunneling tip height control is still challenging. Thermal sensing has very simple structures with no moving parts. It is also very inexpensive and completely foundry CMOS compatible, but its resolution and bandwidth are limited.

Accelerometers can be fabricated in many different ways. The basic process modules include bulk micromachining, surface micromachining, wafer bonding, and DRIE. In most cases, the fabrication involves a combination of two modules or more. The majority of the commercial accelerometers are surface micromachined. One advantage of surface micromachining is its potential of CMOS integration. However, due to some technical challenges, two-chip solutions are still dominant in commercial products. Bulk micromachining is often combined with wafer bonding (glass–silicon or silicon–silicon) to produce high-performance accelerometers. A recent development in which SCS sensing elements are created in CMOS substrate by using DRIE shows some promising results.

In terms of materials, almost all MEMS accelerometers are based on silicon including SOI. Various structural materials have been used, including polysilicon, silicon, SiC, nickel, quartz, PZT, ZnO, poly-SiGe, and oxide/aluminum composite. Poly-SiGe is a promising material for CMOS integration due to its low deposition temperature. There are also a variety of sacrificial materials such as oxide, polysilicon, silicon, metals, photoresist, and various polymers.

Due to the universal presence of motion, vibration, and shock, accelerometers can be applied almost everywhere, from underground drilling to space navigation, from menu selection of cell phones to safety control of vehicles, and from personal assistance of elderly people to performance improvement of athletes. The principal driving force for

accelerometer development has been the automotive industry. With the price decreasing rapidly, we will be witnessing the consumer market emerging as a driving force at a much larger scale. Capacitive accelerometers will continue their dominance overall and especially in consumer markets. Thermal sensing will play a big role in low-bandwidth applications such as physiological activity monitoring in health care. Optical accelerometers may find their niche applications in harsh environments.

The recent competition on monolithic three-axis accelerometers will continue on for a few years. The large volumes have demonstrated a price elasticity and an expectation that three-axis accelerometers will be below $1 each. More and more new applications will be enabled by the small size and low price. Meanwhile, new materials and new processes with CMOS compatibility will continue to be developed especially using SOI wafers. Single-chip integrated wireless accelerometers are not far out of reach. Body motion wireless sensor networks for health monitoring or rehabilitation, household motion sensor networks for security, and self-powered implantable accelerometers are just a few examples that are under development.

Acknowledgments

During the preparation of this manuscript, the authors had help from many people in searching papers, making drawings, providing device pictures, and reviewing the manuscript. The authors are grateful to all of them. In particular, the authors thank Drs. Wilhelm Frey and Zhiyu Pan of Bosch, Drs. Michael Judy and Jinbo Kuang of Analog Devices, Inc., Dr. Dave Monk of Freescale Semiconductor Inc., Mr. Xin Liu and Dr. Yongyao Cai of MEMSIC Inc., Kristine Wiseman of ST Microelectronics Inc., Prof. Hongwei Qu of Oakland University, Dr. Brett Warneke of Dust Networks, and Kemiao Jia and Lei Wu of the University of Florida.

References

Abbaspour-Sani E, Huang R S, Kwok C Y 1995 A wide-range linear optical accelerometer. *Sens. Actuators A Phys* **49**, 149–54

Aikele M, Bauer K, Ficker W, Neubauer F, Prechtel U, Schalk J, Seidel H 2001 Resonant accelerometer with self-test. *Sens. Actuators A Phys* **92**, 161–7

Albert W C 1994 Monolithic quartz structure vibrating beam accelerometer (VBA). *Proc. 1994 IEEE International Frequency Control Symposium,* Boston, MA, USA, pp. 415–420

Amarasinghe R, Dao D V, Toriyama T, Sugiyama S 2007 Development of miniaturized 6-axis accelerometer utilizing piezoresistive sensing elements. *Sens. Actuators A Phys, vol. 134,* pp. 310–320, in press

Amini B V, Pourkamali S, Zaman M, Ayazi F 2004 A new input switching scheme for a capacitive micro-*g* accelerometer. Tech. Dig. 2004 Symp. VLSI Circuits, Honolulu, Hawaii, pp. 310–313

Analog Devices Inc 2005 Analog Devices' 3-axis MEMS accelerometer available now. http://www.analog.com/en/content/0,2886,764%255F%255F87600,00.html

Andersson G I 1995 A novel 3-axis monolithic silicon accelerometer. *Proc. 8th Int. Conf. Solid-State Sensors and Actuators (Transducers '95),* Stockholm, Sweden, pp. 558–561

Ashurst W R, Yau C, Carraro C, Maboudian R, Dugger M T 2001 Dichlorodimethylsilane as an anti-stiction monolayer for MEMS: A comparison to the octadecyltrichlorosilane self-assembled monolayer. *J. Microelectromech. Syst* **10**, 41–9

Atwell A R, Okojie R S, Kornegay K T, Roberson S L, Beliveau A 2003 Simulation, fabrication and testing of bulk micromachined 6H–SiC high-*g* piezoresistive accelerometers. *Sens. Actuators A Phys* **104**, 11–18

Baglio S, Castorina S, Esteve J, Savalli N 2004 Highly sensitive silicon micro-*g* accelerometers with optical output. *Proc. 2004 Int. Symp. Circuits and Systems (ISCAS '04)*, Vancouver, Canada, Vol. 4, pp. 868–71

Barth P W, Pourahmadi F, Mayer R, Poydock J, Petersen K 1988 A monolithic silicon accelerometer with integral air damping and overrange protection. Tech. Dig. 1988 IEEE Solid-State Sensor Actuator Workshop, Hilton Head Island, South Carolina, US, pp. 35–38

Baski A A, Albrecht T R, Quate C F 1988 Tunnelling accelerometer. *J. Microsc.* **152**, 73–6

Beliveau A, Spencer G T, Thomas K A, Roberson S L 1999 Evaluation of MEMS capacitive accelerometers. *IEEE Design Test Comput.* **16**, 48–56

Benech P, Chambered E, Monllor C 1996 Acceleration measurement using PVDF. *IEEE Trans. Ultrason. Ferroelect. Freq. Control* **43**, 838–43

Bernstein J, Miller R, Kelley W, Ward P 1999 Low-noise MEMS vibration sensor for geophysical applications. *J. Microelectromech. Syst.* **8**, 433–8

Berkoff T A, Kersey A D 1996 Experimental demonstration of a fiber Bragg grating accelerometer. *IEEE Photonics Technol. Lett.* **8**, 1677–9

Bosch Sensortec GmbH 2005 SMB360 accelerometer datasheet, Ver. 1.0-112005

Boser B E 1997 Electronics for micromachined inertial sensors. *Tech. Dig. 1997 Int. Conf. Solid-State Sensors and Actuators (Transducers '97),* Chicago, IL, US, pp. 1169–1172

Boser B E, Howe B T 1995 Surface micromachined accelerometers. *Proc. 1995 IEEE Custom Integrated Circuits Conference,* Santa Clara, California, pp. 337–344

Boser B E, Howe R T 1996 Surface micromachined accelerometers. *IEEE J. Solid State Circuits* **31**, 366–75

Boyadzhyan V, Choma J Jr. 1998 High temperature, high reliability integrated hybrid packaging for radiation hardened spacecraft micromachined tunneling accelerometer. *Proc. 1998 IEEE Int. Workshop Integrated Power Packaging (IWIPP '98),* Chicago, Illinois, US, pp. 79–83.

Brown T G 2003 Harsh military environments and microelectromechanical (MEMS) devices. *Proc. 2003 IEEE Sensors Conference,* Toronto, Canada, pp. 753–760

Burns D W, Horning R D, Herb W R, Zook J D, Guckel H 1996 Sealed-cavity resonant microbeam accelerometer. *Sens. Actuators A Phys* **53**, 249–55

Burrer C, Esteve J 1995 A novel resonant silicon accelerometer in bulk-micromachining technology. *Sens. Actuators A Phys* **46**, 185–9

Burrer C, Esteve J, Lora-Tamayo E 1996 Resonant silicon accelerometers in bulk micromachining technology – An approach. *J. Microelectromech. Syst.* **5**, 122–30

Cass S 2001 MEMS in space. *IEEE Spectrum* **38**, 56–61

Chae J, Kulah H, Najafi K 2002 A hybrid silicon-on-glass (SOG) lateral micro-accelerometer with CMOS readout circuitry. *Proc. 15th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2002),* Las Vegas, NV, US, pp. 623–626

Chae J, Kulah H, Najafi K 2004 An in-plane high-sensitivity, low-noise micro-*g* silicon accelerometer with CMOS readout circuitry. *J. Microelectromech. Syst.* **13**, 628–35

Chae J, Kulah H, Najafi K 2005 A monolithic three-axis micro-*g* micromachined silicon capacitive accelerometer. *J. Microelectromech. Syst.* **14**, 235–42

Chaehoi A, Latorre L, Nouet P, Baglio S 2003 Piezoresistive CMOS beams for inertial sensing. *Proc. 2003 IEEE Sensors Conference,* Toronto, Canada, pp. 451–456

Chaehoi A, Latorre L, Mailly F, Nouet P 2005 Monolithic CMOS 3-axis accelerometer combining piezoresistive and heat transfer effects. *Proc. 2005 PhD Research in Microelectronics and Electronics*, Lausanne, Switzerland, pp. 612–615.

Chang S C, Putty M W, Hicks D B, Li C H, Howe R T 1990 Resonant-bridge two-axis microaccelerometer. *Sens. Actuators A Phys* **21**, 342–5

Chen H, Bao M, Zhu H, Shen S 1997 A piezoresistive accelerometer with a novel vertical beam structure. *Proc. 1997 Int. Conf. Solid-State Sensors and Actuators (Transducers '97),* Chicago, IL, US, pp. 1201–1204

Chen P-L, Muller R S, Jolly R D, Halac G L, White R M, Andrews A P, Lim T C, Motamedi M E 1982 Integrated silicon microbeam PI-FET accelerometer. *IEEE Trans. Electron Devices* **29**, 27–33

Chen T D, Kelly T W, Collins D, Berthold B, Brosnihan T J, Denison T, Kuang J, O'Kane M, Weigold J W, Bain D 2005 The next generation integrated MEMS and CMOS process on SOI wafers for overdamped accelerometers. *Tech. Dig. 13th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '05),* Seoul, South Korea, pp. 1122–1125

Cheok A D, Ganesh Kumar K, Prince S 2002 Micro-accelerometer based hardware interfaces for wearable computer mixed reality applications. *Proc. 6th Int. Symp. Wearable Computers (ISWC 2002),* Seattle, Washington, US, pp. 223–230

Cole J C 1991 A new sense element technology for accelerometer subsystems. *Tech. Dig. 1991 Int. Conf. Solid-State Sensors and Actuators (Transducers '91),* San Francisco, California, pp. 93–96

Colibrys Inc. 2006 Si-Flex SF-3000L datasheet, Ver. A01.06

Cournar O, Miller F, Buard N, Poirot P, Gaillard R, Marchand L 2004 Total dose effects and SEE screening on MEMS COTS accelerometers. *Proc. 2004 IEEE Radiation Effects Data Workshop,* Atlanta, Georgia, US, pp. 125–129

Cui T, Wang J 2005 Polymer-based wide-bandwidth and high-sensitivity micromachined electron tunneling accelerometers using hot embossing. *J. Microelectromech. Syst.* **14**, 895–902

Dao R, Morgan DE, Kries HH, Bachelder DM 1996 Convective accelerometer and inclinometer, US Patent No. 5,581,034

Dauderstadt U A, Sarro P M, French P J 1998 Temperature dependence and drift of a thermal accelerometer. *Sens. Actuators A Phys* **66**, 244–9

Davis B S, Denison T, Jinbo K 2004 A monolithic high-*g* SOI-MEMS accelerometer for measuring projectile launch and flight accelerations. *Proc. 2004 IEEE Sensors Conference,* Vienna, Austria, pp. 296–299

Devoe D L, Pisano A P 2001 Surface micromachined piezoelectric accelerometers (PiXLs). *J. Microelectromech. Syst.* **10**, 180–6

Dong H, Jia Y, Hao Y, Shen S 2005 A novel out-of-plane MEMS tunneling accelerometer. *Sens. Actuators A Phys* **120**, 360–4

Donohue L A, Hopkins J, Barnett R, Newton A, Barker A 2004 Developments in Si and SiO2 etching for MEMS based optical applications. *Photonics West 2004, San Jose, CA, USA, SPIE*

Fang D, Qu H, Xie H 2006 A 1 mW dual-chopper amplifier for a 50-/spl mu/g/spl radic/Hz monolithic CMOS-MEMS capacitive accelerometer. Tech. Dig. 2006 Symp. VLSI Circuits, Honolulu, Hawaii, pp. 59–60

Fedder G K, Santhanam S, Reed M L, Eagle S C, Guillou D F, Lu M S C, Carley L R 1996 Laminated high-aspect-ratio microstructures in a conventional CMOS process. *Proc. 9th Int. Workshop Micro Electro Mechanical Systems (MEMS '96),* San Diego, California, pp. 13–18

Fedder G K, Chae J, Najafi K, Denison T, Kuang J, Lewis S, Kulah H 2005 Monolithically integrated inertial sensors. In Baltes H, Brand O, Fedder G K, Hierold C, Korvink J, and Tabata O (eds.) *.CMOS-MEMS,*Weinheim, Germany. Wiley-C, Verlag GmbH & Co., Weinheim, Germany.

Felton L E, Hablutzel N, Webster W A, Harney K P 2004 Chip scale packaging of a MEMS accelerometer. *Proc. 54th IEEE Electronic Components and Technology Conference,* Las Vegas, Nevada, pp. 869–873

Ferrari V, Ghisla A, Marioli D, Taroni A 2005 Silicon resonant accelerometer with electronic compensation of input–output cross-talk. *Sens. Actuators A Phys* **123-4**, 258–66

Franke A E, Heck J M, Tsu-Jae K, Howe R T 2003 Polycrystalline silicon–germanium films for integrated microsystems. *J. Microelectromech. Syst.* **12**, 160–71

Freal J, Zarobila C, Davis C 1987 A microbend horizontal accelerometer for borehole deployment. *J. Lightwave Technol.* **5**, 993–6

Freescale 2006 MMA7261Q, 1.5g–6g low g three axis micromachined accelerometer, data sheet

Fricke J, Obermeier E 1993 Cantilever beam accelerometer based on surface micromachining technology. *J. Micromech. Microeng.* **3**, 190–2

Fujitsu 2005 FAR-S2AB series, data sheet http://www.fujitsu.com

Furtsch M, Offenberg M, Vila A, Cornet A, Morante J R 1997 Texture and stress profile in thick polysilicon films suitable for fabrication of microstructures. *Thin Solid Films* **296**, 177–80

Goustouridis D, Kaltsas G, Nassiopoulou A G 2004 A CMOS compatible thermal accelerometer without solid proof mass, based on porous silicon thermal isolation. *Proc. 2004 IEEE Sensors Conference,* Vienna, Austria, pp. 848–851

Guckel H, Rypstat C, Nesnidal M, Zook J D, Burns D W, Arch D K 1992 Polysilicon resonant microbeam technology for high performance sensor applications. Tech. Dig. 5th IEEE Solid-State Sensor and Actuator Workshop, Hilton Head Island, SC, USA, pp. 153–156

Guldimann B, Thiebaud P, de Rooij N F, Turpin R A 2000 Micromachined, fiber-optic based accelerometer with shutter modulation. *Proc. 13th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2000),* Miyazaki, Japan, pp. 710–714

Hall N A, Okandan M, Degertekin F L 2006 Surface and bulk-silicon-micromachined optical displacement sensor fabricated with the SwIFT-Lite™ Process. *J. Microelectromech. Syst.* **15**, 770–6

Hartwell P G, Bertsch F M, Miller S A, Turner K L, MacDonald, N C 1998 Single mask lateral tunneling accelerometer. *Proc. 11th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '98),* Heidelberg, Germany, pp. 340–344

Helsel M, Gassner G, Robinson M, Woodruff J 1994 A navigation grade micro-machined silicon accelerometer. *Proc. 1994 IEEE Position Location and Navigation Symposium (PLANS '94),* Las Vegas, NV, pp. 51–58

Henrion W, DiSanza L, Ip M, Terry S, Jerman H 1990 Wide dynamic range direct accelerometer. *4th IEEE Solid-State Sensor and Actuator Workshop,* Hilton Head Island, SC, pp. 153–7

Hiratsuka R, van Duyn D C, Otaredian T, de Vries P 1991 A novel accelerometer based on a silicon thermopile. *Tech. Digest. 1991 Int. Conf. Solid-State Sensors and Actuators (Transducers '91),* San Francisco, California, pp. 420–423

Hitachi Metals 2005 H34C Introduction, www.hitachi-metals.co.jp/e/prod/indx_prod.html

Hochst A, Scheuerer R, Stahl H, Fischer F, Metzger L, Reichenbach R, Larmer F, Kronmuller S, Watcham S, Rusu C 2004 Stable thin film encapsulation of acceleration sensors using polycrystalline silicon as sacrificial and encapsulation layer. *Sens. Actuators A Phys* **114**, 355–61

Huang S, Xinxin L, Yuelin W, Jiwei J, Xiaohong G, Deren L, Lufeng C, Kun Z, Bin X 2003 A piezoresistive accelerometer with axially stressed tiny beams for both much increased sensitivity and much broadened frequency bandwidth. *Proc. 12th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '03),* Boston, MA, pp. 91–94

IEEE-STD-337 1972 IEEE standard specification format guide and test procedure for linear, single-axis, pendulous, analog torque balance accelerometer. *IEEE Std* 337–1972

Jakobsen H 1996 Sensor foundries and production of sensors at SensoNor a.s. *J. Micromech. Microeng.* **6**, 193–6

Jean-Michel S 2004 Market opportunities for advanced MEMS accelerometers and overview of actual capabilities vs. required specifications. *Proc. 2004 Position Location and Navigation Symposium (PLANS 2004),* Las Vegas, NV, pp. 78–82

Jia Y-B, Hao Y-L, Zhang R 2004 Double tuning-fork resonant accelerometer. *Proc. 7th Int. Conf. Solid-State and Integrated Circuits Technology,* Beijing, China, pp. 1812–1815

Juan W H, Pang S W 1998 High-aspect-ratio Si vertical micromirror arrays for optical switching. *J. Microelectromech. Syst.* **7**, 207–13

Kajita T, Un-Ku M, Temes G C 2002 A two-chip interface for a MEMS accelerometer. *IEEE Trans. Instrum. Meas.* **51**, 853–8

Kalenik J, Pajak R 1998 A cantilever optical-fiber accelerometer. *Sens. Actuators A Phys* **68**, 350–5

Kemp C J, Spangler L 1995 An accelerometer interface circuit. *Proc. 1995 IEEE Custom Integrated Circuits Conference,* Santa Clara, CA, USA, pp. 345–348

Kenny T W, Waltman S B, Reynolds J K, Kaiser W J 1991 Micromachined silicon tunnel sensor for motion detection. *Appl. Phys. Lett.* **58**, 100–2

Kenny T W, Kaiser W J, Rockstad H K, Reynolds J K, Podosek J A, Vote E C 1994 Wide-bandwidth electromechanical actuators for tunneling displacement transducers. *J. Microelectromech. Syst.* **3**, 97–104

Khammash M, Oropeza-Ramos L, Turner K L 2005 Robust feedback control design of an ultra-sensitive, high bandwidth tunneling accelerometer. *Proc. 2005 American Control Conference,* Portland, Oregon, pp. 4176–4180

Knudson A R, Buchner S, McDonald P, Stapor W J, Campbell A B, Grabowski K S, Knies D L, Lewis S, Zhao Y 1996 The effects of radiation on MEMS accelerometers. *IEEE Trans. Nucl. Sci.* **43**, 3122–26

Kourepenis A, Petrovich A, Weinberg M 1991 Low cost quartz resonant accelerometer for aircraft inertial navigation. *Tech. Dig. 1991 Int. Conf. Solid-State Sensors and Actuators (Transducers '91),* San Francisco, California, pp. 551–553

Kruglick E J J, Warneke B A, Pister K S J 1998 CMOS 3-axis accelerometers with integrated amplifier. *Proc. 11th IEEE Int. Workshop Micro Electro Mechanical Systems (MEMS '98),* Heidelberg, Germany, pp. 631–636

Kubena R L, Atkinson G M, Robinson W P, Stratton F P 1996 A new miniaturized surface micromachined tunneling accelerometer. *IEEE Electron Device Lett.* **17**, 306–8

Kulah H, Yazdi N, Najafi K 2000 A CMOS switched-capacitor interface circuit for an integrated accelerometer. *Proc. 43rd IEEE Midwest Symp. Circuits and Systems,* Lansing, MI, pp. 244–247

Kunz K, Enoksson P, Stemme G 2001 Highly sensitive triaxial silicon accelerometer with integrated PZT thin film detectors. *Sens. Actuators A Phys* **92**, 156–60

Kurtz AD 1976 Transducers employing gap-bridging shim members. US Patent No. 3,995,247.

Kvisteroy T, Jakobsen H 1998 Force sensor device. *US Pat. No. 5,834,646*

Kwon K, Park S 1997 Three axis piezoresistive accelerometer using polysilicon layer. *Proc. 1997 Int. Conf. Solid-State Sensors and Actuators (Transducers '97),* Chicago, IL, US, pp. 1221–1224

Kyung Il L, Takao H, Sawada K, Ishida M 2003 A three-axis accelerometer for high temperatures with low temperature dependence using a constant temperature control of SOI piezoresistors. *Proc. 15th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '03),* Kyoto, Japan

Lacsamana E S, Navarro R M, Mena M G, Felton L E, Webster W A 2005 Very thin packaging of capped MEMS accelerometer device. *Proc. 7th Electronic Packaging Technology Conference (EPTC 2005),* Singapore, pp. 98–102

Le Traon O, Janiaud D, Muller S, Bouniol P 1998 The VIA vibrating beam accelerometer: concept and performance. *Proc. 1998 IEEE Position Location and Navigation Symposium,* Atlanta, GA, USA, pp. 25–29

Lee C I, Johnston A H, Tang W C, Barnes C E, Lyke J 1996 Total dose effects on microelectromechanical systems (MEMS): Accelerometers. *IEEE Trans. Nucl. Sci.* **43**, 3127–32

Lee B-L, Oh C-H, Lee S, Oh Y-S, Chun K-J 2000 A vacuum packaged differential resonant accelerometer using gap sensitive electrostatic stiffness changing effect. *Proc. 13th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2000),* Miyazaki, Japan, pp. 352–357

Lemkin M, Boser B E 1999 A three-axis micromachined accelerometer with a CMOS position-sense interface and digital offset-trim electronics. *IEEE J. Solid State Circuits* **34**, 456–68

Lemkin M A, Ortiz M A, Wongkomet N, Boser B E, Smith J H 1997 A 3-axis surface micromachined $\sum\Delta$ accelerometer. *Dig. Tech. Papers 44th IEEE Int. Solid-State Circuits Conf. (ISSCC '97),* San Francisco, CA, USA, pp. 202–203

Leung A M, Jones J, Czyzewska E, Chen J, Woods B 1998 Micromachined accelerometer based on convection heat transfer. *Proc. 11th IEEE Int. Workshop Micro Electro Mechanical Systems (MEMS '98),* Heidelberg, Germany, pp. 627–630

Lewis S, Alie S, Brosnihan T, Core C, Core T, Howe R, Geen J, Hollocher D, Judy M, Memishian J, Nunan K, Paine R, Sherman S, Tsang B, Wachtmann B 2003 Integrated sensor and electronics processing for >108 "iMEMS" inertial measurement unit components. Tech. *Dig. IEEE Int. Electron Devices Meeting (IEDM '03)*, Washington DC, pp. 39.1.1–39.1.4

Li G, Tseng A A 2001 Low stress packaging of a micromachined accelerometer. *IEEE Trans. Electron. Packaging Manuf.* **24**, 18–25

Li G, Li Z, Wang C, Hao Y, Li T, Zhang D, Wu G 2001 Design and fabrication of a highly symmetrical capacitive triaxial accelerometer. *J. Micromech. Microeng.* **11**, 48–54

Liao K-M, Chen R, Chou B C S 2006 A novel thermal-bubble-based micromachined accelerometer. *Sens. Actuators A Phys* **130-1**, 282–9

Lin L 2000 MEMS post-packaging by localized heating and bonding. *IEEE Trans. Adv. Packaging* **23**, 608–16

Liu C, Tai Y-C 1999 Sealing of micromachined cavities using chemical vapor deposition methods: Characterization and optimization. *J. Microelectromech. Syst.* **8**, 135–45

Liu C-H, Barzilai A M, Reynolds J K, Partridge A, Kenny T W, Grade J D, Rockstad H K 1998 Characterization of a high-sensitivity micromachined tunneling accelerometer with micro-*g* resolution. *J. Microelectromech. Syst.* **7**, 235–44

Llobera A, Plaza J A, Salinas I, Berganzo J, Garcia J, Esteve J, Dominguez C 2004 Technological aspects on the fabrication of silicon-based optical accelerometer with ARROW structures. *Sens. Actuators A Phys* **110**, 395–400

Loh N C, Schmidt M A, Manalis S R 2002 Sub-10 cm3 interferometric accelerometer with nano-*g* resolution. *J. Microelectromech. Syst.* **11**, 182–7

Lotters J C, Bomer J G, Verloop A J, Droog E A, Olthuis W, Veltink P H, Bergveld P 1998 Design, fabrication and characterization of a highly symmetrical capacitive triaxial accelerometer. *Sens. Actuators A Phys* **66**, 205–12

Lu C, Lemkin M, Boser B E 1995 A monolithic surface micromachined accelerometer with digital output. *Dig. Tech. Papers 42nd IEEE Int. Solid-State Circuits Conf. (ISSCC '95),* San Francisco, CA, USA, pp. 160–161

Luo H, Fedder G K, Carley L R 2000 A 1 mG lateral CMOS-MEMS accelerometer. *Proc. 13th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2000),* Miyazaki, Japan, pp. 502–507

Lutz M, Golderer W, Gerstenmeier J, Marek J, Maihofer B, Mahler S, Munzel H, Bischof 1997 A precision yaw rate sensor in silicon micromachining. *Proc. 1997 Int. Conf. Solid-State Sensors and Actuators (Transducers '97),* Chicago, IL, US, pp. 847–850

MacDonald G A 1990 A review of low cost accelerometers for vehicle dynamics. *Sens. Actuators A Phys* **21**, 303–7

Madou M 2002 *Fundamentals of Microfabrication*. CRC Press, Boca Raton and New York

Mailly F, Giani A, Martinez A, Bonnot R, Temple-Boyer P, Boyer A 2003 Micromachined thermal accelerometer. *Sens. Actuators A Phys* **103**, 359–63

Marcal R F M, Kovaleski J L, Suzim A A 1997 A poly vinylidene fluoride (PVF_2) piezoelectric film based accelerometer. *Proc. 1997 IEEE Instrum. Meas. Technol. Conf. (IMTC '97),* Ottawa, Canada, pp. 908–913

Marty J, Baillieu F, Malki A, Renouf C, Lecoy P 1995 Fibre-optic accelerometer using silicon micromachining techniques. *Sens. Actuators A Phys* **47**, 470–3

Matsumoto Y, Nishimura M, Matsuura M, Ishida M 1999 Three-axis SOI capacitive accelerometer with PLL C-V converter. *Sens. Actuators A Phys* **75**, 77–85

Metzger L, Fischer F, Mokwa W 2007 Polysilicon sacrificial layer etching using ClF3 for thin film encapsulation of silicon acceleration sensors with high aspect ratio. *Sens. Actuators A Phys,* **133**, 259–265

Milanovic V, Bowen E, Zaghloul M E, Tea N H, Suehle J S, Payne B, Gaitan M 2000 Micromachined convective accelerometers in standard integrated circuits technology. *Appl. Phys. Lett.* **76**, 508–10

Mineta T, Kobayashi S, Watanabe Y, Kanauchi S, Nakagawa I, Suganuma E, Esashi M 1996 Three-axis capacitive accelerometer with uniform axial sensitivities. *J. Micromech. Microeng.* **6**, 431–5

Najafi K, Junseok C, Kulah H, Guohong H 2003 Micromachined silicon accelerometers and gyroscopes. *Proc. 2003 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS 2003),* Las Vegas, NV, USA, pp. 2353–2358

Ning Y, Loke Y, McKinnon G 1995 Fabrication and characterization of high *g*-force, silicon piezoresistive accelerometers. *Sens. Actuators A Phys* **48**, 55–61

Norling B L 1988 Superflex: A synergistic combination of vibrating beam and quartz flexure accelerometer technology. *J. Inst. Navigation* **34**, 337–53

Offenberg M, Larmer F, Elsner B, Munzel H, Riethmuller W 1995 Novel process for a monolithic integrated accelerometer. *Proc. 8th Int. Conf. Solid-State Sensors and Actuators (Transducers '95),* Stockholm, Sweden, pp. 589–592

Ohlckers P, Holm R, Jakobsen H, Kvisteroy T, Kittilsland G, Nese M, Nilsen S M, Ferber A 1997 An integrated resonant accelerometer microsystem for automotive applications. *Proc. 1997 Int. Conf. Solid-State Sensors and Actuators (Transducers '97),* Chicago, MI, USA

Ohlckers P, Holm R, Jakobsen H, Kvisteroy T, Kittilsland G, Larsen A, Nese M, Nilsen S M, Ferber A 1998 An integrated resonant accelerometer microsystem for automotive applications. *Sens. Actuators A Phys* **66**, 99–104

Okada K 1995 Tri-axial piezoelectric accelerometer. *Proc. 8th Int. Conf. Solid-State Sensors and Actuators (Transducers '95),* Stockholm, Sweden, pp. 566–569

Park W-T, Partridge A, Candler R N, Ayanoor-Vitikkate V, Yama G, Lutz M, Kenny T W 2006 Encapsulated submillimeter piezoresistive accelerometers. *J. Microelectromech. Syst.* **15**, 507–14

Partridge A, Reynolds J K, Grade J D, Kane B J, Maluf N I, Kovacs G T A, Kenny T W 1999 An integrated controller for tunnel sensors. *IEEE J. Solid State Circuits* **34**, 1099–107

Partridge A, Reynolds J K, Chui B W, Chow E M, Fitzgerald A M, Zhang L, Maluf N I, Kenny T W 2000 A high-performance planar piezoresistive accelerometer. *J. Microelectromech. Syst.* **9**, 58–66

Payne R S, Sherman S, Lewis S, Howe R T 1995 Surface micromachining: from vision to reality to vision [accelerometer]. *Dig. Tech. Papers. 42nd IEEE Int. Solid-State Circuits Conf. ISSCC '95,* San Francisco, CA, USA, pp. 164–165

Perez M, Eklund E J, Shkel A M 2005 Designing micromachined accelerometers with interferometric detection. *Proc. 2005 IEEE Sensors Conference,* Irvine, California, pp. 652–655

Plaza J A, Chen H, Esteve J, Lora-Tamayo E 1998 New bulk accelerometer for triaxial detection. *Sens. Actuators A Phys* **66**, 105–8

Plaza J A, Collado A, Cabruja E, Esteve J 2002 Piezoresistive accelerometers for MCM package. *J. Microelectromech. Syst.* **11**, 794–801

Qu W 1999 3D UV-microforming: Principles and applications. *J. Eng. Sci. Educ.* **8**, 13–19

Qu H 2006 Development of DRIE CMOS-MEMS process and integrated accelerometers, PhD Thesis, University of Florida

Qu H, Fang D, Xie H 2004 A single-crystal silicon 3-axis CMOS-MEMS accelerometer. *Proc. 2004 IEEE Sensors Conference,* Vienna, Austria, pp. 661–664

Qu H, Fang F, Xie H 2006 A monolithic CMOS-MEMS 3-axis accelerometer with a low-power, low-noise dual-chopper amplifier. *Proc. 12th Hilton Head Workshop Solid-state Sensors, Actuators, and Microsystems,* Hilton Head Island, SC, USA

Rdjegrd H, Johansson C, Enoksson P, Andersson G 2005 A monolithic three-axis SOI-accelerometer with uniform sensitivity. *Sens. Actuators A Phys* **123–4**, 50–3

Reichenbach R, Schubert D, Gerlach G 2003 Micromechanical triaxial acceleration sensor for automotive applications. *Proc. 12th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '03),* Boston, MA, USA, pp. 77–80

de Reus R, Gullov J O, Scheeper P 1999 Fabrication and characterization of a piezoelectric accelerometer. *J. Micromech. Microeng.* **9**, 123–6

Riethmuller W, Benecke W, Schnakenberg U, Wagner B 1991 Development of commercial CMOS process-based technologies for the fabrication of smart accelerometers. *Dig. Tech. Papers, 1991. Int. Conf. Solid-State Sensors and Actuators (Transducers '91),* San Francisco, California, pp. 416–419

Rines G A 1981 Fiber-optic accelerometer with hydrophone applications. *Appl. Opt.* **20**, 3453–9

Ristic L, Gutteridge R, Dunn B, Mietus D, Bennett P 1992 Surface micromachined polysilicon accelerometer. *Tech. Dig. 5th IEEE Solid-State Sensor and Actuator Workshop,* Hilton Head Island, SC, USA, pp. 118–121

Rockstad H K, Tang T K, Reynolds J K, Kenny T W, Kaiser W J, Gabrielson T B 1996 A miniature, high-sensitivity, electron tunneling accelerometer. *Sens. Actuators A Phys* **53**, 227–31

Rodjegard H, Andersson G I, Rusu C, Lofgren M, Billger D 2005 Capacitive slanted-beam three-axis accelerometer: I. Modelling and design. *J. Micromech. Microeng.* **15**, 1989–96

Roessig T A, Howe R T, Pisano A P, Smith J H 1997 Surface-micromachined resonant accelerometer. *Proc. 1997 Int. Conf. Solid-State Sensors and Actuators (Transducers '97),* Chicago, MI, USA

Roszhart T V, Jerman H, Drake J, de Cotiis C 1995 An inertial-grade, micromachined vibrating beam accelerometer. *Proc. 8th Int. Conf. Solid-State Sensors and Actuators (Transducers '95),* Stockholm, Sweden, pp. 656–658

Roylance L, Angell J 1978 A miniature integrated circuit accelerometer. *Dig. Tech. Papers. 1978 IEEE Int. Solid-State Circuits Conf,* San Francisco, CA, USA, pp. 220–221

Roylance L M, Angell J B 1979 A batch-fabricated silicon accelerometer. *IEEE Trans. Electron Devices* **26**, 1911–17

Rudolf F, Jornod A, Bergqvist J, Leuthold H 1990 Precision accelerometers with μg resolution. *Sens. Actuators A: Phys.* **21**, 297–302

Rudolf F, Bergqvist J 1991 Silicon micromachining for sensor applications. *Microelectron. Eng.* **15**, 399–406

Rusu C, Sedky S, Parmentier B, Verbist A, Richard O, Brijs B, Geenen L, Witvrouw A, Larmer F, Fischer F, Kronmuller S, Leca V, Otter B 2003 New low-stress PECVD poly-SiGe Layers for MEMS. *J. Microelectromech. Syst.* **12**, 816–25

Sawyer G A, Ishii T K 1969 Reflex klystron accelerometer and seismometer. *IEEE Trans. Ind. Electron. Control Instrum.* **IECI-16**, 103–6

Scheeper P, Gullov J O, Kofoed L M 1996 A piezoelectric triaxial acclerometer. *J. Micromech. Microeng.* **6**, 131–33

Schropfer G, Elflein W, de Labachelerie M, Porte H, Ballandras S 1998 Lateral optical accelerometer micromachined in (100) silicon with remote readout based on coherence modulation. *Sens. Actuators A Phys* **68**, 344–9

Seidel H, Fritsch U, Gottinger R, Schalk J, Walter J, Ambaum K 1995 A piezoresistive silicon accelerometer with monolithically integrated CMOS-circuitry. *Proc. 8th Int. Conf. Solid-State Sensors and Actuators (Transducers '95),* Stockholm, Sweden, pp. 597–600

Selvakumar A, Ayazi F, Najafi K 1996 A high sensitivity *z*-axis torsional silicon accelerometer. *Proc. 1996 International Electron Devices Meeting,* San Francisco, CA, USA, pp. 765–768

Seok S, Chun K 2006 Inertial-grade in-plane resonant silicon accelerometer. *Electron. Lett.* **42**, 1092–3

Seok S, Seong S, Lee B, Kim J, Chun K 2002 A high performance mixed micromachined differential resonant accelerometer. *Proc. 2002 IEEE Sensors Conference,* Orlando, Florida, pp. 1058–1063

Seshia A A, Palaniapan M, Roessig T A, Howe R T, Gooch R W, Schimert T R, Montague S 2002 A vacuum packaged surface micromachined resonant accelerometer. *J. Microelectromech. Syst.* **11**, 784–93

Shaw K A, Zhang Z L, MacDonald N C 1993 SCREAM I: A single mask, single-crystal silicon process for microelectromechanical structures. *Proc. 1993 IEEE Micro Electro Mechanical Systems (MEMS '93),* Fort Lauderdale, FL, USA, pp. 155–160

Shemansky F, Ristic Lj, Koury D, Joseph E 1995 A two-chip accelerometer system for automotive applications. *Microsyst. Technol.* **1**, 121–3

Sherman S J, Tsang W K, Core T A, Payne R S, Quinn D E, Chau K H L, Farash J A, Baum S K 1992 A low cost monolithic accelerometer; product/technology update. *Tech. Dig. 1992 Int. Electron Devices Meeting,* San Francisco, CA, USA, pp. 501–504

Shi J-J, Zhang W, Hao Y-L, Zeng Z-J 2004 Study to the integrated micro piezoresistive accelerometer for high *g* application with amplifying circuit. *Proc. 7th Int. Conf. Solid-State and Integrated Circuits Technology,* Beijing, China, pp. 1820–1823

Soref R A, McMahon D H 1984 Tilting-mirror fiber-optic accelerometer. *Appl. Opt.* **23**, 486–91

Sparacin D K, Spector S J, Kimerling L C 2005 Silicon waveguide sidewall smoothing by wet chemical oxidation. *J. Lightwave Technol.* **23**, 2455–61

Spillman W B 1982 Multimode fiber-optic accelerometer based on the photoelastic effect. *Appl. Opt.* **21**, 2653–4

Srinivasan U, Houston M R, Howe R T, Maboudian R 1998 Alkyltrichlorosilane-based self-assembled monolayer films for stiction reduction in silicon micromachines. *J. Microelectromech. Syst.* **7**, 252–60

ST Microelectronics 2004 LIS3L02AQ3, 2g/6g linear 3-axis accelerometer, data sheet

Stauffer J M 2006 Current capabilities of MEMS capacitive accelerometers in harsh environment. *Proc. 2006 IEEE/ION Position, Location, and Navigation Symposium (PLANS 2004),* San Diego, CA, USA, pp. 23–25

Stemme G 1991 Resonant silicon sensors. *J. Micromech. Microeng.* **1**, 113–25

Stewart R, Thede R, Couch P, Tarrant D 2004 High *G* MEMS accelerometer for Compact Kinetic Energy Missile (CKEM). *Proc. 2004 Position, Location, and Navigation Symposium (PLANS 2004),* Las Vegas, NV, pp. 20–25

Storgaard-Larsen T, Bouwstra S, Leistiko O 1996 Opto-mechanical accelerometer based on strain sensing by a Bragg grating in a planar waveguide. *Sens. Actuators A Phys* **52**, 25–32

Su S X P, Yang H S, Agogino A M 2005 A resonant accelerometer with two-stage microleverage mechanisms fabricated by SOI-MEMS technology. *IEEE Sensors J.* **5**, 1214–23

Sulouff R E Jr. 1991 Silicon sensors for automotive applications. *Dig. Tech. Papers 1991 Int. Conf. Solid-State Sensors and Actuators (Transducers '91),* San Francisco, California, pp. 170–176

Suminto J T 1991 A simple, high performance piezoresistive accelerometer. *Dig. Tech. Papers 1991 Int. Conf. Solid-State Sensors and Actuators (Transducers '91),* San Francisco, California, pp. 104–107

Szaniawski K, Napieralski A, Sekalski P, Podsiadly P 2004 Design of a prototype of a 3-axis capacitive acceleration sensor. *Proc. 24th Int. Conf. Microelectronics,* Nis, Serbia and Montenegro, pp. 219–222

Takao H, Matsumoto Y, Hee-Don S, Tanaka H, Ishida M, Nakamura T 1995 Three dimensional vector accelerometer using SOI structure for high temperature. *Proc. 1995 Int. Conf. Solid-State Sensors and Actuators (Transducers '95),* Stockholm, Sweden, pp. 683–686

Takao H, Matsumoto Y, Ishida M 1997 A monolithically integrated three axial accelerometer using stress sensitive CMOS differential amplifiers. *Proc. 1997 Int. Conf. Solid-State Sensors and Actuators (Transducers '97),* Chicago, IL, US, pp. 1173–1176

Takao H, Fukumoto H, Ishida M 2001 A CMOS integrated three-axis accelerometer fabricated with commercial submicrometer CMOS technology and bulk-micromachining. *IEEE Trans. Electron Devices* **48**, 1961–8

Takeda N 2000 Ball semiconductor technology and its application to MEMS. *Proc. 13th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2000),* Miyazaki, Japan

Tien N C 1998 Micro-optical inertial sensors using silicon MEMS. *Proc. 1998 IEEE Aerospace Conference,* Aspen, CO, USA, pp. 437–443

Toda M, Thompson M.L 2006 Contact-type vibration sensors using curved clamped PVDF film. *IEEE Sens. J.* **6**, 1170–77

Toda R, Takeda N, Murakoshi T, Nakamura S, Esashi M 2002 Electrostatically levitated spherical 3-axis accelerometer. *Proc. 15th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2002),* Las Vegas, NV, USA, pp. 710–713

Todd M D, Johnson G A, Althouse B A, Vohra S T 1998 Flexural beam-based fiber Bragg grating accelerometers. *IEEE Photonics Technol. Lett.* **10**, 1605–7

Uenishi Y, Tsugai M, Mehregany M 1994 Micro-opto-mechanical devices fabricated by anisotropic etching of (110) silicon. *Proc. 1994 IEEE Workshop Micro Electro Mechanical Systems (MEMS '94),* Oiso, Japan, pp. 319–324

Waltman S B, Kaiser W J 1989 Electron tunneling sensor. *Sens. Actuators* **19**, 201–10

Wang J, Zavracky P M, McGruer N E, Morrison R H 1997 Study of tunneling noise using surface micromachined tunneling tip devices. *Proc. 1997 Int. Conf. Solid-State Sensors and Actuators (Transducers '97),* Chicago, IL, US, pp. 467–470

Wang L P, Deng K, Zou L, Wolf R, Davis R J, Trolier-McKinstry S 2002 Microelectromechanical systems (MEMS) accelerometers using lead zirconate titanate thick films. *IEEE Electron Device Lett.* **23**, 182–4

Wang Z, Zong D, Lu D, Xiong B, Li X, Wang Y 2003 A silicon micromachined shock accelerometer with twin-mass-plate structure. *Sens. Actuators A Phys* **107**, 50–6

Warren K 1994 Navigation grade silicon accelerometer with sacrifically etched SIMOX and BESOI structure. *Tech. Dig. Solid-State Sensors and Actuators Workshop,* Hilton Head Island, SC, pp. 69–72

Warren K 1996 High performance silicon accelerometers with charge controlled rebalance electronics. *Proc. 1996 IEEE Position, Location, and Navigation Symposium,* Atlanta, GA, USA, pp. 27–30

Wu J, Fedder G K, Carley L R 2002 A low-noise low-offset chopper-stabilized capacitive-readout amplifier for CMOS MEMS accelerometers. *Dig. Tech. Papers 2002 IEEE Int. Solid-State Circuits Conference (ISSCC 2002),* San Francisco, CA, USA, pp. 428–429

Xie H, Fedder G K 2000 A CMOS z-axis capacitive accelerometer with comb-finger sensing. *Proc. 13th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2000),* Miyazaki, Japan, pp. 496–501

Xie H, Fedder G K 2002 Vertical comb-finger capacitive actuation and sensing for CMOS-MEMS. *Sens. Actuators A Phys* **95**, 212–21

Xie H, Erdmann L, Zhu X, Gabriel K J, Fedder G K 2002 Post-CMOS processing for high-aspect-ratio integrated silicon microstructures. *J. Microelectromech. Syst.* **11**, 93–101

Yazdi N, Najafi K 1997 An all-silicon single-wafer fabrication technology for precision microaccelerometers. *Proc. 1997 Int. Conf. Solid-State Sensors and Actuators (Transducers '97),* Chicago, IL, USA, pp. 1181–1184

Yazdi N, Salian A, Najafi K 1999 A high sensitivity capacitive microaccelerometer with a folded-electrode structure. *Proc. 12th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '99),* Orlando, FL, USA

Yazdi N, Najafi K 2000 An all-silicon single-wafer micro-*g* accelerometer with a combined surface and bulk micromachining process. *J. Microelectromechan. Syst.* **9**, 544–50

Yazdi N, Najafi K, Salian A S 2003 A high-sensitivity silicon accelerometer with a folded-electrode structure. *J. Microelectromech. Syst.* **12**, 479–486

Yazdi N, Kulah H, Najafi K 2004 Precision readout circuits for capacitive microaccelerometers. *Proc. 2004 IEEE Sensors Conference,* Vienna, Austria, pp. 28–31

Yeh C, Najafi K 1995 A low-voltage bulk-silicon tunneling-based microaccelerometer. *Proc. 1995 International Electron Devices Meeting (IEDM '95),* Washington DC, USA, pp. 593–596

Yeh C, Najafi K 1997 A low-voltage tunneling-based silicon microaccelerometer. *IEEE Trans. Electron Devices* **44**, 1875–82

Yeh C, Najafi K 1998 CMOS interface circuitry for a low-voltage micromachined tunneling accelerometer. *J. Microelectromech. Syst.* **7**, 6–15

Yun W, Howe R T, Gray P R 1992 Surface micromachined, digitally force-balanced accelerometer with integrated CMOS detection circuitry. *Tech. Dig. 5th IEEE Solid-State Sensor and Actuator Workshop,* Hilton Head Island, SC, USA, pp. 126–131

Yun C H, Brosnihan T J, Webster W A, Villarreal J 2005 Wafer level packaging of MEMS accelerometers with through wafer interconnects. *Proc. 55th IEEE Electronic Components Technology Conference,* Orlando, FL, USA, pp. 320–323

Zavracky P M, McClelland B, Warner K, Wang J, Hartley F, Dolgin B 1996 Design and process considerations for a tunneling tip accelerometer. *J. Micromech. Microeng.* **6**, 352–8

Zhang G, Xie H, de Rosset L E, Fedder G K 1999 A lateral capacitive CMOS accelerometer with structural curl compensation. *Proc. 12th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '99),* Orlando, FL, USA, pp. 606–611

Zhao Y, Cai Y 2005 *Z*-axis thermal accelerometer. In: No USPA (ed.) *U.S. Patent Application No. 20050274187*

Zhao Y, Hua Y 2004 Method of etching a deep trench in a substrate and method of fabricating on-chip devices and micro-machined structures using the same. *U.S. Patent No.6,712,983*

Zhu Y, Ping S, Chao L, Lacquet B M, Swart P L, Spammer S J 2003 Temperature-insensitive fiber Bragg grating accelerometer. *IEEE Photonics Technol. Lett.* **15**, 1437–9

Zou Q, Tan W, Kim E S, Loeb G E 2004 Highly symmetric tri-axis piezoelectric bimorph accelerometer. *Proc. 17th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2004),* Maastricht, Netherlands, pp. 197–200

Biographies

Huikai Xie is an Associate Professor at the Department of Electrical and Computer Engineering of the University of Florida. He received his MS in electrical engineering and computer science from Tufts University in 1998, and Ph.D. degree in electrical and computer engineering from Carnegie Mellon University in 2002. He also holds BS and MS degrees in electronics engineering from Beijing Institute of Technology. From 1992 to 1996, he was a research faculty member and lecturer at Tsinghua University, Beijing, China, working on various silicon-based chemical and mechanical sensors. He spent the summer of 2001 at Robert Bosch Corporation designing 6-DOF (degree of freedom) inertial measurement units. He received the 1996 Tsinghua University Motorola Education Award and the 1996 Best Paper Award from *Chinese Journal of Semiconductors*. He was also named the 2006 *Small Times*® magazine Best of Small Tech Researcher of the Year Finalist. He has published over 80 technical papers, and has two US patents granted and eight patents pending. He also serves in the technical program committees of several international conferences in the areas of MEMS and Photonics. His present research interests include micro/nanofabrication, integrated inertial sensors, microactuators, integrated power converters, optical MEMS, biomedical imaging, and fiberoptic sensors.

Mr. Robert E Sulouff is the Director of Business Development for the Micromachined products Division of Analog Devices Inc. (ADI), Cambridge, MA. He has led the start-up of the MEMS business area at ADI and is now addressing the longer-term strategic and business development opportunities of MEMS products. He joined ADI in 1993 as Product Line Director for Accelerometers and took the business from its first sales to the world's major market share producer of accelerometers and gyros. Before ADI, for seven years he was the Manager of Advanced Sensor Technology for Siemens Automotive. Preceding Siemens he was the Manager of Advance Sensors for Honeywell, Solid State Electronics Center. Mr. Sulouff was the VP of Engineering and Technology at Insouth Microsystems, a start-up which he co-founded and where he invented a MEMS accelerometer (G-Chip). He has over 35 published papers and two patents in accelerometer design. He is an active organizing member of several international technical conferences in MEMS including Hilton Head Sensors and Actuators. Educated at the University of Central Florida with BS and MS degrees in engineering materials with executive course work at MIT Sloan.

Gary K Fedder is the Howard M Wilkoff Professor of Electrical and Computer Engineering and the Director of the Institute for Complex Engineered Systems at Carnegie Mellon University (CMU). He joined CMU in 1994 with a joint appointment in ECE and the Robotics Institute. He received the BS and MS degrees in electrical engineering from MIT in 1982 and 1984, respectively. From 1984 to 1989, he worked at the Hewlett-Packard Company on circuit design and printed-circuit modeling. In 1994, he obtained the Ph.D. degree from UC-Berkeley, where his research resulted in the first demonstration of multimode control of an underdamped surface micromachined inertial device. He received the 1993 AIME Electronic Materials Society Ross Tucker Award, the 1996 Carnegie Institute of Technology GT Ladd Award, and the 1996 NSF CAREER Award. Currently, he serves as a subject editor for the *IEEE/ASME Journal of Microelectromechanical Systems*, on the editorial board of the IoP *Journal of Micromechanics and Microengineering* and as coeditor of the Wiley-VCH *Sensors Update* and *Advanced Micro- and Nanosystems* book series. He served as general cochair of the 2005 IEEE MEMS Conference. He has contributed to over 100 research publications and several patents in the MEMS area. His research interests include microsensor and microactuator design and modeling, integrated MEMS manufactured in CMOS processes, and structured design methodologies for MEMS.

2.06 Vibrating Gyroscopes

Farrokh Ayazi, Mohammad Faisal Zaman, and Ajit Sharma, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Published by Elsevier B.V.

Glossary

DARPA Defense Advanced Research Projects Agency

IEEE Institute of Electrical and Electronics Engineers

2.06.1 Introduction

Gyroscope is a sensor that measures the rate or the angle of rotation. Micromachined gyroscopes constitute one of the fastest growing segments of the microsensor market. The application domain of these devices is rapidly expanding from automotive to consumer electronics industry and personal navigation systems. A multitude of applications exist in the automotive sector including navigation, antiskid and safety systems, rollover detection, and next-generation airbag and antilock brake systems. Consumer electronics applications include image stabilization in digital cameras and smart user interfaces in handhelds, gaming, and inertial pointing devices.

Miniature gyroscopes can also be used for navigation purposes. Inertial navigation is the process of determining the position of a body in space by using the measurements provided by accelerometers and gyroscopes installed on the body. Inertial measurement units (IMUs) are vital components in aircraft, unmanned aerial vehicles, global positioning system (GPS)-augmented navigation, and personal heading references. An IMU typically uses three accelerometers and three gyroscopes placed orthogonally along the coordinate axes to gather information about an object's direction and heading. The components of acceleration and rotation rate can consequently be interpreted to yield the object's accurate position in space. An IMU is self-contained and can aid in accurate short-term navigation of a craft or an object in the absence of GPS-assisted inertial navigation (Lawrence 1993).

The applications space of microgyroscopes can be divided based on performance requirements. Table 1 summarizes the performance requirements for rate-grade and navigation-grade gyroscopes. The majority of automotive and consumer electronics applications require rate-grade performance (as indicated in Table 1), whereas higher-precision navigation-grade devices are suitable for IMUs and high-end applications in aerospace and petroleum industry. The petroleum industry uses gyroscopic sensors for real-time monitoring and correction of drilling in offshore rigs. Guidance systems and platform stabilization of missiles are but a few of the military applications that require accurate angular measurements.

Table 1 Performance requirements for rate-grade and navigation-grade gyroscopes

	Rate grade	*Navigation grade*
Angle random walk ($\deg h^{-1/2}$)	>0.5	<0.005
Bias drift ($\deg h^{-1}$)	10–1000	<0.1
Scale factor accuracy	0.1–1%	<10 ppm
Full scale range ($\deg s^{-1}$)	50–1000	>500
Shock level (g)	1000	10^3–10^4
Bandwidth (Hz)	>70	~10–100

Micromachined gyroscopes are projected to become practical alternatives to expensive and bulky conventional inertial sensors in the near future. Current high-performance rotation sensors including precision fiber-optic gyroscopes, ring laser gyroscopes, and conventional rotating wheel gyroscopes are too expensive and/or too large for use in most emerging applications. With micromachining processes allowing mass production of micromechanical systems on a chip together with their control and signal conditioning electronics, low-cost and micromachined silicon gyroscopes will provide high-accuracy rotation measurements. Moreover, advances in the fabrication techniques allow the detection and control electronics to be integrated on the same silicon chip together with the mechanical sensor elements. Thus, miniaturization of vibratory gyroscopes through innovative microfabrication processes and gyroscope designs is expected to become a very attractive solution to current inertial sensing market needs as well as to open new market opportunities.

In the past decade, much of the effort in developing micromachined gyroscope has concentrated on rate-grade devices driven by investments from the automotive industry. These applications require a resolution of about $0.1°\deg s^{-1}$ in a 50-Hz bandwidth. Small form factor, lightweight, and low power consumption make micromachined microgyros ideal for use in handheld applications, many of which requiring multiaxis devices. As mentioned earlier, short-range navigation and guidance require resolution and bias drift properties approaching or better than $0.1°\deg h^{-1}$. Such performances have yet to be demonstrated by current silicon micromachined gyroscopes. However, in recent years, several Defense Advanced Research Projects Agency (DARPA)-sponsored initiatives have been made to examine the feasibility of navigation-grade performance in micromachined silicon gyroscopes.

Although bulk mechanical and optical gyroscope technologies have successfully demonstrated navigation-grade performance, their constituent components

are difficult to miniaturize and develop on silicon substrate using conventional CMOS fabrication technique. Hence, microelectromechanical systems (MEMS) research has almost exclusively focused on the third class of devices referred to as vibratory gyroscopes.

2.06.2 Coriolis Vibratory Gyroscopes

Vibratory gyroscopes use vibrating mechanical elements to sense rotation. They have no rotating parts that require bearings and can easily be miniaturized and batch fabricated using micromachining techniques. Vibratory gyroscopes are based on the transfer of energy between two vibration modes of a structure induced by Coriolis acceleration. Coriolis acceleration, named after the French scientist and engineer Gaspard Gustave de Coriolis (1792–1843), is an apparent acceleration that arises in a rotating reference frame and is proportional to the rate of rotation. The Coriolis effect may be illustrated with an example of a flying object as viewed through the eyes of an observer located on a rotating reference frame as shown in Figure 1. The observer located along the x-axis of the xyz coordinate system views an object traveling above him/her in space with a certain velocity $\mathbf{v}$. If the observer's reference frame rotates normally (about the z-axis) with an angular velocity $\vec{\Omega}$, to the observer the object appears to move toward the x-axis with an acceleration $2\vec{v} \times \vec{\Omega}$. Although no real force has been exerted on the flying object, to the observer located on the rotating reference frame, an apparent force has resulted which is directly proportional to the rate of rotation.

2.06.2.1 Coriolis Acceleration

To better understand Coriolis acceleration, it is often helpful to study it analytically. In Figure 1(b), frames A and B are the inertial and the rotating reference frames, respectively. The component $\vec{r}_A$ and $\vec{r}_B$ are position vectors relative to the inertial and rotating frames. Finally, $\vec{\theta}$ and $\vec{R}$ refer to the orientation and the position of the rotating frame B, respectively.

The time derivative of any vector $\vec{r}$, which is defined in frames A and B, is indicated by the transport theorem (Haim 1998) as

$$\frac{d}{dt}\overrightarrow{r_A(t)} = \frac{d}{dt}\overrightarrow{r_B(t)} + \frac{d}{dt}\theta \times \overrightarrow{r_B(t)} \quad [1]$$

Taking the second time derivative of the position vector, the acceleration of a body moving with the rotating reference frame may be calculated as

$$\overrightarrow{r_A(t)} = \overrightarrow{R(t)} + \overrightarrow{r_B(t)} \quad [2]$$

$$\frac{d}{dt}\overrightarrow{r_A(t)} = \frac{d}{dt}\overrightarrow{R(t)} + \frac{d}{dt}\overrightarrow{r_B(t)} + \frac{d}{dt}\theta \times \overrightarrow{r_B(t)} \quad [3]$$

$$\begin{aligned}\frac{d^2}{dt^2}\overrightarrow{r_A(t)} = {} & \frac{d^2}{dt^2}\overrightarrow{R(t)} + \frac{d^2}{dt^2}\overrightarrow{r_B(t)} + \frac{d}{dt}\theta \times \frac{d}{dt}\overrightarrow{r_B(t)} + \frac{d}{dt}\theta \\ & \times \left(\frac{d}{dt}\theta \times \overrightarrow{r_B(t)}\right) + \frac{d^2}{dt^2}\theta \times \overrightarrow{r_B(t)} \\ & + \frac{d}{dt}\theta \times \overrightarrow{r_B(t)}\end{aligned} \quad [4]$$

By replacing the first and second time derivatives of the reference frames with velocity and acceleration components, respectively, the acceleration expression reduces to

$$\vec{a}_A = \vec{A} + \vec{a}_B + \Omega \times \vec{r}_B + \Omega \times \left(\Omega \times \vec{r}_B\right) + 2\Omega \times \vec{v}_B \quad [5]$$

The term $2\vec{\Omega} \times \vec{v}_B$ is the Coriolis term. The vibratory rate gyroscope can essentially be viewed as an accelerometer measuring the Coriolis acceleration to

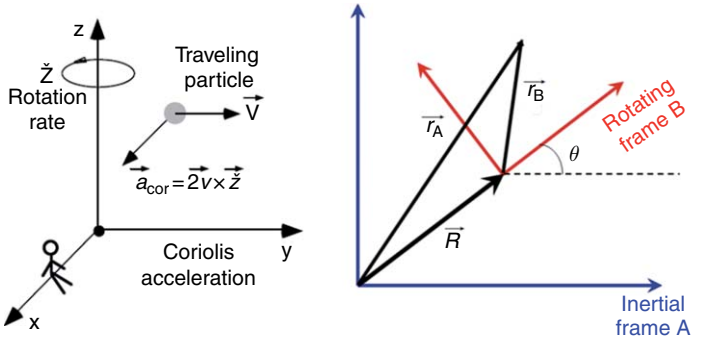


Figure 1 (a) Visualization of the Coriolis effect and (b) derivation of Coriolis acceleration.

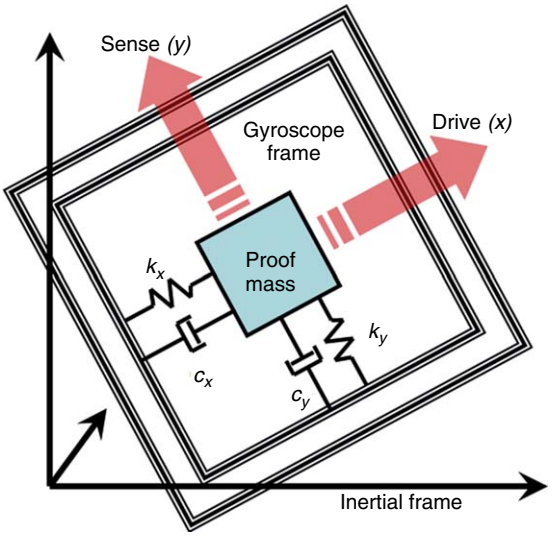


Figure 2 Illustration of a rotating gyroscope frame along an inertial frame.

calculate the rotation rate. Using **Figure 2**, we examine the equation of motion on a single proof mass subjected to rotation about an inertial reference frame.

$$\vec{F} = M\left(\vec{A} + \vec{a}_B + \Omega \times \vec{r}_B + \Omega \times (\Omega \times \vec{r}_B) + 2\Omega \times \vec{v}_B\right) \quad [6]$$

In eqn [6] $\vec{A}$ and $\vec{\Omega}$ are the linear acceleration and angular velocity of the rotating gyroscope frame, respectively, while $\vec{v}_B$ and $\vec{a}_B$ are the velocity and acceleration vectors of the gyroscope proof mass with respect to the inertial reference frame, respectively. When the proof mass motion is decomposed along its two principal vibratory axes and linear accelerations are canceled out, the two equations of motions are expressed as (Acar and Shkel 2005)

$$\begin{aligned} m\frac{d^2}{dt^2}x + c_x\frac{d}{dt}x + \left(k_x - M\left(\Omega_y^2 + \Omega_z^2\right)\right)x \\ + M\left(\Omega_x\Omega_y + \frac{d}{dt}\Omega_z\right)y \\ = \tau_x + 2M\Omega_z\frac{d}{dt}y \end{aligned} \quad [7]$$

$$\begin{aligned} m\frac{d^2}{dt^2}y + c_y\frac{d}{dt}y + \left(k_y - M\Omega_x^2 + \Omega_z^2\right)y \\ + M\left(\Omega_x\Omega_y + \frac{d}{dt}\Omega_z\right)x \\ = \tau_y + 2M\Omega_z\frac{d}{dt}x \end{aligned} \quad [8]$$

For a constant angular rate, and neglecting the minimal contributions of second-derivative components of off-axis angular acceleration, the simplified 2-degrees of freedom equations of motion for the gyroscope proof mass may be expressed as follows:

$$m\frac{d^2}{dt^2}x + c_x\frac{d}{dt}x + k_x x = \tau_x + 2M\Omega_z\frac{d}{dt}y \quad [9]$$

$$m\frac{d^2}{dt^2}y + c_y\frac{d}{dt}y + k_y y = \tau_y + 2M\Omega_z\frac{d}{dt}x \quad [10]$$

The terms $2M\Omega_z\frac{d}{dt}x$ and $2M\Omega_z\frac{d}{dt}y$ are the rotation induced and causes dynamic coupling between the vibratory axes and are utilized for angular rate measurements in vibratory gyroscopes.

As can be seen from the previous discussion, it is necessary to devise a mechanism to excite the mass into its reference vibration. A second mechanism that must be implemented is a scheme to pick off the Coriolis-induced signal. Various schemes have been used to implement the drive and the pickoff mechanisms for Coriolis microgyroscopes. The reference drive excitation can be achieved using piezoelectric, capacitive electrostatic, or electromagnetic actuation schemes. Similarly, the Coriolis signal can be picked off using piezoresistive or capacitive means. The subject matter of the following sections refers mainly to the case of capacitive, silicon vibratory Coriolis microgyroscopes.

2.06.2.2 Classifications of Vibratory Gyroscopes

Vibratory gyroscopes are based on the transfer of energy between two vibration modes of a structure and can operate in either matched-mode or split-mode condition. Under matched-mode condition, the sense mode is designed to have the same (or nearly the same) resonant frequency as the drive mode. Hence, the rotation-induced Coriolis signal is amplified by the mechanical quality factor (Q) of the sense mode. In split-mode condition, the drive and the sense modes are separated in resonant frequency. The sense mode is then a controlled mode that operates similar to an accelerometer and measures the Coriolis acceleration. Owing to Q-amplification, gyroscopes operated under matched-mode configuration offer higher sensitivity and better resolution. Resonant matched devices are broadly classified into two types depending on the nature of their operating modes (**Figure 3**). Type I devices rely on nondegenerate vibration modes for driving and sensing. Tuning fork and frame gyroscopes are examples of type I gyroscopes. Type II devices on the contrary function with

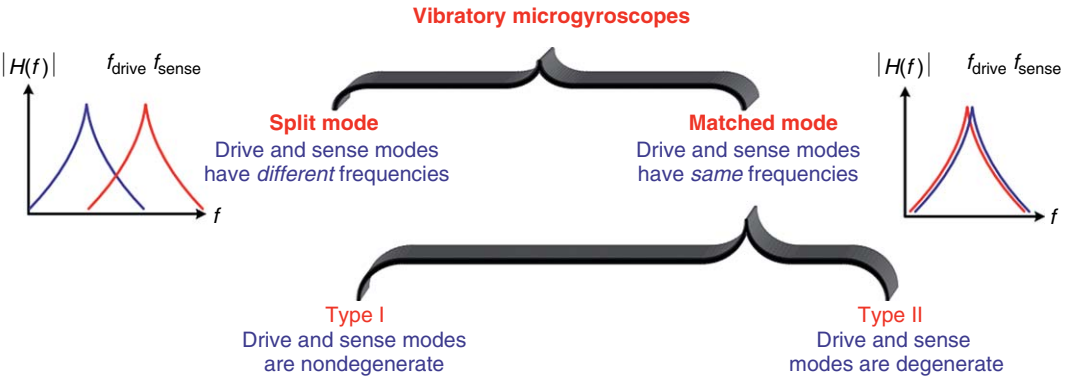


Figure 3 Classification of gyroscopes based on operational modes.

degenerate vibration modes and are invariably easier to match and operate under matched condition. Shell-type gyroscopes are examples of type II gyroscopes.

2.06.2.3 Gyroscope Performance Parameters

2.06.2.3.1 Resolution

The resolution of a gyroscope refers to the minimum rotation rate that can be distinguished from the noise floor of the gyroscope system. The overall resolution of the micromachined vibrating silicon gyroscope is determined by two uncorrelated components: the mechanical (or Brownian) and the electronic noise floors.

$$\Omega(\text{Overall}) = \sqrt{\Omega^2(\text{Brownian}) + \Omega^2(\text{Electronic})} \quad [11]$$

Brownian motion of the structure caused by molecular collisions from surrounding medium represents the mechanical noise component of any vibratory gyroscope (Yazdi *et al.* 1998). By equating the displacement caused by Brownian motion to the displacement induced by Coriolis acceleration, one may derive the mechanical resolution of a vibratory microgyroscope as given by eqn [12]. The quantities k_B, T, and BW represent the Boltzmann constant (1.38e–23 J K^{-1}), operating temperature (K), and measurement bandwidth (Hz), respectively. In the expression, ω_0 is the resonant frequency of the sensor, M its mass, q_{drive} the amplitude of vibration along the reference axis, and Q_{EFF} the effective quality factor of the system. For a given mechanical structure, Q_{EFF} can be maximized by matching the frequencies of the drive and the sense modes, as will be explained in a subsequent section.

$$\Omega(\text{Brownian}) \propto \frac{1}{q_{\text{drive}}} \sqrt{\frac{4k_B T}{\omega_0 M Q_{\text{EFF}}}} \sqrt{BW} \quad [12]$$

Evaluating the electronic noise floor is dependent on the interface electronics architecture. However, assuming the noise at the output V_n has a white spectrum near the operating frequency, the minimum detectable rotation rate limited by electronic noise can in general be simplified as follows (for a capacitive detection system):

$$\Omega(\text{Electronic}) \propto \frac{g_0^2}{V_P} \frac{\omega_0 C_{\text{parasitics}} V_n}{q_{\text{drive}} Q_{\text{EFF}}} \sqrt{BW} \quad [13]$$

In eqn [13], g_0 and V_P represent the rest sense gap and the polarization voltage applied to the substrate, respectively. $C_{\text{parasitics}}$ is the total parasitic capacitance between the sensor and the electronic front-end and that of pads, bond wires, and circuit gate. It is evident from both equations that drive amplitude, device mass, and sense gaps play a key role in determining the overall noise floor in a vibratory gyroscope system. Resolution is expressed in units of noise floor (deg h^{-1} Hz$^{-1/2}$) or in terms of angle random walk (ARW) (deg h$^{-1/2}$) (Ayazi 2000).

2.06.2.4 Sensitivity and Scale Factor

From eqn [10], the motion along the sensitive axis (y) in response to a normal rotation rate (Ω_z) may be expressed as follows:

$$y = \frac{2\Omega_z \omega_x q_{\text{drive}}}{\omega_y^2} \quad [14]$$

In this expression, ω_x and ω_y are the drive and the sense mode frequencies and q_{drive} is the displacement along the driven axis. It is quite evident that increasing q_{drive} has beneficial attributes beyond device resolution. As was with the case of the noise floor, for a given mechanical sensor, the sensitivity can be maximized by maximizing Q_{EFF}. The Coriolis-induced displacement produces a capacitance change that can be converted into a voltage using an electronic front-end. The sensitivity to (volts.degree^{-1}.second^{-1}) of the microgyroscope can then be expressed as

$$\text{Sensitivity} \propto \frac{2V_{\text{P}}}{C_{S0} + C_{\text{parasitics}}} \frac{C_{S0}}{g_0} \frac{Q_{\text{EFF}} q_{\text{drive}}}{\omega_0} \qquad [15]$$

Sense mode oscillations are typically detected through capacitive, piezoresistive, piezoelectric, or optical means. Rate sensitivity or scale factor in vibratory gyroscope is most commonly expressed in units of volts per degree per second. Another aspect to be considered with regard to scale factor is the linearity. This is a measure of the deviation from the normalized sensitivity line.

2.06.2.4.1 Bias stability

An important measure of long-term stability of a gyroscope is its bias drift. Very similar to offset in circuits, the bias drift of a gyroscope is composed of systematic and random components. The systematic components arise due to temperature variations, linear accelerations, vibrations, and other environmental factors (Institute of Electrical and Electronics Engineers (IEEE) 1997). Empirical studies have tried to formulate an expression that can predict drift in gyroscopes (Watson and Henke 2002), but considerable research remains to be done.

The random component of bias drift has a $1/f$ noise characteristic and is related to the ARW of the device, which in turn depends on the system noise floor. The bias stability of the gyroscope measures the drift of the output offset over time. The drift can be measured in several ways, but the most commonly used means at present is the Allan variance technique (IEEE 1997). A sample of this analysis will be examined during the case study of the matched-mode tuning fork gyroscope (M^2-TFG). The bias drift is usually expressed in units of degrees per hour.

The bias drift of a microgyroscope becomes important when the fact that it is used along with accelerometers to obtain heading information, i.e., position and orientation, is considered. The rate information from the microgyroscope is integrated to obtain the angle. Any long-term variations in the rate (of which the bias drift is an accurate indicator) add up and can cause a large error in angle and orientation information. Modern gyroscope systems periodically calibrate themselves with GPS to ensure that the heading information is accurate. The longer a system can function accurately without the need for calibration, the better and more accurate heading and orientation information it can deliver. A long interval between calibration sequences is crucial for applications such as deep-sea navigation and oil exploration where it is not possible to resurface very easily to calibrate with GPS.

2.06.2.4.2 Bandwidth and dynamic range

The bandwidth of the gyroscope system determines the response time of the system. This is the time required for the output to settle within a certain range of the expected value for an input step function.

The final performance criterion regularly used in specifying gyroscope performance is the dynamic range. This refers to the range of input values over which the output is detectable. Typically, it is computed as the ratio between the maximum input rotation rate (full scale rate) that the sensor can tolerate and the system noise floor.

2.06.2.5 Gyroscope Interface Electronics

As micromachined silicon gyroscopes scale aggressively to meet higher performance parameters, the interface electronics that actuate, sense, and control these micromechanical structures play a vital role in determining the overall performance of the microgyroscope system. **Figure 4** shows a schematic overview of the constituent system blocks in a typical microgyroscope system. Based on functionality, the electronics can be divided into the following subsystems:

(1) Drive oscillator: Sets up the reference vibration of the structure/proof-mass along the reference axis.
(2) Sense channel: Extracts the rotation-induced Coriolis signal along the sensitive axis.
(3) Quadrature nulling scheme: Measures and minimizes quadrature error in the sense channel.
(4) Automatic frequency tuning: Eliminates the frequency separation between the drive and sense resonant modes.
(5) Self-test and trimming: Calibrates the sensor periodically.

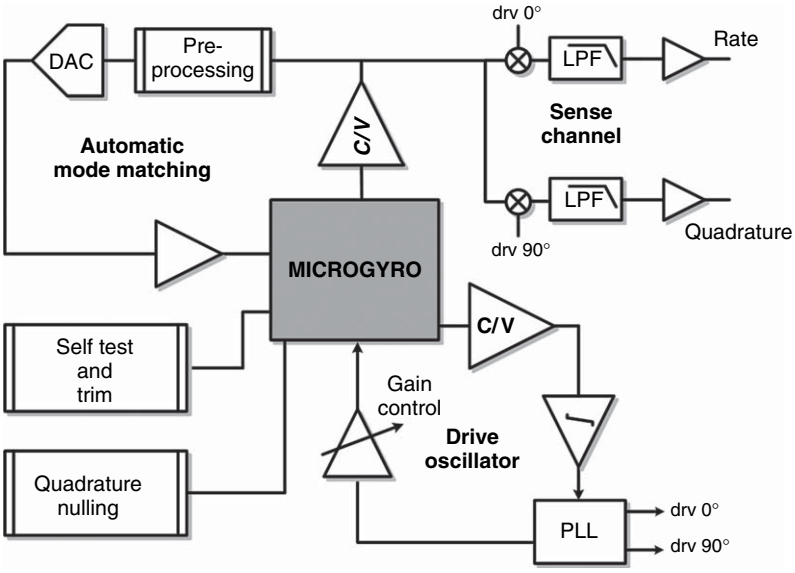


Figure 4 Overview of a typical micromachined vibratory gyroscope system.

2.06.2.5.1 Drive loop

The drive loop electronics is responsible for starting and sustaining oscillations along the reference axis at a constant amplitude. It is essential that a constant drive amplitude be maintained, as any variation in the drive amplitude manifests itself as a change in the velocity of the mechanical structure (along the driven axis). Velocity fluctuations modulate the sensor output and can result in a false or an inaccurate rate output. The drive loop uses an automatic level control (ALC) circuit to achieve and maintain a constant drive amplitude. Typically, there are two approaches to implement the drive loop:

(1) An electromechanical oscillator approach
(2) A phase-locked loop (PLL)-based approach

A high Q for the drive resonant mode can significantly ease the design of the drive oscillator. A large drive-Q enables the oscillations to be built up and sustained using much smaller voltage levels. Because the vibratory microgyroscope is a mass-spring system, it can be modeled using second-order dynamics. The drive mode of the microgyroscope can be modeled in a fashion similar to that for a micromachined resonator (Pourkamali *et al.* 2004) using a series RLC circuit. While the L (inductance) and C (capacitance) components determine the modal frequency and get canceled at resonance, the motional impedance, R, represents a loss mechanism. The loss due to this motional impedance must be compensated for by an electromechanical oscillator. A large drive-Q implies low motional impedance. A low motional impedance drive loop allows for reduced voltages and alleviates gain requirements for the sustaining electronics.

The microgyroscope can be driven to oscillation by placing it in a feedback loop satisfying the Barkhausens criterion, i.e., maintaining unity gain and zero phase shift. The microgyroscope is the frequency-determining element in this loop, as shown in **Figure 5(a)**. The front-end is a current-to-voltage (I-V) converter cascaded with a phase-shifting buffer. In **Figure 5(b)**, the reference drive vibrations are setup using a PLL. A PLL locks on to the output of the front-end I-V converter. The PLL output is amplified or attenuated to achieve the desired voltage amplitude and used to drive the microgyroscope. Alternately, a variable gain amplifier can be used to implement the ALC.

The electromechanical oscillator approach is more attractive than the PLL approach as it has lower power dissipation. However, the mechanical motional impedance of the drive resonator can be on the order of a few megaohms in certain vibratory structures and thereby limit the use of this approach.

2.06.2.5.2 Sense channel

The rotation-induced Coriolis force is an amplitude-modulated (AM) signal where the carrier frequency is the drive oscillation frequency and the input rotation rate modulates the amplitude. The resulting Coriolis acceleration is a dual-sideband signal centered at the drive oscillation frequency. Synchronous

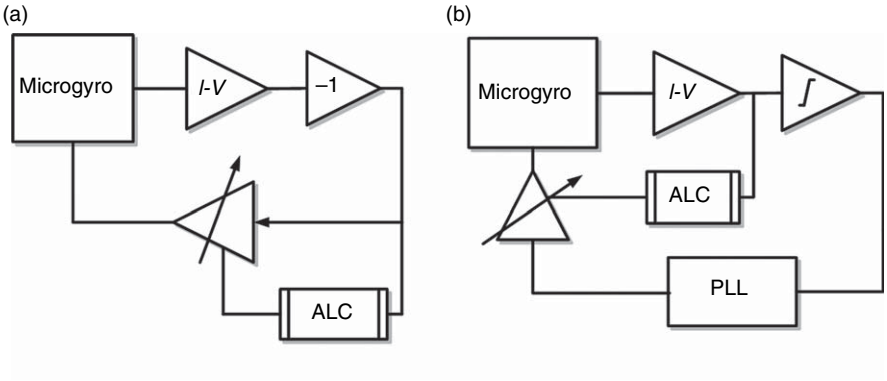


Figure 5 Drive loop implemented using (a) electromechanical oscillator and (b) phase-locked loop (PLL). ALC, automatic level control.

demodulation widely used in AM communication receivers can be utilized for extracting the rate information from the microgyroscope. The AM Coriolis output is demodulated using the drive oscillator signal. This phase-sensitive demodulation allows for rejection of the interfering mechanically generated quadrature error. The low-pass-filtered baseband signal is proportional to input rotation rate and may be amplified if necessary at a later stage. Because synchronous demodulation allows for phase-sensitive detection and therefore rejection of quadrature error, it is preferred to other techniques such as envelope detection.

The primary challenge in implementing the sense channel is to design a low-noise front-end for I-V conversion. The Coriolis-induced displacement of the microgyroscope along the sense axis is on the order of a few angstroms. For resolving such minute rate signals, the front-end noise floor and therefore the minimum detectable capacitance change must be minimized to the order of a few attofarads. The task of designing circuits that can accurately measure such capacitance changes is made all the more difficult by the presence of large parasitic capacitances. In a typical two-chip solution (where the IC and device are on separate substrates), the input parasitic capacitances are on the order of 1 to 10 pF.

As summarized in **Table 1**, commercial rate-grade applications require noise floors on the order of 10 to 100 deg h^{-1} and should typically be able to sense rotation rates as large as ± 500 deg s^{-1}. Navigation-grade gyroscopes typically have similar full-scale range, but the noise floor specifications are sub-0.1 deg h^{-1}. This is roughly two orders of magnitude lower than their commercial counterparts and calls for the sensor to have a large dynamic range. While mechanical structures can attain dynamic ranges greater than 120 dB, designing front-end electronics with such a large dynamic range is challenging. The dynamic range is limited by the supply voltage on the one hand and the noise floor on the other. With the continued scaling of CMOS process to narrower channel lengths, flicker noise in transistors and the reduced voltage headroom make the achievement of dynamic ranges greater than 100 dB extremely challenging.

There are a number of approaches for Coriolis rate detection (Boser 1997). Quartz rate sensors (QRSs) use piezoresistive techniques (Voss *et al.* 1997), while voltage buffers have also been commonly used (Ayazi and Najafi 2001, Putty and Najafi 1994) as front-ends for capacitive microgyroscopes. Transimpedance front-ends are the most promising candidates for tactical-grade systems as they offer low current noise and therefore subattofarad capacitance resolutions (Saukoski *et al.* 2005, Sharma *et al.* 2006). Because the output of the microgyroscope is a motion-induced current, charge amplifier integrators are preferred to voltage buffers as they allow a trade-off between gain and bandwidth (Clark *et al.* 1996).

2.06.2.5.3 Quadrature nulling

There exists a large error signal common to all vibratory rate gyroscopes that is caused by imbalances in the mechanical elements due to fabrication imperfections (Clark *et al.* 1996). These imbalances cause the proof mass to vibrate at an angle of θ_Q from the actual drive axis. This subsequently results in mechanical coupling of the drive resonator motion to the sensitive axis and an oscillatory deflection that is measured along the sense axis with the

Coriolis-induced deflections. Both the quadrature error and the Coriolis deflection are AM signals centered at the drive resonant frequency. The only distinguishing feature between the two signals is that they have a relative phase difference of 90°, and hence, the origin of the term quadrature error. The 90° phase difference is because the Coriolis acceleration is sensitive to proof mass velocity along the drive axis, while the quadrature error is proportional to the proof mass position along the drive axis.

The ratio of quadrature-induced displacement to the Coriolis-induced displacement at matched mode is expressed as

$$\left.\frac{\mathcal{Y}_{\text{Quadrature}}}{\mathcal{Y}_{\text{Coriolis}}}\right|_{\text{Matched-Mode}} \approx \frac{\omega_0 \cdot \theta_Q}{2\Omega_z} \qquad [16]$$

Eqn [16] indicates that even a small angle θ_Q can cause a significant quadrature error, much greater than the Coriolis-induced displacement. Micromachining silicon technology has much lower mechanical tolerances as compared with advanced CMOS processing. A 1.7-μm-wide flexure will typically be defined to no better than 0.2 μm (Geen *et al.* 2002). Consequently, the majority of the quadrature error results from off-diagonal elements in the spring stiffness matrix as shown in **Figure 6**. For this reason, reducing quadrature error is best achieved by ensuring that the spring stiffness matrix is diagonalized by applying specific quadrature nulling techniques.

Quadrature nulling of any gyroscopic sensor depends significantly on the nature of the mechanical structure. A number of techniques have been used to null quadrature error. Some of the earliest works (Barnaby and Morrow 1956, Barnaby and Reinhardt 1951) involved trimming and bucking to control quadrature error, but these techniques did not track over temperature and life. Other techniques involved servomechanisms whereby a force was applied to the mass so as to null any displacement of the proof mass that is in phase with position (Clark *et al.* 1996). The surface micromachined gyroscopes developed by researchers at HSG-IMIT are based on the decoupled angular velocity detector (DAVED) approach (Geiger *et al.* 2001). The technique has been shown to reduce quadrature error by a factor of 5, by the use of two types of one-dimensional springs. The first restricts the drive resonator motion to one dimension, while the second restricts the motion of the sensing mass relative to the driven mass to one dimension as well. The Analog Devices gyroscope uses a set of optimally designed levers (Geen *et al.* 2002) to reduce quadrature error to less than 1 ppm by improving the selectivity of the suspension flexures. Yet another technique that has been implemented is the use of torque cancelation electrodes (Putty and Najafi 1994, Sharma *et al.* 2007) to correct for any misalignment between the drive and the sense axes.

2.06.2.5.4 Self-test and trim

To be viable in a large-scale manufacturing environment, the microgyroscope systems must be equipped with a self-test capability to ensure quality and reliability. Apart from significantly reducing test times at the production facility, these systems allow for calibration of the sensor in the field. Currently, most gyroscope systems use postfabrication trimming to a certain extent to account for microfabrication imperfections. With the development of more advanced lithographic tools and etching systems, the use of these trimming procedures can be minimized.

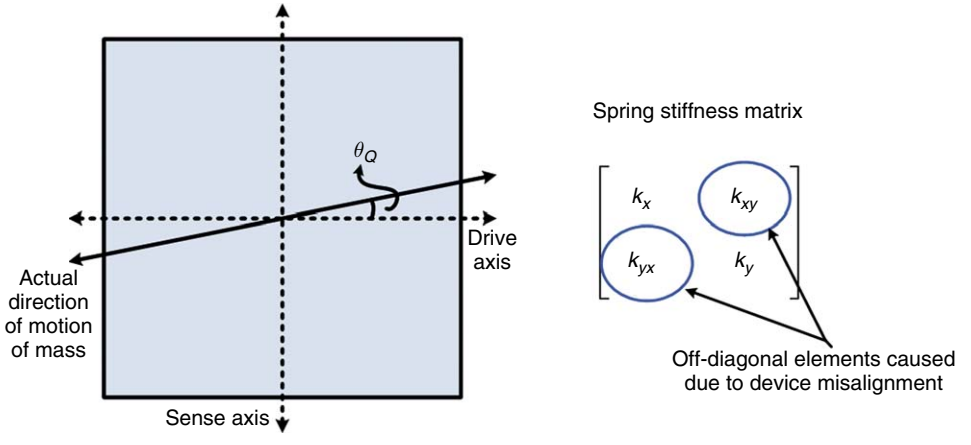


Figure 6 Demonstration of quadrature error in a proof mass of a vibratory gyroscope.

2.06.2.5.5 Advantages of mode matching

The system noise floor and the sensitivity of the microgyroscope depend on the Q_{EFF} of the system. The microgyroscope is a resonant MEMS sensor, and the Q_{EFF} represents the mechanical gain of the sensor. From eqns [11] through [13], it is evident that the noise floor and the sensitivity are maximized by maximizing Q_{EFF}. Q_{EFF} depends on the inherent mechanical quality factors of the structure as well as on the frequency split between the drive and sense modes. For a given mechanical structure, Q_{EFF} can be maximized by equalizing the drive and the sense resonant mode frequencies, i.e., drive resonant frequency (ω_x) = sense resonant frequency $(\omega_y) = \omega_0$. When the drive and the sense frequencies are matched, there is maximum energy coupling between the drive and the sense modes and therefore maximum sensitivity.

Mode matching is therefore essential for leveraging the maximum mechanical gain (Q_{EFF}) of the sensor. Empirical studies have suggested eqn [17] for the bias drift of gyros (Watson and Henke 2002), once again highlighting the need for mode matching.

$$N_B \frac{\omega_o^2}{Q_{drive} Q_{sense} Area_{electrodes}} \qquad [17]$$

Mode matching, however, is not without its own problems. First, Q-amplification increases the sensitivity of the gyro to the Coriolis-induced acceleration (by Q_{sense}). However, any spurious displacements in the y-direction, such as quadrature error, are also amplified by the same mechanical gain Q_{sense}. Second, after the modes have been matched, it is necessary to continuously monitor the modes and maintain matching. This demands continuous-time (CT) monitoring and correction of any mismatch that can occur as a result of environmental changes, shock, etc. Finally, from a commercial point of view, an automatic mode-matching circuit is more cost-effective as it can replace any expensive trimming that needs to be done. Hence, a robust technique and a feasible CMOS-compatible implementation that can achieve automatic mode matching must be developed.

2.06.2.5.6 Automatic mode matching and quadrature revisited

Mode matching of a gyroscope is closely related to the quadrature error problem. Quadrature error in vibratory gyroscopes has been recognized as the single most important factor that precludes perfect mode matching (i.e., zero frequency split between the drive and the sense resonant frequencies) (Clark *et al.* 1996).

It must be noted that a matched-mode device is one in which the drive and the sense modes are eigenmodes – i.e., they are degenerate. The quadrature-induced off-diagonal terms in the spring stiffness matrices prevent the eigenmodes to be degenerate, leading consequently to a mismatch in the resonant frequencies. In order to achieve mode degeneracy, it is necessary that these terms be electronically eliminated and the spring stiffness matrix be diagonalized. **Figure 7** shows how the K-matrix can be made symmetric by applying quadrature cancelation voltages to eliminate the off-diagonal terms. Hence, quadrature cancelation is essential to attain and maintain mode matching. Once the quadrature error has been nulled, the drive and the sense stiffness coefficients can be made equal to ensure that the drive and the sense frequencies are equal. This is achieved through electrostatic spring softening.

An automatic mode-matching system must be able to match the frequencies of the drive and the sense modes. There has been considerable work in this respect as such a system would be invaluable in applications where frequent calibration of the gyroscope is not possible, for instance, in IMUs for aircraft, unmanned surveillance vehicles, etc. Boeing and Jet Propulsion Laboratory (JPL) have implemented a mode-matching ASIC (Keymeulen *et al.* 2005) that uses evolutionary computation to achieve mode matching. The use of a dual PLL approach (Sung *et al.* 2006) has also been suggested to implement automatic mode matching.

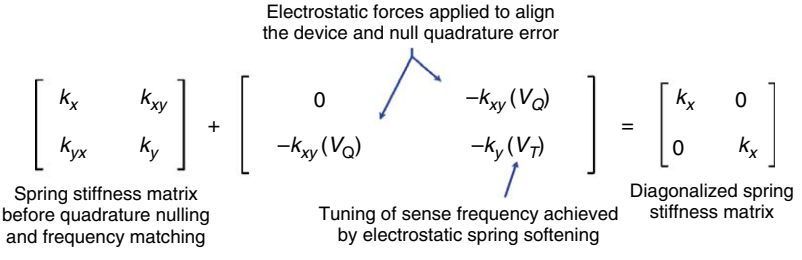


Figure 7 Application of a balancing voltage explicitly cancels off the off-diagonal terms.

2.06.3 Overview of Silicon Micromachined Vibratory Gyroscopes

The first batch-fabricated silicon micromachined rate gyroscope was demonstrated by the Charles Stark Draper Laboratory in 1991. This p-doped silicon bulk mechanical device was a double-gimbal vibratory gyroscope supported by torsional flexures (Greiff *et al.* 1991). In this design, an outer gimbal was electrostatically excited at a constant amplitude using drive electrodes. This oscillatory motion was transferred to the inner gimbal along the stiff axis of the inner flexures. When exposed to a rotation input normal to the plane of the device, Coriolis force causes the inner gimbal to oscillate about its sensitive axis. A rotation rate resolution of $4\,\mathrm{deg\,s^{-1}}$ was realized using this structure.

Two years later, in 1993, the Charles Stark Draper Laboratory also reported a silicon-on-glass tuning fork gyroscope (**Figure 8**) (Bernstein *et al.* 1993) fabricated using the dissolved wafer process (Gianchandani and Najafi 1992). This gyroscope was electrostatically vibrated in its plane using a set of interdigitated comb drives to achieve drive amplitudes as large as 10 μm. Any rotation signal normal to the drive mode would then excite the out-of-plane rocking mode of the structure that is capacitively monitored. The noise equivalent rate observed by this structure was $0.1\,\mathrm{deg\,s^{-1}}$ in a 60-Hz bandwidth.

Miniaturized gyroscopes, however, existed before 1993. Quartz-based vibratory gyroscopes had already been around in the early 1980s. Examples of these devices are fused-quartz hemispherical resonator gyro (HRG) by Delco (Ragan and Lynch 1984), quartz tuning forks such as the QRS by Systron Donner (Soderkvist 1990), and a piezoelectric vibrating disk gyro (Burdess and Wren 1986). Although quartz vibratory gyroscopes can yield high quality factors at atmospheric pressure and demonstrate high sensitivity, the fabrication technology is not compatible with IC fabrication technology.

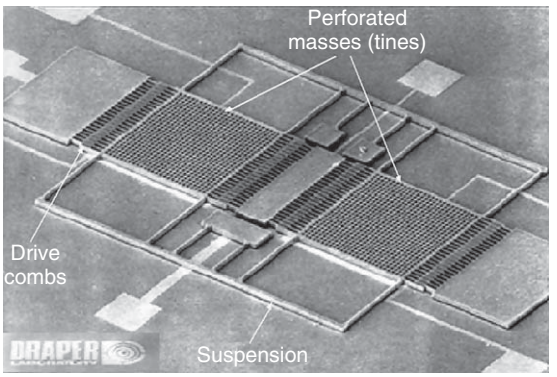


Figure 8 SEM view of Draper's silicon-on-glass tuning fork gyroscope. (Source: Bernstein J, Cho S, King T, Kourepenis A, Maciel P, Weinberg M 1993 A micromachined comb-drive tuning fork rate gyroscope. *Proc. IEEE Micro Electro Mechanical Systems Workshop (MEMS 93)*, Fork Lauderdale, FL, USA, pp. 143–148.

In the late 1980s, after successful demonstration of batch-fabricated silicon accelerometers, extensive efforts were initiated to replace quartz with silicon in micromachined vibratory gyroscopes. A number of vibratory gyroscope structures were demonstrated in the process: (i) tuning fork structures, (ii) vibrating beams or proof masses, (iii) vibrating shells, and (iv) frame gyroscopes. In Section 2.06.4.1 various capacitive micromachined gyroscopes in terms of the process technology are discussed. It was mentioned earlier that piezoresistive and capacitive techniques are the primary means of detecting Coriolis-based motion of vibratory gyroscopes. Although piezoresistive devices are easier to fabricate and require a simpler electronic interface because of their lower output impedance when compared with capacitive devices, they have high temperature sensitivity and consequently suffer from poor resolution and stability. For this reason this chapter focuses on capacitive vibratory gyroscopes. In general, micromachining processes for fabrication of capacitive vibratory gyroscopes fall into one of three categories: (i) silicon bulk micromachining and wafer bonding, (ii) polysilicon surface micromachining, and (iii) combined bulk–surface micromachining also commonly referred to as mixed processes.

2.06.3.1 Surface Micromachined Gyroscopes

The field of surface micromachined vibratory gyroscopes has been extensively researched (Juneau and Pisano 1996, Oh *et al.* 1997, Park *et al.* 1997). Implementing micromachined polysilicon structures in an integrated IC fabrication process is very attractive in an economic sense. A polysilicon structural material offers high quality factor and an orientation-independent Young's modulus. Furthermore, by integrating the readout electronic circuitry on a single silicon chip, it is possible to reduce the parasitic capacitances and hence increase the signal-to-noise ratio.

Single- and dual-axis polysilicon surface micromachined gyroscopes have been realized by researchers at the University of California, Berkeley. Berkeley's z-axis vibratory rate gyroscope (Clark *et al.* 1996) resembles a vibrating beam design and consists of

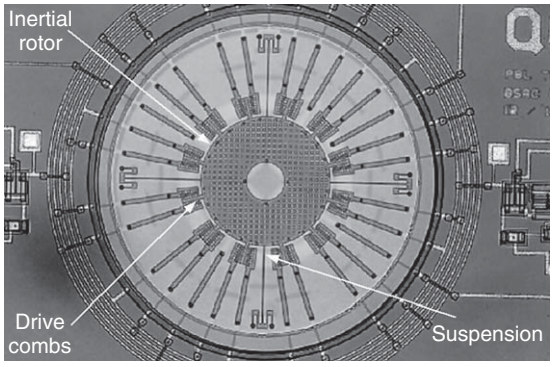


Figure 9 Close-up die shot of Berkeley's dual-axis rate gyroscope, integrated with sense and drive electronics. (Source: Juneau T, Pisano A 1996 Micromachined dual input axis angular rate sensor. *Tech. Dig. Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, pp. 299–302.)

an oscillating mass that is electrostatically driven into resonance using comb drives. Any deflections that result from Coriolis acceleration are detected differentially in the sense mode using interdigitated comb fingers. This device, 1 mm across, was integrated with a transresistance amplifier on a single die using the Analog Devices BiMEMS process. The remaining control and signal processing electronics such as the quadrature error nulling and mode matching was implemented off chip. This device demonstrated a resolution of 1 deg s^{-1} $Hz^{-1/2}$.

Berkeley has also reported a dual-axis surface micromachined gyroscope based on a rotational resonance of a 2-μm-thick polysilicon rotor disk as shown in **Figure 9** (Juneau and Pisano 1996). Because the disk is symmetric in two orthogonal axes, the sensor can sense rotation equally about both these axes. This device, integrated with electronics, yielded an ARW of 10 deg $h^{-1/2}$.

Perhaps the most successful commercial micromachined silicon gyroscope is the Analog Devices ADXRS iMEMS gyroscope series (Geen *et al.* 2002). The device implemented using the unique BiMEMS surface micromachining process integrates both the mechanical elements and electronics on a single die, thereby ensuring improved performance and robustness. The MEMS component as shown in **Figure 10** consists of two polysilicon sensing structures, each containing a dither frame that is electrostatically driven into resonance. This produces the necessary velocity element to produce a Coriolis force during angular rotation. At two of the outer extremes of each frame, orthogonal to the dither motion, lie movable fingers that are placed between pickoff fingers to capacitively monitor Coriolis

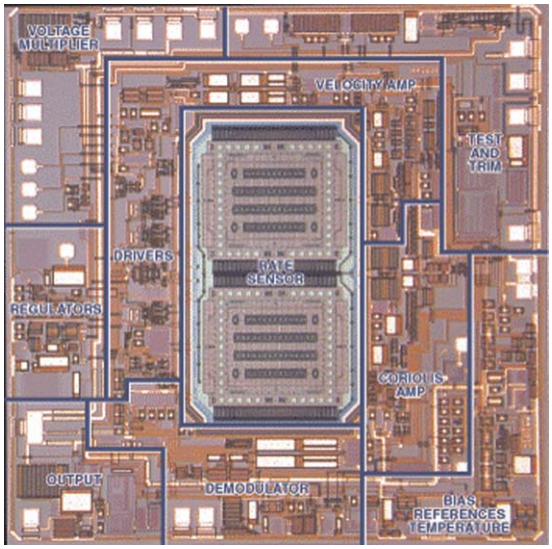


Figure 10 Die picture of an Analog Device ADXRS series microgyroscope. (Source: Geen J, Sherman S, Chang J, Lewis S 2002 Single-chip surface micromechanical integrated gyroscope with 50° Allen deviation. *IEEE J. Solid State Circuits*, **37**(12), 1860–6.)

motion. The ADXRS gyroscope demonstrates a rate sensitivity of 12.5 mV deg^{-1} s^{-1} and an estimated Allan deviation (bias drift) of 50 deg h^{-1}.

The final surface micromachined gyroscope discussed in this section has been developed by Robert Bosch GmbH (Gomez *et al.* 2005). The MEMS element consists of two identical mass-spring structures connected by a coupling spring. Each structure consists of three frames: (i) drive, (ii) Coriolis, and (iii) detection. Primary drive mode oscillations comprise movement of the drive and Coriolis frame in-plane along the x-direction, excited via comb-drive electrodes. The secondary mode (detection) is excited by Coriolis forces as a result of an angular rate around the z-axis; it comprises oscillation of the Coriolis and detection frame. Coriolis signals are detected by parallel plate capacitors within the detection frame. Decoupling of the detection frame from the primary oscillation reduces electrical crosstalk significantly. The Bosch gyroscope represents a two-die approach, with the IC interfaced with the MEMS element in a vacuum-sealed package. The Bosch gyroscope in testing yielded bias stability values approaching 1 deg h^{-1}, which is the highest reported for surface micromachined gyroscopes to date.

Surface micromachined gyroscopes discussed in the previous section suffer from thin-film residual stress, squeeze-film damping, and other problems associated with low mass. The low-mass problem implies that resolution levels of sub-0.1 deg h^{-1} typically required

for navigation-grade performance cannot be achieved. With limited device thickness, surface micromachined gyroscopes also suffer from a reduced capacitance area and thereby device sensitivity.

2.06.3.2 Bulk Micromachined Gyroscopes

The need to address the capacitance area and mass inadequacies of surface micromachined gyroscopes has directly led to research in bulk micromachined gyroscopes. The issues that originally plagued bulk micromachining such as high aspect ratio trench etching, wafer bonding, and vacuum packaging have been addressed in recent years with advancement in micromachining tools. Another trend in bulk micromachined gyroscope that is gaining popularity is the utilization of silicon-on-insulator (SOI) wafers. Because of their thick single-crystal silicon device layers, SOI wafers provide the necessary substrates for bulk mechanical structures, which are critical to gyroscope performance. Additionally, they also allow the possibility of integrating the mechanical structures of the gyroscope with the electronics, as in the case of integrated surface micromachined gyroscopes.

Bulk micromachined gyroscopes have made significant progress since 1995. In 1999, Murata reported a deep reactive ion etching (DRIE) gyroscope with decoupled sense and drive modes with a resolution of $0.07\,\mathrm{deg\,s^{-1}}$ with a 10-Hz bandwidth (Park 1999). Also reported by Samsung was a 40-μm-thick single-crystal silicon gyroscope fabricated using SOI technology (Tanaka *et al.* 1995). This device demonstrated a resolution of $0.015\,\mathrm{deg\,s^{-1}}$ in 25 Hz bandwidth, a bias drift of $500\,\mathrm{deg\,h^{-1}}$, a sensitivity of $145\,\mathrm{mV\,deg^{-1}\,s^{-1}}$, and −3 dB bandwidth of 12.5 Hz.

During the earlier development of bulk micromachined vibratory gyroscopes, no on-chip readout electronics were present because the fabrication process required Si–Si or Si–glass anodic bonding and two-side alignment. Carnegie Mellon University addressed this issue by implementing a lateral-axis angular rate sensor with in-plane vibration and out-of-plane Coriolis acceleration sensing. The sensor plus on-chip CMOS circuitry is fabricated using a post-CMOS micromachining process that utilizes interconnected metal layers as etching masks and a single-crystal silicon layer as the structural material (Xie and Fedder 2003).

The NASA JPL microgyroscope consists of three major components: (i) a silicon clover-leaf vibrating structure, (ii) a silicon baseplate that is bonded to the clover-leaf structure, and (iii) a metal post that is epoxied inside a hole on the silicon resonator. The last reported data on the electrostatically tuned and optimized JPL microgyroscope indicate an ARW of $0.1\,\mathrm{deg\,h^{-1/2}}$ with bias stability of $2\,\mathrm{deg\,h^{-1}}$ operating in vacuum under matched-mode condition (Bae *et al.* 2002).

So far the emphasis has been on gyroscopes operating with proof masses suspended by a flexural network and operating with nondegenerate vibratory modes. The focus now shifts to shell-type gyroscopes. This is a family of devices that rely on the Coriolis-induced transfer of energy between degenerate vibratory modes. Perhaps the most commonly known shell-type structure is the vibrating ring gyroscope. The vibrating ring gyroscope consists of a ring; semicircular support springs; and drive, sense, and control electrodes. Symmetry considerations require at least eight springs to result in a balanced device with two identical flexural modes that have equal natural frequencies. The ring is electrostatically vibrated into an elliptically shaped primary flexural mode with a fixed amplitude. When the device is subjected to rotation around its normal axis, Coriolis force causes energy to be transferred from the primary mode to the secondary flexural mode that is located 45° apart from the primary mode, causing amplitude to build up proportionally in the latter mode; this buildup is capacitively monitored.

The first micromachined version of the vibrating ring gyroscope was fabricated by electroforming nickel into a thick polyimide (or photoresist) mold on a silicon substrate in a postcircuit process (Putty and Najafi 1994). The gyroscope implemented by the University of Michigan demonstrated a resolution of approximately $0.5\,\mathrm{deg\,s^{-1}}$ in a 10-Hz bandwidth (limited by the readout electronic noise). However, the nickel electroforming technology used for this device has a major drawback. The thermal expansion coefficient of the silicon substrate is different from that of the nickel sensor element. Therefore, with temperature fluctuations, the ring expands (or shrinks) more than do the adjacent electrodes, which are attached to the substrate, causing the ring-to-electrode gap spacing to change; this in turn results in an increased temperature sensitivity of offset and scale factor.

In an effort to eliminate temperature sensitivity and low quality factor issues that plagued the nickel implementation, BAE systems developed a single-crystal silicon bulk micromachined implementation of the vibrating ring gyroscope. In 2002, the University of Michigan reported a high-resolution vibrating ring

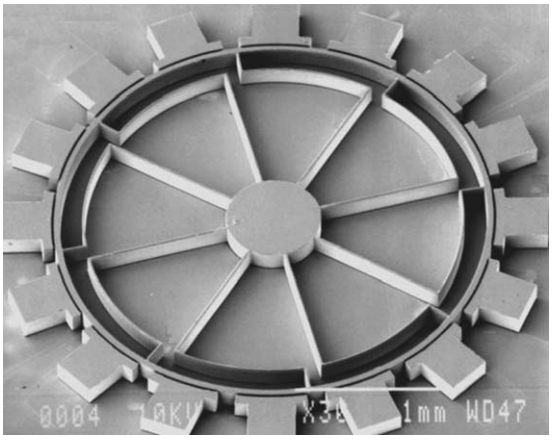


Figure 11 SEM image of the vibrating ring gyroscope implemented on (111) silicon substrate. (Source: He G, Najafi K 2002 A single-crystal silicon vibrating ring gyroscope. *Proc. IEEE MEMS 2002*, Las Vegas, USA, pp. 718–21.)

gyroscope implemented using (111) single-crystal silicon structural material (**Figure 11**) (He and Najafi 2002). The 150-μm structural device was defined by DRIE and it demonstrated an impressive noise floor of $10\,\mathrm{deg\,h^{-1}\,Hz^{-1/2}}$ in a 2-Hz bandwidth with a reported rate sensitivity of approximately $130\,\mathrm{mV\,deg^{-1}\,s^{-1}}$.

2.06.3.3 Mixed-Mode Micromachined Gyroscopes

As the trend toward greater device thickness increases, the need to maintain and improve the aspect ratio of the capacitive operating gaps also becomes essential. Even the most advanced deep reactive ion silicon etching tools are unable to define in a repeated fashion trench aspect ratio greater than 30:1. This severely limits the performance capability of silicon bulk micromachined vibratory gyroscopes. The solution to the aspect ratio issue is the development of novel fabrication technology that combines the aspects of both surface and bulk micromachined processes. The mixed-mode process technology utilizes bulk silicon structural material to define device geometries while relying on a sacrificial layer used in surface micromachining process to define the operating capacitive gaps.

To eliminate the myriad of problems faced by its nickel electroforming technology, the University of Michigan developed the high aspect ratio poly- and single-crystal silicon (HARPSS) process (**Figure 12**) (Ayazi and Najafi 2000). This novel, mixed-mode fabrication technology is capable of producing thick MEMS structures while implementing submicron high aspect ratio sensing and tuning capacitive gaps. This unique feature improves the device sensitivity and noise floor while dramatically reducing the operating voltages when compared with its nickel predecessor. The polysilicon HARPSS vibrating ring gyroscope demonstrated a rate sensitivity of $0.2\,\mathrm{mV\,deg^{-1}\,s^{-1}}$ in a dynamic range of $\pm 250\,\mathrm{deg\,s^{-1}}$. The measured noise floor of the gyroscope system was approximately $0.1\,\mathrm{deg\,s^{-1}\,Hz^{-1/2}}$ limited primarily by the poor quality factor associated with trench-refilled structural polysilicon.

An alternate shell structure was displayed by the Georgia Institute of Technology by utilizing the HARPSS process (Zaman *et al.* 2005). The RSG may be visualized as a superposition of two identical square shells that are 45° apart (**Figure 13**). This ensures pairs of degenerate flexural vibratory modes

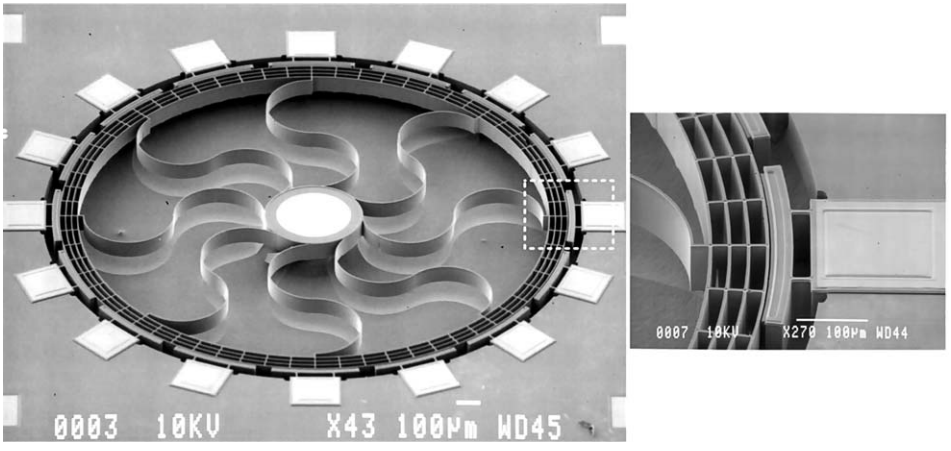


Figure 12 Implementation of the vibrating ring gyroscope using the high aspect ratio poly- and single-crystal silicon (HARPSS) process.

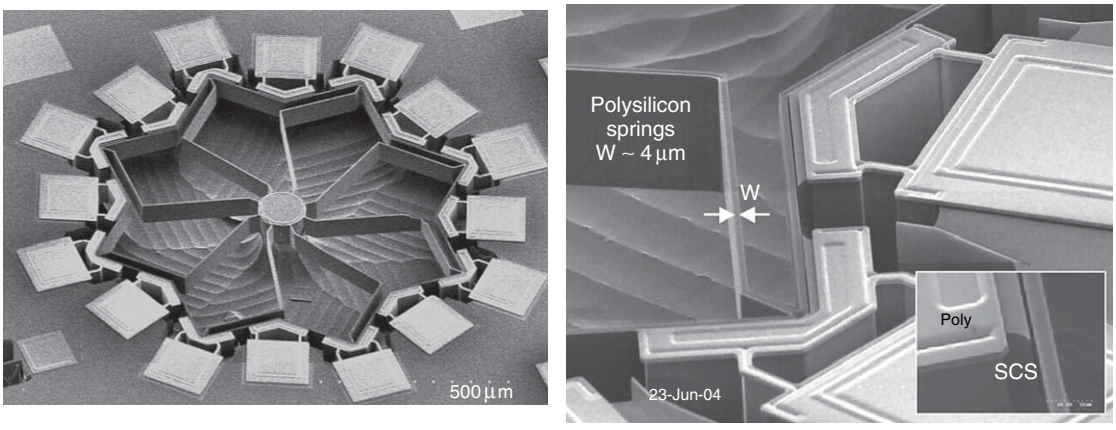


Figure 13 SEM overview of a 1-mm polysilicon resonating star gyroscope. *Right*: View of the electrode. *Inset*: Close-up of the 1-µm actuation gap defined through sacrificial oxide.

in the resulting eightfold star shell, which is anchored to a central post through flexural springs. The primary advantages of the RSG are improved quadrature minimization and increased electrode area. The RSG demonstrated a resolution and a sensitivity similar to those of the HARPSS vibrating ring gyroscope.

2.06.4 Case Study: An In-Plane M^2-TFG

The M^2-TFG (Zaman *et al.* 2006) is an example of a type I vibratory gyroscope. The M^2-TFG as illustrated in **Figure 14** consists of two proof masses supported by flexural springs and anchored at a

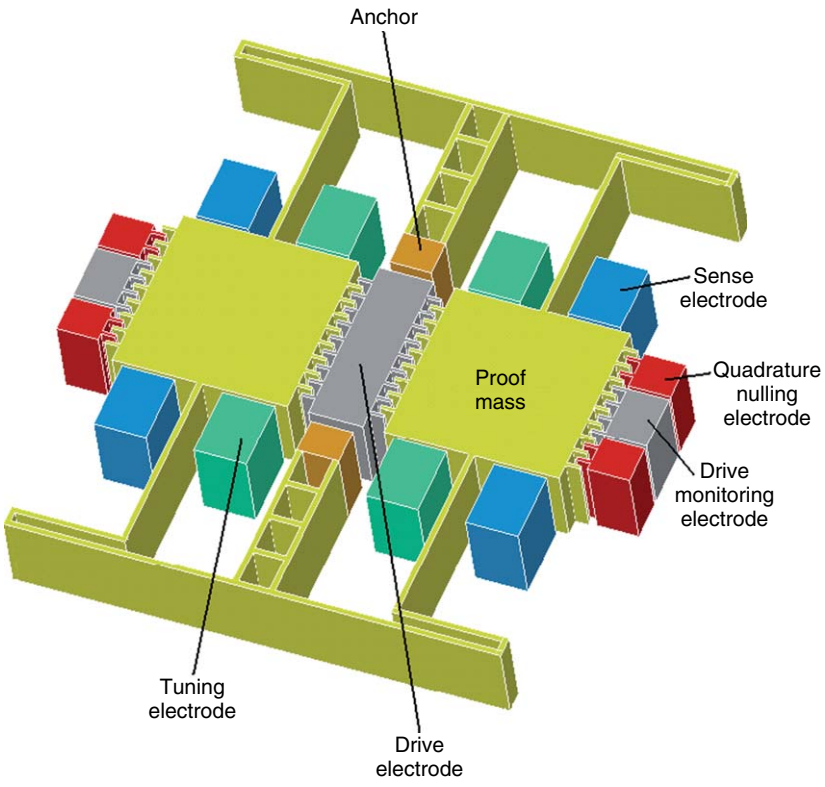


Figure 14 Schematic of the matched-mode tuning fork gyroscope (M^2-TFG).

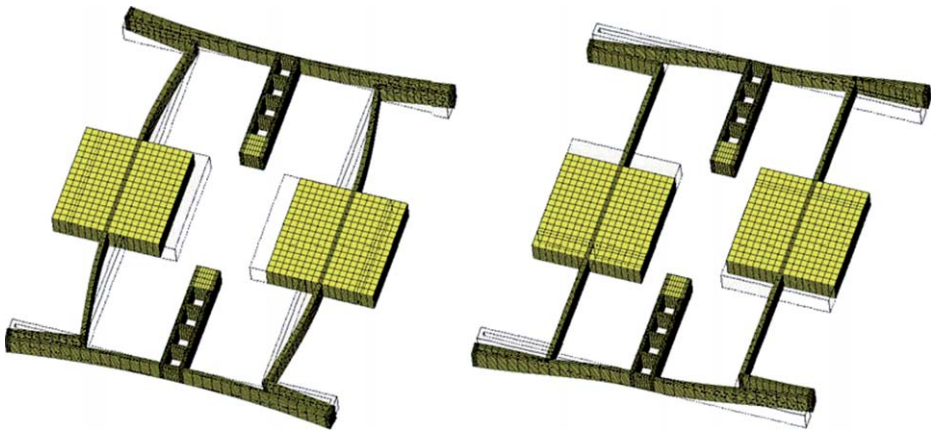


Figure 15 Resonant operating modes of the matched-mode tuning fork gyroscope (M^2-TFG): (a) drive mode and (b) sense mode vibrations.

central post. Actuation, sensing, and tuning electrodes are distributed around the proof masses and flexures. The operating principle is based on a standard tuning fork's response to rotation. The proof masses (akin to the tines of a tuning fork) are driven at resonance along the x-axis using comb-drive electrodes, and the Coriolis acceleration induced by rotation about the z-axis is sensed capacitively using electrodes along the y-axis.

The in-plane drive and sense resonant modes are illustrated in **Figure 15**. A major incentive in utilizing such a symmetric tuning fork architecture as opposed to a frame or single-mass design, is the differential sensing capability. As a result, linear acceleration signals are rejected as common mode without the need for complex electronics.

2.06.4.1 Flexural Spring Design

The flexural spring designs must satisfy a host of critical properties. The ideal spring design must ensure that there is zero/minimal coupling between the two resonant modes in the absence of rotation. To this effect, a fishhook architecture was developed, which ensures that the mode shapes are aligned laterally (two-directional flexibility), and more importantly, the two mode shapes move along independent axis during operation (Park *et al.* 1997). The spring structure design makes it possible to drive the solid proof masses linearly with displacement amplitudes in the range of 4–6 μm with minimal displacement along the sense axis. The second important criterion is to ensure that the two operating mode frequencies lie in close proximity to one another while remaining isolated from the other resonant modes (such as out-of-plane resonant modes). Detailed finite element analysis simulations of the flexural springs were performed to optimize the dimensions to allow the sense mode to occur 50–100 Hz higher than the drive mode. Upon fabrication, the sense mode frequency is electrostatically tuned (using the tuning electrodes) to match the drive mode frequency.

A key determinant of performance for a gyroscope is its Q. In order to achieve higher rate resolution and better bias stability, the design of gyroscopes with high sense mode Q is consistently pursued. A high Q in the drive mode is also necessary to obtain large drive amplitudes using small drive voltages – a highly desirable feature required in low-power CMOS interfacing. Several energy dissipation mechanisms exist in micromachined gyroscopes, such as air damping, thermoelastic damping (TED), surface loss, and support loss. Among them, air damping can be eliminated by operating the gyroscope in vacuum, whereas surface loss is often negligible through optimized processing. The measured unloaded Q is hence a combination of two dissipation mechanisms expressed as

$$\begin{aligned} 1/Q &= 1/Q_{\text{TED}} + 1/Q_{\text{support}} + 1/Q_{\text{surface}} + 1/Q_{\text{air}} \\ &\approx 1/Q_{\text{TED}} + 1/Q_{\text{support}} \end{aligned} \quad [18]$$

During resonant flexural vibrations, the proof masses exert a time-harmonic load on their support structure. Acting as an excitation source, this time-harmonic load excites elastic waves propagating through the support. Part of the vibration energy

dissipated through elastic wave propagation in the support media is commonly referred to as support loss (Hao *et al.* 2003). The lack of thermal equilibrium between various parts of the vibrating solid is responsible for TED in solids. In order for the equilibrium to be reinstated, irreversible heat flow driven by the temperature gradient occurs, also referred to as thermal relaxation. The irreversibility of such a relaxation process is associated with energy dissipation or damping. Q_{TED} defines the upper limit for maximum achievable Q in an M^2-TFG structure and is a function of the flexural spring width and the device operating frequency (Zaman *et al.* 2004).

The drive mode and the sense mode vibrations depicted in **Figure 15** show that the two proof masses vibrate in the opposite directions and consequently alleviate support loss. The vibrations in the drive mode cause the forces applied on the anchors in the opposite directions and thus lead to cancelation in support loss to a great extent. Similarly, the vibrations in the sense mode introduce forces to the anchors. However, the forces at the two anchors are off the center and lead to a net torque applied on the support structure. The flexural spring dimensions and operating frequency regime are carefully selected to alleviate support loss and operate in a region of high Q_{TED}.

2.06.4.2 Device Fabrication

The M^2-TFGs are fabricated on thick SOI substrates. The moving sections of the structure and the areas under the comb drives are first released from the backside of the wafer by etching the handle silicon layer through to the buried oxide layer using the Bosch process. The buried oxide is then removed in an RIE system. Finally, the device layer is patterned all the way through, leaving behind a suspended structure whose anchor is attached to the substrate via several support posts.

The fabrication process is very simple and precludes the requirement of any perforations in the proof mass, resulting in a larger mass per unit area (**Figure 16**). The simultaneous elimination of the ground plane under the comb drives prevents the excitation of the out-of-plane modes and detrimental effects associated with comb levitation. The fact that all the electrodes and the structure are defined using a single mask ensures that there are no horizontal misalignment errors of actuation/sense gaps. **Figure 17** shows a few SEM images of fabricated devices implemented on a thick SOI substrate.

2.06.4.3 Interface Electronics

The complete M^2-TFG system is shown in **Figure 18**. It consists of the drive, sense, quadrature nulling, and automatic mode-matching system blocks, implemented in a 0.6-μm CMOS process (Sharma *et al.* 2007). While the function of each of these systems had been highlighted previously, the following section contains a brief description of the M^2-TFG electronics and highlights some of the system aspects that are unique to the M^2-TFG interface system.

2.06.4.3.1 Front-end I-V conversion

A CT low-noise transimpedance amplifier (TIA) is used as the front-end interface (Sharma *et al.* 2006) as indicated in **Figure 19**. Such a circuit topology is capable of measuring subattofarads of capacitance change at low power, without the need for complex switched capacitor techniques. A CT approach is adopted to avoid the kT/C noise associated with switched capacitor front-ends (Jiang *et al.* 2000). The shunt-shunt feedback of a TIA represents low input impedance to the capacitive sensor, thereby mitigating the effects of the increased input parasitic capacitance (C_I) of a two-chip approach and minimizing Q-loading. **Figure 19** shows the major noise contributors in a TIA that is interfaced to a capacitive

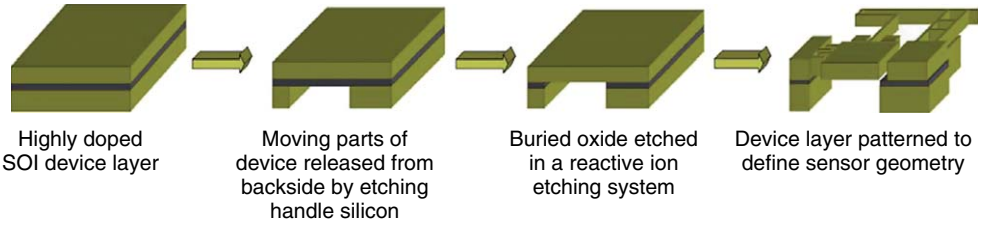


Figure 16 Outline of the fabrication process flow for silicon-on-insulator (SOI) matched-mode tuning fork gyroscope (M^2-TFG).

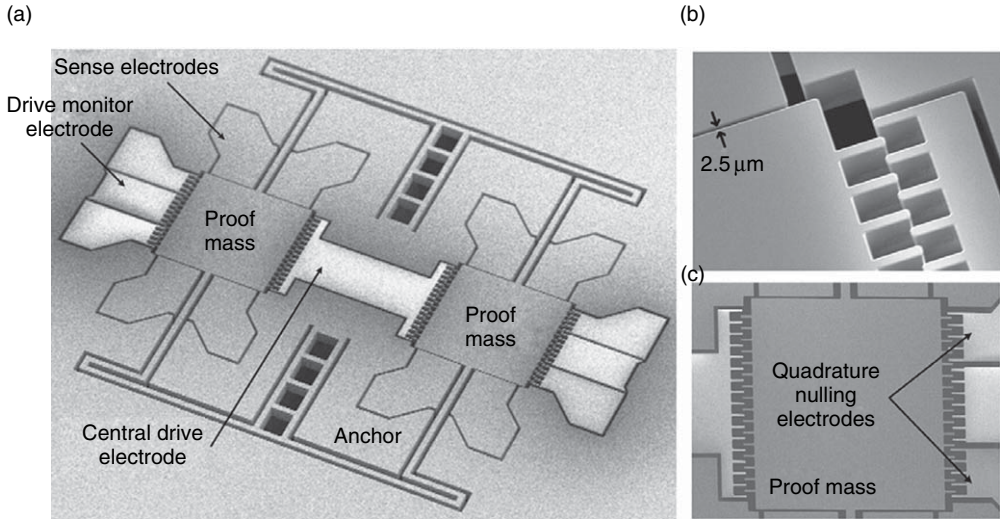


Figure 17 (a) A SEM image of a fabricated 60-μm-thick matched-mode tuning fork gyroscope (M^2-TFG), (b) a close-up of drive and sense gaps, and (c) a view of the quadrature nulling electrodes.

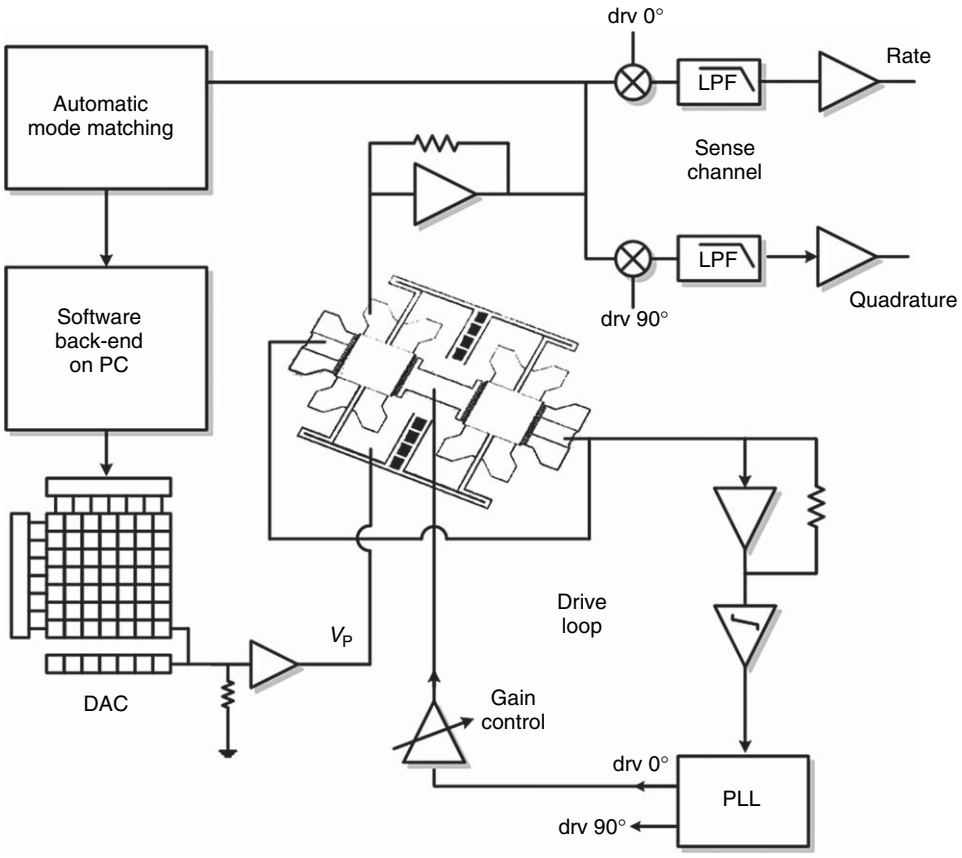


Figure 18 Schematic of the complete matched-mode tuning fork gyroscope (M^2-TFG) interface system.

sensor. $v^2_{\text{op-amp}}$ and $i^2_{\text{op-amp}}$ are the input-referred voltage and current noise of the core amplifier, respectively.

The advantage of using a TIA front-end for low-noise capacitive sensing is apparent from the expression for the TIA SNR. Increasing R_F not only allows

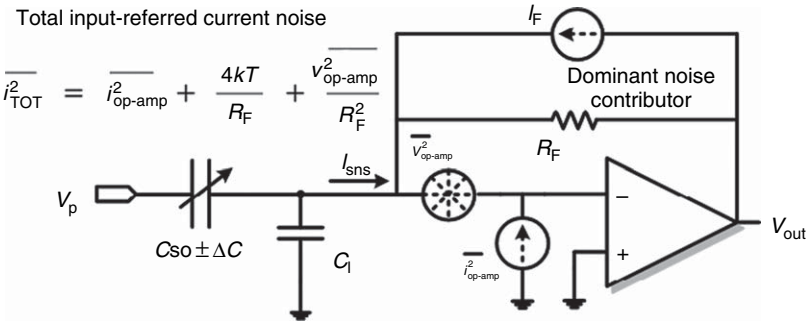


Figure 19 Schematic of transimpedance amplifier (TIA) front-end and various noise sources.

for a high first-stage gain but also improves the system SNR by a factor of $\sqrt{R_F}$ as indicated in eqn [19].

$$SNR = \frac{S_O}{N_O} = \frac{I_{SIGNAL}}{I_{NOISE}} = \frac{I_{SIGNAL}\sqrt{R_F}}{\sqrt{4kT}} \qquad [19]$$

Plots of the measured total input current noise of the T-network TIA for different values of V_{CNTRL} are shown in **Figure 20**. In the case of T-network TIA, in the signal band (10 to 20 kHz), the flicker plus thermal noise from the MOS resistor is the dominant noise contributor. As was explained earlier, the input-referred current noise decreases with increasing transresistance (R_F), and this is clearly evident from **Figure 20**.

In case of a high-Q resonant sensor, the spot noise at the signal frequency is the quantity of interest. At 10 kHz, the T-network TIA with a V_{CNTRL} of 0.96 V has a transimpedance gain of 1.6 MΩ and an input-referred current noise of 99 fA $Hz^{-1/2}$. This corresponds to a capacitive resolution of 0.04 aF $Hz^{-1/2}$ at 10 kHz. For this case, the measured output voltage noise at 10 kHz is −136 dB $Hz^{-1/2}$, which corresponds to a signal to noise plus distortion ratio (SNDR) of 104 dB in a 10-Hz bandwidth.

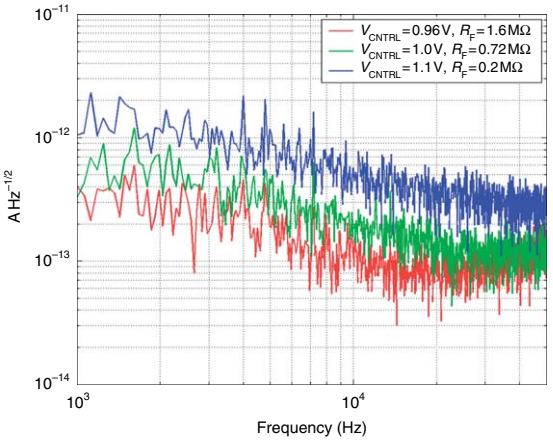


Figure 20 Measured input-referred current noise of a T-network transimpedance amplifier (TIA) for different V_{CNTRL}.

2.06.4.3.2 Synchronous I-Q demodulation

Synchronous I-Q demodulation is used to extract the rate signal from the sensor output electrodes and differentiate it from quadrature error. The sensor output is multiplied with a square wave that is in phase with the proof mass velocity and then low pass filtered to obtain an analog voltage proportional to the input rotation rate. An analog voltage indicative of the quadrature error is obtained by multiplying the sensor output with a square wave in phase with the proof mass position. The signals in phase with and in quadrature with the velocity are generated by using a frequency divider in the PLL as shown in **Figure 21(a)**. A CMOS Gilbert cell (Gilbert 1968) is used for the multiplication as shown in **Figure 21(b)**. The output of the four-quadrant multiplier is low pass filtered to yield analog signals proportional to the rotation rate and quadrature error.

2.06.4.3.3 Drive oscillator

The drive amplitude (q_{drive}), which is the maximum displacement of the oscillating proof mass, is a critical parameter in determining the noise floor and Coriolis sensitivity. The drive mode of the M^2-TFG is excited into resonance using interdigitated comb fingers. In a comb-driven structure, the maximum

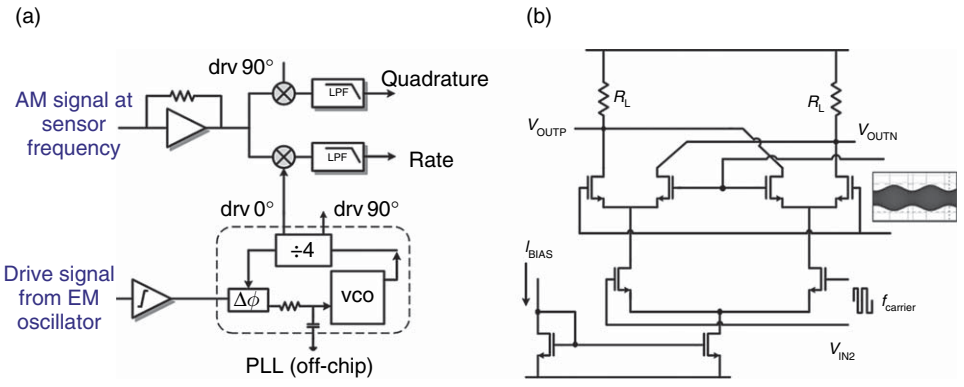


Figure 21 Schematic of the sense channel showing (a) I-Q demodulation and (b) Gilbert cells used for multiplication.

achievable drive amplitude is proportional to the square of the gap between adjacent combs. In the case of the M^2-TFG, this is further restricted by operation in the matched mode.

$$q_{\text{drive}} = \frac{-w_0 + \sqrt{w_0^2 + 2\left(k_x/k_y\right)d^2}}{2} \cong \frac{-w_0 + \sqrt{w_0^2 + d^2}}{2} \quad [20]$$

In eqn [20], w_0 indicates the initial comb overlap while d represents adjacent comb gap. An increased d also translates to increased mechanical motional impedance along the drive axis as given by eqn [21].

$$R_{\text{motional}} \propto \frac{\omega_0 M d^2}{Q V_P^2} \quad [21]$$

To illustrate with an example, in order to achieve drive amplitudes of 3 μm, the required comb gap is 7 μm, which corresponds to a motional impedance of 12 MΩ (at 15 kHz in vacuum). Consequently, larger drive voltages will have to be applied and larger on-chip loop gains need to be realized for sustaining the drive oscillations. Larger drive voltages imply higher power consumption, whereas larger loop gain implies more silicon real estate – both of which place constraints on electronic integration. Hence, there is a need for a front-end that can provide significantly high transimpedance gain with a low area overhead. The T-network TIA described in the previous section satisfies the above requirements. In the implemented M^2-TFG drive loop, the T-network TIA is cascaded with a phase-shifting buffer to satisfy Barkhausen's criterion and sustain oscillations in the series resonant electromechanical drive loop, as shown in **Figure 22**.

Once drive oscillations have been started, it is necessary to maintain a constant amplitude and prevent the modulation of the sensor output. An ALC circuit (Lee and Nguyen 2003) is therefore used to keep q_{drive} constant, thereby preventing false rate outputs. As shown in **Figure 22**, the ALC circuit consists of a full-wave rectifier, envelope detector, and instrumentation amplifier to supply the level control voltage. The ALC output voltage is used to control the MOS transistor in the T-network. **Figure 22** also shows the buffered closed-loop drive oscillation waveform and the spectrum of the signal when the drive loop is interfaced with a prototype M^2-TFG.

2.06.4.4 Device Characterization

Optimizing the flexural designs ensures minimal support loss and TED in flexures to result in high-Q in-plane drive and sense resonant modes, as well as in minimal coupling and frequency separation between the two modes (**Figure 23**). Typical open-loop quality factors measured for M^2-TFG operating modes are on the order of 80 000 and 50 000 for the drive and the sense modes, respectively.

Figure 24 shows the tuning characteristics of the sense resonant mode as the V_P on the MEMS structure is increased. This mode shows a variation of approximately 1600 ppm V^{-1} over a tuning voltage range of 20 V. The tuning characteristics of the sense mode are important as the sense frequency is tuned to match the drive frequency for mode matching. In addition, the tuning characteristics of the sense mode can be used to electrostatically maintain mode matching in the presence of temperature variations. Parabolic temperature compensation of micromachined resonators has been demonstrated to achieve frequency stabilities as low as 39 ppm $°C^{-1}$ (Sundaresan *et al.* 2005). A similar technique can be applied for the temperature compensation of micromachined silicon gyroscopes.

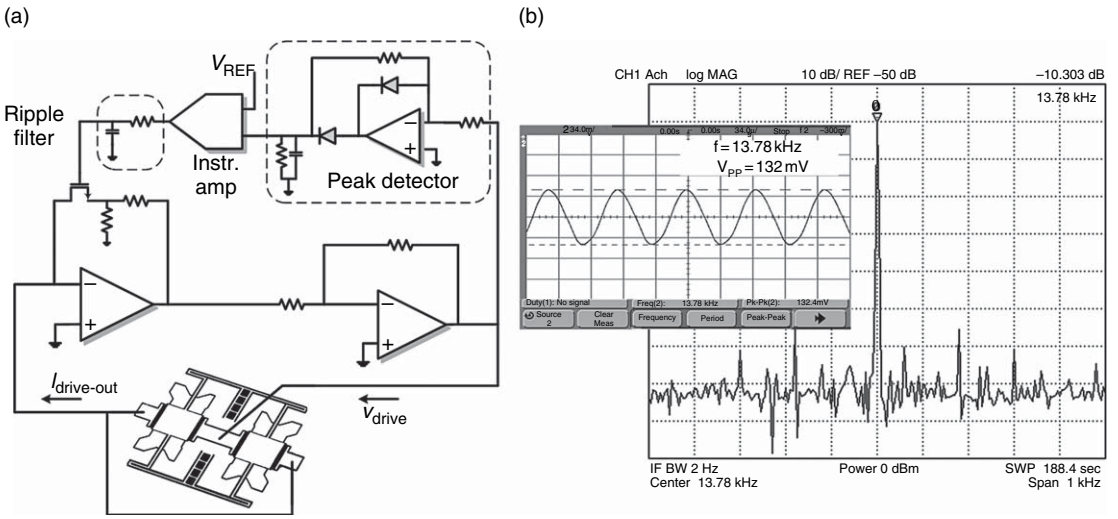


Figure 22 (a) Schematic of drive resonant oscillator and automatic level control (ALC) circuit and (b) waveform and spectrum.

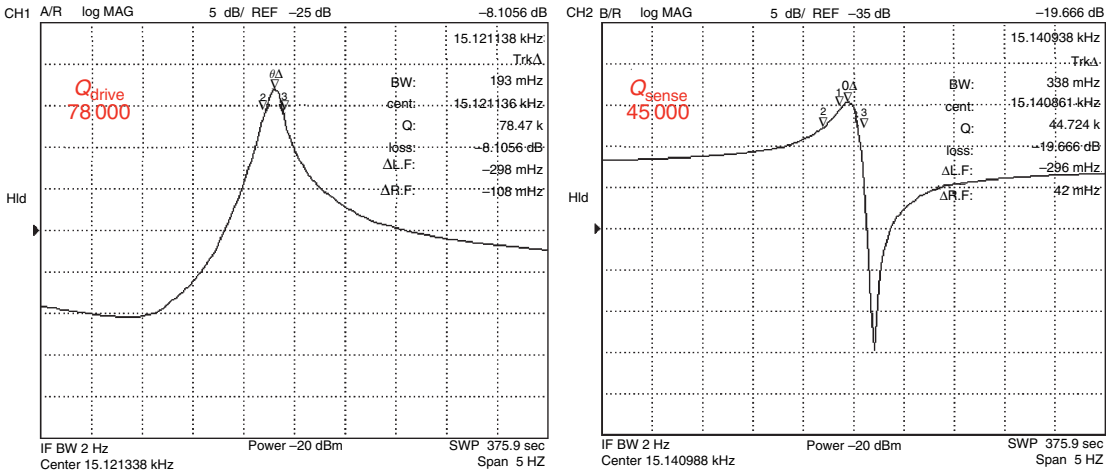


Figure 23 Measured drive and sense resonant mode *Q* (in mTorr vacuum).

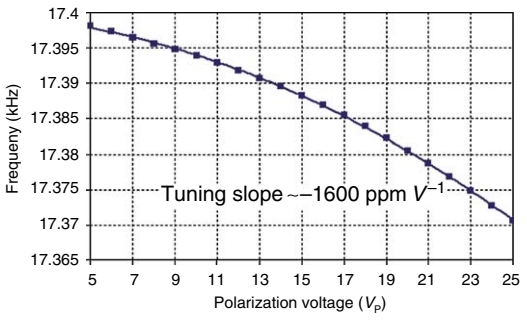


Figure 24 Variation of the sense resonant mode frequency as a function of the polarization voltage, V_P.

2.06.4.5 Quadrature Nulling and Mode Matching

The sense mode is designed to be slightly higher in frequency than the drive mode to enable mode matching even in the presence of process variations. Mode matching is achieved by increasing V_P on the MEMS structure until electrostatic spring softening decreases the sense mode frequency to become equal to that of the drive mode (0 Hz split). This feature differentiates the M^2-TFG from other MEMS gyros (Geen *et al.* 2002).

Despite a mechanically decoupled design, in practice, fabrication imperfections can lead to nonzero off-diagonal elements in the spring stiffness matrix resulting in an undesirable zero-rate output (ZRO) at the sensor operating frequency. As discussed earlier, this quadrature error signal prevents perfect mode matching. As shown in **Figure 25**, an initial frequency split of 60 Hz is reduced to 10 Hz without the need for any external quadrature cancelation.

Only through the electronic minimization of the quadrature error is the perfect matching of the drive and the sense resonant modes possible (Clark *et al.* 1996). The ZRO is minimized in a postfabrication trimming step by applying DC voltages at specific quadrature nulling electrodes to properly align the device. These electrodes are shown in **Figure 17**. Despite quadrature nulling, there exists a finite amount of ZRO. The term perfect mode matching practically implies that the drive and the sense modes are so close in frequency that one is indistinguishable from the other (i.e., frequency separation of less than 100 mHz). Once matched, synchronous I-Q demodulation (Putty and Najafi 1994) is used to distinguish between the ZRO and the Coriolis signal. **Figure 25** shows a perfectly matched M^2-TFG device after quadrature nulling. It is generally observed that mode matching and quadrature nulling do not degrade the operating Q.

2.06.4.6 Performance Analysis

2.06.4.6.1 Scale factor measurements

The sensor output voltage was measured at varying angular speeds to characterize the sensitivity of the M^2-TFG system. The scale factor of the M^2-TFG under matched-mode operation was measured to be 24 mV deg^{-1} s^{-1} as indicated in **Figure 26(a)**. **Figure 26(b)** shows the time domain response of the sensor to clockwise and counterclockwise rotation steps respectively, whereas **Figure 26(c)** shows the response of the M^2-TFG to an input sinusoidal rotation.

Scale factor linearity is of significant concern in angular rate sensing. Typically, scale factor linearity is expressed in terms of percentage of full scale (% of FS) range. There are a number of factors that affect scale factor linearity. Driving the proof masses beyond their linear regime along the drive axis can result in nonlinearity in the drive velocity and hence in the Coriolis-induced sense axis displacement. Fabrication-induced imperfections along the sidewalls of the proof mass and sense electrodes can lead to a nonuniform sense rest capacitance and can cause significant variation in scale factor linearity. Distortion in the front-end capacitance-to-voltage converter can easily result in significant distortion of the Coriolis signal. The bandwidth of the signal chain that is used to amplify and demodulate the Coriolis signal must be higher than the gyro resonant frequency. Otherwise there can be excess phase change (Clark *et al.* 1996) that can result in unwanted phase variations. Finally, the multiplier must have sufficient linearity to perform the synchronous demodulation. As discussed earlier, there are a number of factors that affect SF linearity. Therefore, the design of a linear signal processing chain for Coriolis detection is nontrivial.

The scale factor stability is directly affected by the fluctuations in the matched-mode Q (Q_{EFF}) over time. It was, however, observed that the measured Q_{EFF} for the M^2-TFG remained constant over a period of 24 h at a fixed room temperature and pressure and over the course of week only varied by less than one-tenths of percent.

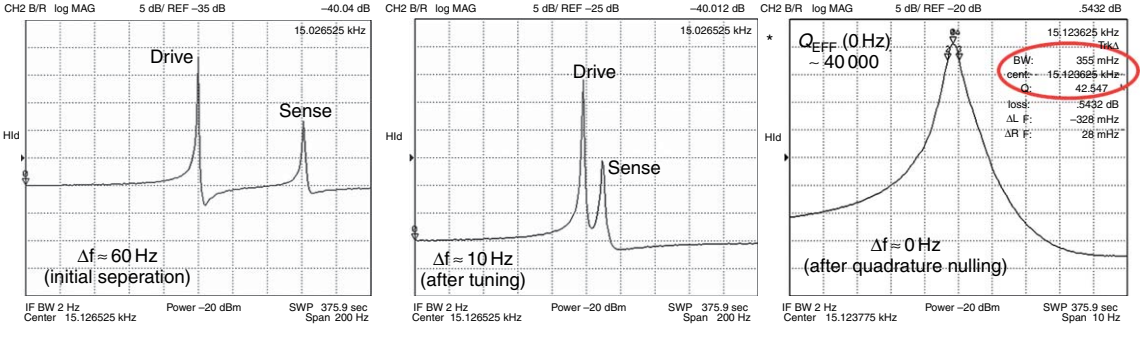


Figure 25 Pair of frequency response plots showing initial electrostatic tuning to compensate frequency split and perfect matched-mode tuning fork gyroscope (M^2-TFG) after quadrature nulling.

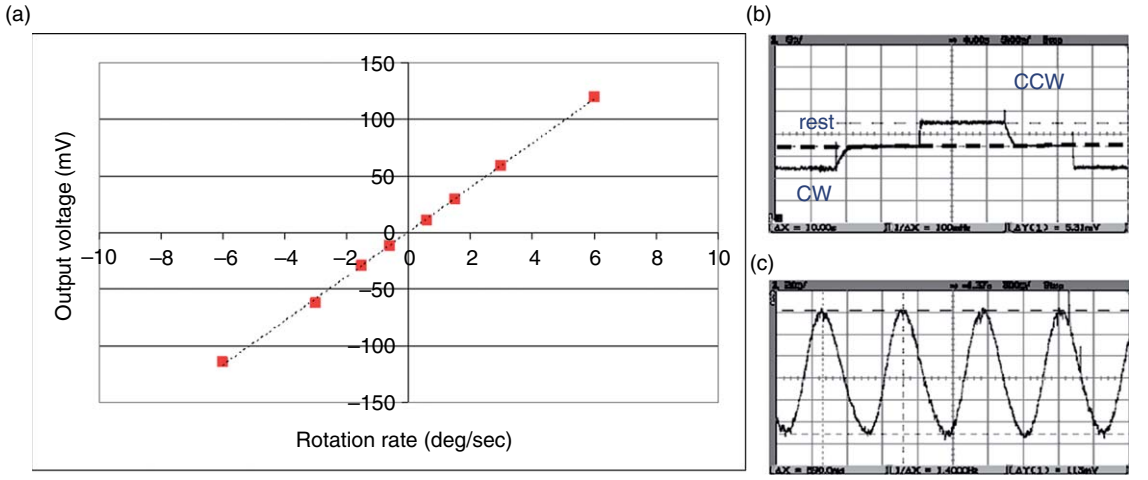


Figure 26 (a) The rate sensitivity plot for the matched-mode tuning fork gyroscope (M^2-TFG) under matched-mode operation, (b) response of the gyro to clockwise (CW) and counterclockwise (CCW) step inputs, and (c) response of the gyro to a sinusoidal input rate.

2.06.4.6.2 Bias drift measurements

Bias drift is a fundamental performance gauging parameter in any IMU. The bias drift of the M^2-TFG system interfaced with the electronics was evaluated using the guidelines set by the IEEE (1997). The bias drift has been estimated using the Allan variance technique. The Allan variance curve was subsequently curve-fitted to determine the different noise contributions of the gyro system.

The ZRO of the device was sampled every 100 ms for a period of 12 h (at room temperature and vacuum). Using the collected ZRO data, an Allan variance analysis was performed to characterize the long-term stability of the matched-mode device. A time slice of the ZRO from one electrode is shown in Figure 27(a). The ZRO level may be minimized through proper vibration isolation and elimination of other extraneous common-mode noise. The root Allan variance plot of the M^2-TFG is shown in Figure 27(b).

The minima point of the Allan variance graph is a measure of the bias stability of the system. From the Allan variance graph, the measured bias instability of the system (IEEE 1997) is 0.96 deg h^{-1}. The bias drift of a microgyroscope becomes important when one considers the fact that it is used along with accelerometers to obtain heading information – i.e., position and orientation. The rate information from the microgyroscope is integrated to obtain the angle. Any long-term variations in the rate (of which the

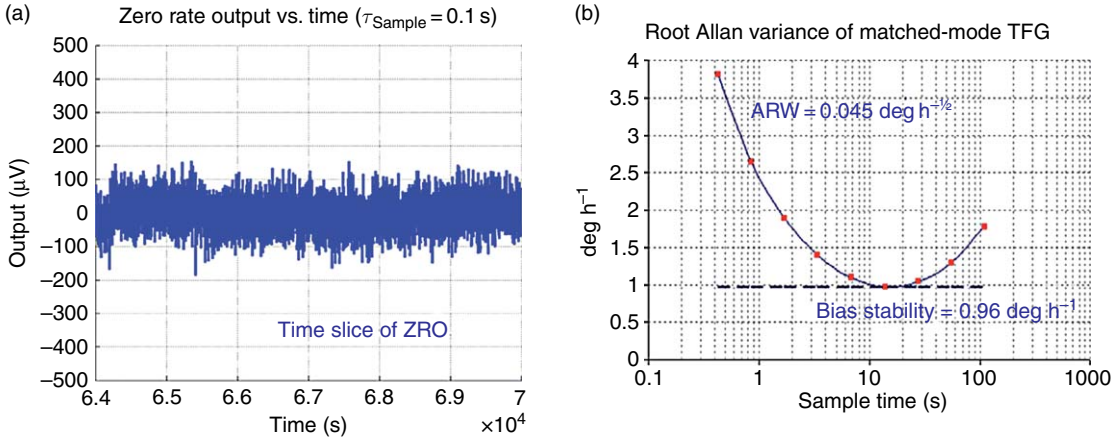


Figure 27 (a) Time slice of the zero-rate output (ZRO) and (b) root Allan variance plot of the matched-mode tuning fork gyroscope (M^2-TFG) interfaced with electronics.

bias drift is an accurate indicator) add up and can cause a large error in angle and orientation information. Modern gyroscope systems periodically calibrate themselves with GPS to ensure that the heading information is accurate. The longer a system can function accurately without the need for calibration, the better and more accurate heading and orientation information it can deliver. A long interval between calibration sequences is crucial for applications such as deep-sea navigation and oil exploration where it is not possible to resurface very easily to calibrate with the GPS.

Apart from the long-term stability of the microgyro, the Allan variance plot also gives information about the short-term stability of the device. The short-term stability of the device is directly dependent on the noise floor or resolution of the sensor. The $\tau^{-0.5}$ region of the Allan variance curve gives a measure of the ARW of the microgyroscope. From the Allan variance curve, the measured ARW is $0.045\ \text{deg h}^{-1/2}$.

2.06.4.6.3 Temperature measurements

In order to guarantee scale factor stability and ensure low bias drift, it is essential that the mode matching be maintained over temperature. The M^2-TFG was tested under both matched and split-mode conditions to determine accurately the effects of temperature on device performance. **Figure 28(a)** shows the measured frequency variation in the individual (unmatched) resonant modes of an M^2-TFG sample as a function of temperature. Such a frequency shift is attributed to the temperature dependency of Young's modulus of signal crystal silicon (SCS). The drive mode temperature coefficient of frequency (TCF) is measured to be about $-22\ \text{ppm}\ ^\circ\text{C}^{-1}$. As indicated by **Figure 28(a)**, the TCF of the drive and of the sense resonant modes vary by less than $1\ \text{ppm}\ ^\circ\text{C}^{-1}$.

The TCF data for a matched-mode device are shown in **Figure 28(b)**. Matched-mode condition is maintained over temperature because the two modes have similar TCFs and track with temperature. Q_{EFF} was observed to slightly decrease at elevated temperature and may be attributed to TED or finite frequency split generated due to variation in TCF between the individual modes. The effect of degradation in Q is reflected in the Allan variance analysis at 70°C which yielded a bias instability of $1.1\ \text{deg h}^{-1}$.

2.06.4.6.4 Comparison between matched- and split-mode operations

In this section, performance parameters of an M^2-TFG operating in split-mode condition are examined. **Figure 29(a)** depicts the modes of an M^2-TFG matched to within 2 Hz of each other. The measured Q_{EFF} in this case is roughly 10 000. The measured scale factor is approximately $7.2\ \text{mV deg}^{-1}\ \text{s}^{-1}$. As indicated in eqn [22], the sensitivity is a direct function of the Q_{EFF}.

$$\text{Sensitivity} = \frac{2V_P}{C_{S0} + C_{\text{Parasitics}}} \frac{C_{S0}}{d_0} \frac{Q_{\text{EFF}} x_{\text{drive}}}{\omega_{\text{sense}}} \qquad [22]$$

.

Another subtle effect of mode mismatch is that the scale factor of the microgyroscope can vary with the input signal frequency. As the modes are mismatched, the sensitivity varies as a function of bandwidth. Bandwidth enhancement techniques that maintain relatively constant sensitivity (although reduced as compared with the matched-mode case) are an area of active research (Acar and Shkel 2005).

Table 2 lists the main performance metrics of the M^2-TFG and compares the values at matched-mode

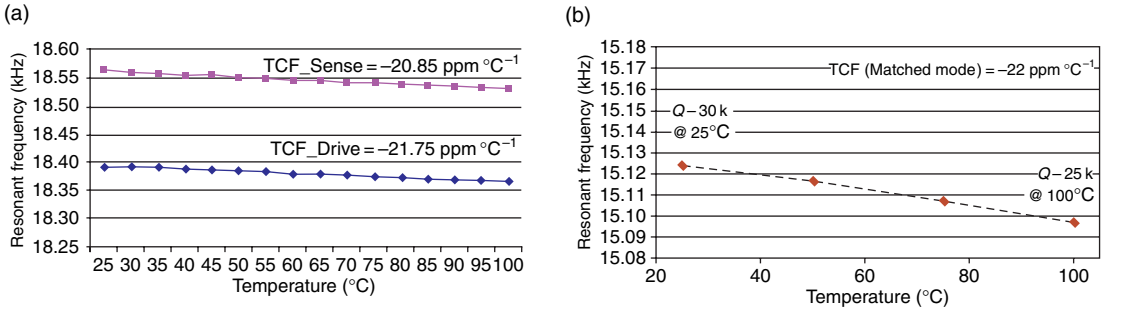


Figure 28 (a) Temperature coefficient of frequency (TCF) measurements of individual operating modes of a matched-mode tuning fork gyroscope (M^2-TFG) operating at 18.4 kHz and (b) TCF measurements of a device operating at a matched-mode condition.

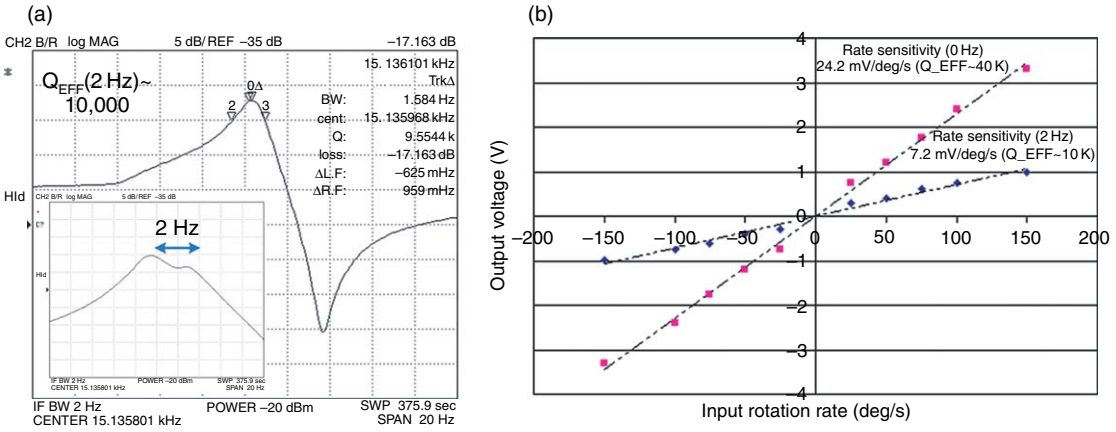


Figure 29 (a) Drive and sense modes matched to within 2 Hz and effective quality factor and (b) scale factor at 12 Hz separation where effective sense *Q* is 10 000.

Table 2 Performance comparison of the matched-mode tuning fork gyroscope (M^2-TFG) under two separate matching conditions and basic device parameters

Device parameter	*Mismatched*	*Matched*
Frequency split (Hz)	2	0
Effective sense quality factor	10 000	40 000
Theoretical Brownian noise floor (deg °h^{-1} $Hz^{-1/2}$)	1.1	0.5
Measured bias drift (at 25°C) (deg °h^{-1})	5.4	0.96
Rate sensitivity (mV deg^{-1} s^{-1})	7.2	24.2
Effective mass (μg)	30	
Drive amplitude (μm)	3	
Temperature coefficient of frequency (ppm °C^{-1})	−22	
Operating frequency (kHz)	15	

condition with those obtained when the modes are mismatched by about 2 Hz (corresponding to the case illustrated in **Figure 29**).

2.06.5 Future Trends in Micromachined Silicon Gyroscopes

Most of the research and development of silicon micromachined gyroscopes since the early 1990s has been driven by the automotive industry. The performance of micromachined gyroscopes has drastically improved over a rather short period, as illustrated in **Figure 30**, which is based on a sample of devices reported in the literature. Since 1991, the performance of micromachined gyroscopes, indicated by the ARW, has improved by a factor of 10 every 2 years. It is anticipated that a continuing improvement in the performance of micromachined gyroscopes (resolution, bias stability, dynamic range, etc.) will be seen over the next 10–20 years. As mentioned earlier, this will open up new avenues and market segments for micromachined gyroscopes, specifically inertial navigation for military and civilian applications.

Performance improvement is but one aspect of microgyroscope development. Considerable effort is under way for large-volume production of these micromachined sensors. Packaging and reliability are the two most important factors that must be considered before micromachined gyroscopes can be viably commercialized. The Analog Devices iMEMS® gyroscopes are hermetically sealed but operate at atmospheric pressure. Although this might suffice for rate-grade sensors, robust vacuum packaging will be required for navigation-grade gyroscopes. Additionally, to facilitate low-cost production of microgyroscopes, there is an increased emphasis on the design-for-manufacture and design-for-test.

The importance of robust vacuum packaging that can withstand temperature variations and shock has been recognized, and there is considerable research effort in that direction (Chae *et al.* 2005, Yoon *et al.* 2006). Researchers at the University of Michigan and Georgia Tech are developing hermetically packaged, environmentally protected, near-inertial-grade micromachined silicon gyroscopes. This work aims at developing wafer-scale micromachining technologies with high reliability, long lifetime, low power

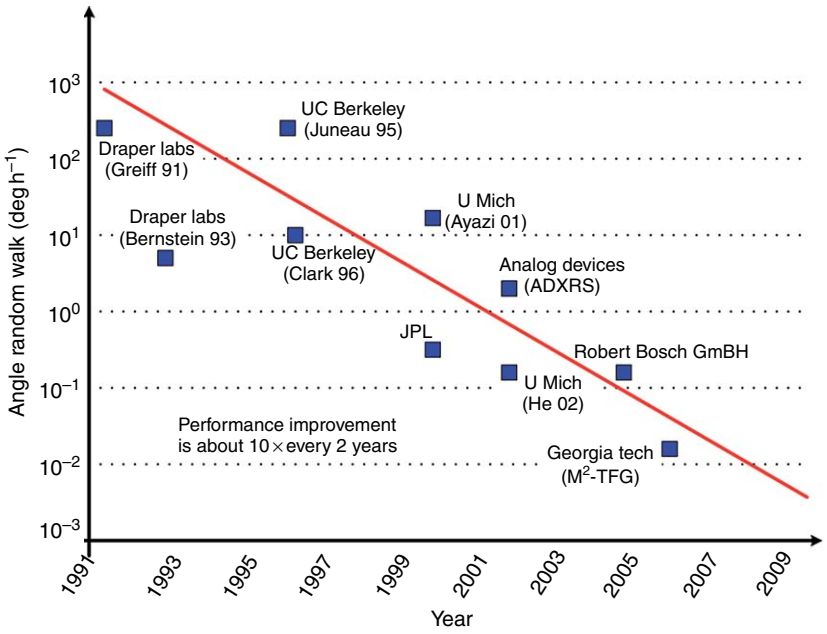


Figure 30 Performance of micromachined gyroscopes has improved by a factor of 10 every 2 years since 1991.

consumption, and high resistance to shock and environmental variations such as temperature.

While high aspect ratio micromachining provides significant improvement in sensor performance, it also enables the development of MEMS gyroscopes that can operate at lower voltages. The MEMS devices can then leverage the benefits offered by state-of-the-art CMOS processes and operate off a low-voltage battery.

References

Acar C, Shkel A 2005 Structurally decoupled micromachined gyroscopes with post-release capacitance enhancements. *J. Micromech. Microeng.* **15**, 1092–101

Ayazi F 2000 A high aspect-ratio high-performance polysilicon vibrating ring gyroscope. Ph.D. dissertation, The University of Michigan

Ayazi F, Najafi K 2000 High aspect-ratio combined poly and single-crystal silicon (HARPSS) MEMS technology. *IEEE/ASME J. Microelectromech. Syst.* **9(5)**, 288–94

Ayazi F, Najafi K 2001 A HARPSS polysilicon vibrating ring gyroscope. *IEEE J. Microelectromech. Syst.* **10**, 169–79

Bae S, Hayworth K, Shcheglov K, Yee K, Wiberg D 2002 JPL's MEMS gyroscope – Fabrication, 8-electrode tuning and performance results. *Tech. Dig.* Solid-State Sensors and Actuators Workshop 2002, Hilton Head Island, SC, USA

Barnaby R, Morrow C 1956 *US Patent* 2 753 173

Barnaby R, Reinhardt A 1951 *US Patent* 2 544 646

Bernstein J, Cho S, King T, Kourepenis A, Maciel P, Weinberg M 1993 A micromachined comb-drive tuning fork rate gyroscope. *Proc. IEEE Micro Electro Mechanical Systems Workshop (MEMS '93)*, Fort Lauderdale, FL, USA pp. 143–8

Boser B E 1997 Electronics for micromachined inertial sensors. *Tech. Dig. 9th Int. Conf. Solid-State Sensors and Actuators (Transducers '97)*, Chicago, IL, USA, pp. 1169–72

Burdess J, Wren T 1986 The theory of a piezoelectric disc gyroscope. *IEEE Trans. Aerosp. Electron. Syst.* **AES-22(4)**, 410–18

Chae J, Giachino J M, Najafi K 2005 Wafer-level vacuum package with vertical feed throughs. *Proc. IEEE Int. Micro Electro Mechanical Systems (MEMS '05)*, Miami, FL, USA, pp. 548–51

Clark W, Howe R, Horowitz R 1996 Surface micromachined z-axis vibratory rate gyroscope. *Tech. Dig. Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, pp. 283–7

Geen J, Sherman S, Chang J, Lewis S 2002 Single-chip surface micromachined integrated gyroscope with 50°/h Allan deviation. *IEEE J. Solid State Circuits* **37(12)**, 1860–6

Gianchandani Y, Najafi K 1992 A bulk silicon dissolved wafer process for microelectromechanical systems. *IEEE J. Microelectromech. Syst.* **1(2)**, 77–85

Gilbert B 1968 A precise far-quadrant multiplier with subnanosecond response. *IEEE J. Solid-State Circuits* **3(4)**, 365–73

Gomez U, Kuhlman B, Classen W, Bauer W 2005 New surface micromachined angular rate sensor for vehicle stabilizing systems in automotive applications. *Tech. Dig. Int. Conf. Solid-State Sensors and Actuators (Transducers '05)*, Seoul, South Korea pp. 184–7

Greiff P, Boxenhorn B, King T, Niles L 1991 Silicon monolithic micromechanical gyroscope. *Tech. Dig. 6th Int. Conf. Solid-State Sensors and Actuators (Transducers '91)*, San Francisco, CA, USA, pp. 966–8

Haim B 1998 *Analytical Dynamics*. McGraw-Hill, New York

Hao Z, Erbil A, Ayazi F 2003 An analytical model for support loss in micromachined beam resonators with in-plane flexural vibrations. *Sens. Actuators A. Phys.* **109**, 156–64

He G, Najafi K 2002 A single-crystal silicon vibrating ring gyroscope. *Proc. IEEE MEMS 2002*, Las Vegas, NV, USA pp. 718–21

IEEE Standards No. 952 1997 IEEE standard specifications format guide and test procedure for single axis IFOG

Jiang X, Seeger J I, Kraft M, Boser B E 2000 A monolithic surface micromachined z-axis gyroscope with digital output. *Proc. Symp. VLSI Circuits,* Honolulu, Hawaii, USA, 2000, pp. 16–19

Juneau T, Pisano A 1996 Micromachined dual input axis angular rate sensor. *Tech. Dig. Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, pp. 299–302

Keymeulen D, Fink W, Ferguson M, Peay C, Oks B, Terrile R, Yee K 2005 Evolutionary computation applied to the tuning of MEMS gyroscopes. *Proc. GECCO 2005*, Washington DC, pp. 927–35

Lawrence A 1993 *Modern Inertial Technology; Navigation, Guidance, and Control*. Springer-Verlag, New York

Lee S, Nguyen C T-C 2003 Influence of automatic level control on micromechanical resonator oscillator phase noise. *Proc. IEEE Frequency Control Symposium and PDA Exhibition*, Tampa, FL, USA, pp. 341–9

Lutz M, Golderer W, Gerstenmeier J, Marek J, Maihofer B, Mahler S, Munzel H, Bischof U 1997 A precision yaw rate sensor in silicon micromachining. *Tech. Dig. 9th Int. Conf. Solid-State Sensors and Actuators (Transducers '97)*, Chicago, IL, USA, pp. 847–50

Oh Y, Lee B, Baek S, Kim H, Kim J, Kang S, Song C 1997 A surface micromachined tunable vibratory gyroscope. *Proc. IEEE Micro Electro Mechanical Systems Workshop (MEMS '97)*, Nagoya, Japan, pp. 272–7

Park K Y 1999 Lateral gyroscope suspended by two gimbals through high aspect ratio ICP etching. *Tech. Dig. 10th Int. Conf. Solid-State Sensors and Actuators (Transducers '99)*, Sendai, Japan, pp. 972–5

Park K, Lee C, Oh Y, Cho Y 1997 Laterally oscillated and forcebalanced micro vibratory rate gyroscope supported by fish hook shape springs. *Proc. IEEE Micro Electro Mechanical Systems Workshop (MEMS '97)*, Nagoya, Japan, pp. 494–9

Pourkamali S, Hao Z, Ayazi F 2004 VHF single crystal silicon elliptic bulk-mode capacitive disk resonators: Part II: Implementation and characterization. *IEEE/ASME J. Microelectromech. Syst.* **13(6)**, 1054–62

Putty M W, Najafi K 1994 A micromachined vibrating ring gyroscope. *Tech. Dig. Solid State Sensors and Actuators Workshop 1994*. Hilton Head Island, SC, USA

Ragan R, Lynch D 1984 Inertial technology for the future, part X: Hemispherical resonator gyro. *IEEE Trans. Aerosp. Electron. Syst.* **AES-20(4)**, 432

Saukoski M, Aaltonen L, Halonen K, Salo T 2005 Fully integrated charge sensitive amplifier for readout of micromechanical capacitive sensors. *Proc. ISCAS 2005*, Kobe, Japan, pp. 5377–80

Sharma A, Zaman M, Ayazi F 2006 A 104 dB SNDR transimpedance-based CMOS ASIC for tuning fork microgyroscopes. *Proc. IEEE CICC 2006*, San Jose, CA, USA, pp. 655–9

Sharma A, Zaman M, Ayazi F 2007 A 0.2°/hr microgyro with automatic CMOS mode-matching. *Tech. Dig. ISSCC 2007*, San Francisco, CA, USA

Soderkvist J 1990 Design of a solid-state gyroscopic sensor made of quartz. *Sens. Actuators* **A21–A23**, 293–6

Sundaresan K, Ho G K, Pourkamali S, Ayazi F 2005 A two-chip, 4-MHz electromechanical reference oscillator. *Proc. ISCAS 2005*, Kobe, Japan, pp. 5461–4

Sung W T, Lee J Y, Lee J G, Kang T 2006 Design and fabrication of an automatic mode controlled vibratory gyroscope. *Proc. IEEE MEMS 2006*, Istanbul, Turkey, pp. 674–7

Tanaka K, Mochida Y, Sugimoto M 1995 A micromachined vibratory gyroscope. *Sens. Actuators A Phys.* **50**, 111–15

Voss R, Bauer K, Ficker W, Gleissner T, Kupke W, Rose M, Sassen S, Schalk J, Seidel H, Stenzel E 1997 Silicon angular rate sensor for automotive applications with piezoelectric drive and piezoresistive read-out. *Tech. Dig. 9th Int. Conf. Solid-State Sensors and Actuators (Transducers '97)*, Chicago, IL, USA, pp. 879–82

Watson W S, Henke T J 2002 Coriolis gyro configuration effects on noise and drift performance. *Symp. Gyro Technol.* Stuttgart, Germany

Xie H, Fedder G 2003 Fabrication, characterization, and analysis of a DRIE CMOS-MEMS gyroscope. *IEEE Sens. J.* **3**, 622–31

Yazdi N, Ayazi F, Najafi K 1998 Micromachined inertial sensors. *Proc. IEEE* **86(8)**, 1640–59

Yoon S W, Yazdi N, Chae J, Perkins N C, Najafi K 2006 Shock protection using integrated nonlinear spring shock stops. *Proc. IEEE MEMS 2006*, Istanbul, Turkey, pp. 702–5

Zaman M, Sharma A, Amini B V, Ayazi F 2004 Towards inertial grade vibratory microgyros: A high-Q in-plane silicon-on-insulator tuning fork device. *Tech. Dig. Solid-State Sensors, Actuators and Microsystems Workshop*, Hilton Head Island, SC, USA, pp. 384–5

Zaman M, Sharma A, Amini B, Ayazi F 2005 Resonating star gyroscope. *Proc. IEEE Int. Micro Electro Mechanical Systems (MEMS '05)*, Miami FL, USA, pp. 355–8

Zaman M, Sharma A, Ayazi F 2006 A high performance matched-mode tuning fork gyroscope. *Proc. IEEE MEMS 2006*, Istanbul, Turkey, pp. 66–9

Biographies

Farrokh Ayazi is an associate professor in the School of Electrical and Computer Engineering at the Georgia Institute of Technology. He received his BS degree in electrical engineering from the University of Tehran, Iran, in 1994, and his MS and Ph.D. degrees in electrical engineering from the University of Michigan, Ann Arbor, in 1997 and 2000, respectively. He joined the faculty of Georgia Tech in December 1999. Professor Ayazi's research interests are in the areas of integrated micro- and nanoelectromechanical resonators, RF MEMS, MEMS inertial sensors, VLSI analog circuits and techniques, and microfabrication technologies.

Professor Ayazi is a 2004 recipient of the NSF CAREER award, the 2004 Richard M. Bass Outstanding Teacher Award (determined by the vote of the ECE senior class), and the Georgia Technical College of Engineering Cutting Edge Research Award for 2001–2002. He received a Rackham Predoctoral Fellowship from the University of Michigan during 1998–1999. He serves on the technical program committee of the IEEE International Solid State Circuits Conference (ISSCC).

Mohammad Faisal Zaman received his BS degree (High Honors) in electrical engineering at the Georgia Institute of Technology (Atlanta, GA), in 2001. He is currently pursuing his Ph.D. degree at the IMEMS lab at the Georgia Institute of Technology. His doctoral research involves the design, fabrication, and analysis of high-precision micromachined gyroscopes. He has held internship positions with General Electric and Sprint-Nextel. He is a member of the Eta Kappa Nu and IEEE.

Ajit Sharma received his BE (Hons.) degree in electrical engineering from the Birla Institute of Technology and Science, Pilani, India, in 2001 and his MS degree in electrical engineering from Oregon State University, Corvallis, OR, in 2003. His masters' research focused on the prediction of substrate noise coupling in mixed-signal SOCs and methodologies for accurate substrate parasitic extraction in CMOS substrates. He is currently pursuing his Ph.D. degree at the IMEMS lab at the Georgia Institute of Technology, Atlanta, GA. His doctoral work involves the design of low-power and low-noise CMOS analog interface circuits for high-precision MEMS gyroscopes and modeling of microsystems. He has held co-op positions at Infineon Technologies, Singapore, and Biotronik Inc., Lake Oswego, OR. He received a Texas Instruments Analog Fellowship during 2003–2006.

2.07 Flow Sensors

Sjoerd Haasl[1] **and Göran Stemme**[2], [1]Imego AB, Gothenburg, Sweden, [2]Microsystem Technology, School of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden

Glossary

AC Alternating Current
BiCMOS Integrated circuit that uses both bipolar and CMOS technologies
BJT Bipolar Junction Transistor
CC Constant Current
CCCV Constant Current, Constant Voltage
CMOS Complementary Metal Oxide Semiconductor
CNT Carbon Nanotube
CP Constant Power
CT Constant Temperature
CTA Constant Temperature Anemometry
CTD Constant Temperature Difference
CV Constant Voltage
DC Direct Current
DOE Diffractive Optical Element
EHD Electrohydrodynamic

ISFET Ion-Sensitive Field Emission Transistor
LDA Laser Doppler Anemometry
LIGA *Litografi Galvanisierung Abformung* – a combination of deep-etch X-ray lithography, electroforming, and molding techniques
LTCC Low-Temperature Cofired Ceramics
MESFET Metal–Semiconductor Field-Effect Transistor
MIS Minimally Invasive Surgery
MOS Metal Oxide Semiconductor
MWCNT Multiwalled Carbon Nanotube
PBS Phosphate-Buffered Saline
PDMA Plastic Deformation Magnetic Assembly
PIV Particle Image Velocimetry
PMMA Polymethylmethacrylate
PMOS Positive-Channel MOS
PWM Pulse Width Modulation
RF Radio Frequency
SAW Surface Acoustic Wave
SOI Silicon on Insulator
SWCNT Single-Walled Carbon Nanotube
TBA Temperature-balance anemometry
TCR Temperature Coefficient of Resistance
TTL Transistor–Transistor Logic

2.07.1 Introduction

The flow of fluids has always interested mankind. From rudimentary weather predictions based on the direction of the winds to the simulation of fluid flow over a jet foil, measuring and modeling flow has played a central role in predicting its behavior and its effects on the surroundings. Even prior to the emergence of microsystem technology, a wealth of different flow sensors based on varying principles was developed. The advent of micromachining has allowed the miniaturization of existing sensors and resolutions that were not possible before. The batch manufacturability of many micromachined flow sensors makes them suitable for commercialization. Reproducibility has opened up the field of array sensing at relatively low cost, offering the possibility of obtaining an instantaneous representation of a complex flow. New sensor principles have emerged as well that take advantages of new materials such as carbon nanotubes (CNTS). Over the past few decades, an entirely new field, microfluidics, has emerged and with it the need for flow sensing on microscales. But not all of the problems have been solved, and no one sensor is applicable in all situations.

The goal of this chapter is to provide an overview of the micromachined flow sensors presented in the literature to date. The sensors are classified according to the domain in which they function and their operating principle. Due to the close relationship between micromachining and integrated circuits, a section is devoted to on-chip electronics. Two applications are also discussed.

The study of fluids in motion, fluid dynamics, is a complex field and plays an important role in the design of flow sensors. The large theoretical background to this field will be touched upon only when deemed necessary to clarify the flow-sensing principles presented. The theory behind the principles used to measure the flow, such as heat transfer, will also be introduced but not covered in detail. For more in-depth information on these subjects, references to the specialized literature have been provided.

2.07.2 Thermal Flow Sensors

Among micromachined flow sensors, thermal flow sensors make up the largest group, accounting for more than half of all the sensors described in the literature. They are also the oldest type of micromachined flow sensors, as a natural evolution from the integrated circuits, with which the observation of the air cooling of a simple heated resistor is sufficient to obtain a measurement of the flow. In this section, some basic fluid dynamics and heat transfer theory are introduced to serve as reference for the subsequent discussions on thermal flow sensors. At the end of the section, an overview table covering thermal flow sensors is given.

2.07.2.1 Laminar and Turbulent Flow

Figure 1 shows a time-averaged 1D flow profile near a wall. U is the speed of the flow in the direction of the flow, and y is the direction normal to the surface. The flow in the figure is the average of a flow that fluctuates with time. A measure for the amount of fluctuation, in other words the turbulence, is the Reynolds number. The Reynolds number at a certain

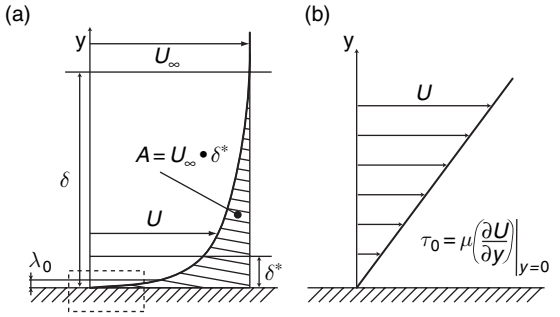


Figure 1 Time-averaged 1D flow profile near a wall. (b) Close-up of the region in the square of (a): the viscous sublayer, where the flow velocity is linearly dependent on the distance to the wall.

point is defined as $Re \equiv UD/\nu$, where U is the fluid velocity at that point, D is the characteristic length, for example, the distance of the point to the wall, and ν is the kinematic viscosity of the fluid (defined below). This dimensionless number is the ratio between the inertial and viscous forces in a fluid and is a measure of how turbulent a fluid flow is at a certain point. Roughly, one can say that for Reynolds numbers less than 10^3, the flow is not turbulent, that is, it is laminar, and for Reynolds numbers greater than 10^5, the flow is completely turbulent.

For a straight, parallel, and uniform flow, the shear stress, τ, between layers is proportional to the velocity gradient in the direction perpendicular to the layers:

$$\tau = \mu \frac{\partial U}{\partial y}$$

This is Newton's law of viscosity, and the proportionality constant μ is called the dynamic viscosity. Most gases as well as many liquids follow this law of viscosity. The kinematic viscosity, ν, is defined as the quotient of the fluid's dynamic viscosity and its density:

$$\nu = \frac{\mu}{\rho}$$

When talking about fluids near a wall, the term boundary layer is often used to define the layer close to the wall where a certain quantity, such as velocity or temperature, changes the most. The thickness of the layer is defined as an arbitrary value, e.g., the edge of the hydrodynamic boundary layer is often defined as the position off the wall where the fluid velocity is 99% of the free-stream velocity.

2.07.2.2 Heat Transfer

When a sensor is placed in a fluid with a different temperature, it will exchange heat with the fluid and with its attachment point. There are three mechanisms of heat transfer: conduction, convection, and radiation. Conduction is defined as the heat transfer from molecule to molecule. For an isotropic and homogenous medium, the behavior of the temperature T due to conduction can be described as follows:

$$\frac{\partial T}{\partial t} = \kappa \nabla^2 T$$

where κ is the thermal diffusivity, a material-specific quantity dependent on the material's thermal conductivity k, density ρ, and specific heat capacity c_p

$$\kappa = \frac{k}{\rho c_p}$$

Convection of the fluid increases the effect of conduction, as molecules with larger temperature differences are brought into contact with each other. Mathematically, this can be described by adding a convection term to the heat transport equation:

$$\frac{\partial T}{\partial t} + U \nabla T = \kappa \nabla^2 T$$

For sensors, the practical effect of convection is that, at the surface of a body in a fluid, heat transfer, Q, occurs that can be described as follows:

$$Q = hA(T_b - T_f)$$

where A is the surface area of the body, T_b the temperature of the body, T_f the temperature of the fluid, and h the heat transfer coefficient. The heat transfer coefficient is dependent on the fluid as well as on the flow. Since the flow velocity is zero at the wall, the heat transfer into the fluid is defined by conduction and can also be expressed as a function of the temperature gradient at the wall ($y = 0$) (Holman 1997):

$$\frac{Q}{A} = h(T_b - T_f) = -k \frac{\partial T}{\partial y}\bigg|_{y=0}$$

A dimensionless parameter used to characterize the enhancement of the heat transfer due to convection is the Nusselt number. It is the ratio of heat transferred from a surface to heat conducted away by the fluid. Defined as $Nu \equiv hL/k$, where L is the characteristic length scale and k is the thermal conductivity of the fluid, the Nusselt number is often correlated with the Reynolds and Prandtl numbers.

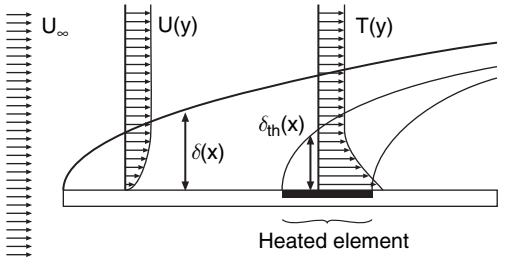


Figure 2 Typical average velocity and temperature profiles for a plate with a heater placed in the flow. $\delta(x)$ is the hydrodynamic boundary layer thickness, $\delta_{th}(x)$ is the thermal boundary thickness.

The Prandtl number is defined as the ratio of the fluid velocity boundary layer thickness to the fluid temperature boundary layer thickness and can be expressed as $\mathrm{Pr} = \mu c_p / k$.

The third mechanism for heat transfer is radiation. The amount of heat, P_{rad}, that is radiated follows the Stefan–Boltzmann law:

$$P_{\mathrm{rad}} = e\sigma A T^4$$

where σ is the Stefan–Boltzmann constant, A the area of the radiating body, T the temperature, and e the emissivity, a value between 0 and 1 where 1 is an ideal radiator. Since the surroundings (T_∞) also emit radiation, the net heat transfer of a body is $P_{\mathrm{net}} = e\sigma A(T^4 - T_\infty^4)$. Most often, for micromachined flow sensors, the heat emitted by radiation can be neglected as it is much smaller than the heat losses due to conduction and convection.

Figure 2 shows the typical average velocity and temperature profiles for a plate with a heater placed into the flow. The illustrated case is of the common situation where the Prandtl number is approximately 1 or greater, resulting in the thermal boundary layer being completely enveloped in the hydrodynamic boundary layer. This is the case for gases (where $Pr \sim 1$) and most liquids (water 6.75, glycerol 7250), with the exception of liquid metals (e.g., mercury 0.044) (Elwenspoek and Wiegerink 2001).

2.07.2.3 Thermal Flow Sensor Principles

Thermal flow sensors can be classified in function of three different principles: thermal anemometers, calorimetric flow sensors, and time-of-flight flow sensors. Figure 3 illustrates the principles.

The defining feature of thermal anemometers is that the flow is measured by its cooling effect on a heated entity. For calorimetric flow sensors, a temperature sensor is placed separate from the heated entity yet still within the thermal boundary layer of the heated body, and the heating effect of the flow is measured. By placing a sensor both upstream and downstream of the heater, the flow direction can also be measured. Time-of-flight flow sensors have the temperature sensors placed outside the thermal boundary layer of the heater and measure the time it takes for a thermal pulse to reach the sensors. Like calorimetric flow sensors, time-of-flight flow sensors can be made direction sensitive by adding sensing elements.

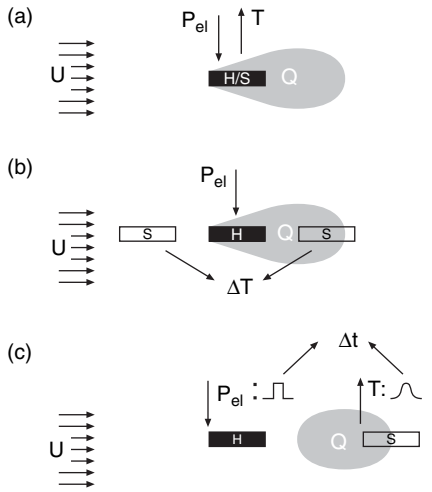


Figure 3 The three principles used in thermal flow sensing: (a) thermal anemometry, (b) calorimetric flow sensing, and (c) time-of-flight flow sensing (H, heater; S, sensor). (Source: Ashauer M, Glosch H, Hedrich F, Hey N, Sandmaier H, Lang W 1999 Thermal flow sensor for liquids and gases based on combinations of two principles. *Sens. Actuators A (Phys.)* **A73(1–2)**, 7–13. © 1999 Elsevier.)

An ideal thermal flow sensor is sensitive only to the thermal effects of the flow. All sensors, however, need to be supported in some way. Therefore, there is always some heat conduction through the supporting structure. In the worst case, the amount of heat conducted away is dependent on the flow outside the volume that the sensor is intended to measure. This results in reduced resolution and nonlinearities.

2.07.2.3.1 Thermal anemometers (alternative name: heat-loss flowmeters)

Depending on the heated element in a thermal anemometer, sensors of this type are called hot-wire, hot-chip, or hot-film anemometers. Conventional

hot-wire anemometers have been in use for more than a century and range from 1 µm diameter and 0.25 mm length upward (Bruun 1995). Though microsystem technology has enabled them to be miniaturized even more, most micromachined thermal flow sensors are of similar dimensions as the smallest conventional thermal flow sensors. They provide other advantages such as the possibility to form arrays to integrate control electronics on the sensor chip. **Table 2** provides an overview of micromachined anemometers.

In 1914, King (1914) derived an equation for infinitely long hot-wire anemometers in low Reynolds number flows that describes the relationship between the dissipated power (Q_{diss}) and the free-stream velocity (U). This equation is referred to as King's law:

$$Q_{\text{diss}} = A + B\sqrt{U}$$

where A and B are functions of the temperature and the dimensions of the wire. A dimensionless expression for the same equation is $Nu = A' + B'\sqrt{Re}$. For wires of finite length and with a linear temperature dependence on resistivity, the resistance at temperature T_w can be written as $R_w = R_0(1 + \alpha_0(T_w - T_0))$ where R_0 is the resistance and α_0 the temperature coefficient of resistance (TCR), both at temperature T_0. Let $V_w = R_w I_w$ be the voltage over the heated wire, then

$$\frac{V_w^2}{R_w} = (A + BU^n)(T_w - T_0)$$

where the exponent n differs from 0.5 due to the finite length of the wire (Bruun 1995).

A hot wire does not have an ideal single-axis flow direction sensitivity. For a circular wire, suspended by a pair of prongs that do not interfere with the flow, the response for flow components in the plane normal to the wire will be independent of the direction. The tangential component of the flow will give a small response as well (i.e., for wires with a length short of infinity). Let U_N be the flow's component normal to the hot wire and parallel to the prongs suspending it, U_T the component tangential to the hot wire, and U_B the binormal component perpendicular to the first two. Then the effective velocity, U_e, that the wire will measure can be described by Jørgensen's equation (Jørgensen 1971):

$$U_e^2 = U_N^2 + k_T^2 U_T^2 + k_B^2 U_B^2$$

where k_T and k_B are probe-dependent coefficients. For a standard plated conventional hot wire, $k_T = 0.2$ and $k_B = 0.2$ (Bruun 1995). For further information on hot wires, the interested reader is referred to the specialized literature and references therein (Bruun 1995).

Micromachined thermal anemometers differ from each other in the heater and sensor type, thermal insulation, whether or not they compensate for fluid temperature, and in the mode in which they are run. Nearly all thermal anemometers use some kind of resistor as a heating element. The most straightforward and common heating element is a doped silicon resistor. The high resistivity of silicon enables heating with low currents. Different metals have also been used, platinum being the most common. For measuring the temperature of the heated element, often the same resistor is used. A high TCR is an advantage. **Table 1** provides a short overview of the TCR values of metals commonly used for flow sensing.

Alternative temperature sensors used for thermal anemometers are thermopiles (e.g., n/p Si or Al–Si) and diodes. An advantage of thermopiles is that they measure the temperature difference. Assuming that the bulk of the sensor has the slowly varying temperature of the fluid, the temperature of the heated element can be kept more easily at a constant temperature difference (CTD) with the fluid, which has certain advantages over constant temperature (CT) (see below). Diodes are used for the linear temperature dependence of the threshold voltage (MacHattie 1979). Their advantage over implanted or diffused resistors lies in their temperature behavior being almost independent of the fabrication process (van der Wiel *et al.* 1993). Among the more exotic sensing variants, one can count capacitive sensors, where the temperature dependence of the dielectric's permittivity is used (Kwok *et al.* 1996, Lin *et al.* 1996), and surface acoustic wave (SAW) sensors, where temperature changes cause a resonance frequency shift (Ahmad 1985, Joshi 1994, Rebiere *et al.* 1994).

Table 1 TCR values of metals commonly used for micromachined thermal flow sensors

Metal	*TCR (Goodfellow, 2007) (‰ K^{-1})*
Pt	3.92
Au	4.0
Mo	4.35
Ni	6.8
Ta	3.5
Al	4.5
Ti	3.8
Cr	2.14

Table 2 Overview of thermal anemometers

Reference	*Heater*[a]	*Sensor*	*Insulation*[b]	*Fluid temp Compensation*	*Fluid type*	*Mode*[c]	*Sensitivity*[d,e,f]	*Sensitivity range*[g]	*Measured range*[h]	*Resolution*	*ΔT/I/V/P*[i]	*Time constant*[j]
Ahmad (1985)	Thin-film resistor	SAW	None, hot chip: 128° $LiNbO_3$	No	Air	CP	11 Hz ccm^{-1} $= 7.7 \times 10^{-8}$ sccm^{-1}		2000–4000 ccm in 4″ × 5″ cross section		49°C	
Al Khalfioui *et al.* (2003)	Semiconductive alloy	Thermopile (n-alloy/ p-alloy: Bi_2Te_3-Sb_2Te_3, Bi_2Te_3-Sb_2Se_3)	PI substrate	No	Air	CC (AC)	Laminar: $V = 0.9$ mV (m s^{-1})$^{-0.7}$ Turbulent: $V = 1.07$ mV (m s^{-1})$^{-0.1}$					
Betzner *et al.* (1996)	Monocrystalline Si	Same	SiO/air/SiN	Yes: upstream resistor	Air	Full Wheatstone bridge	61 mV m^{-1} s		0–4.6 m s^{-1}			5 ms
Bouwstra *et al.* (1989)	Polysilicon resistor	Polysilicon piezo resistor	SiN bridge	No	Air	Resonant	0.8 kHz at 10 sccm				85 kHz, 20 K	
Bouwstra *et al.* (1990)	Polysilicon resistor	Polysilicon piezo resistor	SiN membrane	No	Air	Resonant	13 Hz at 10 sccm				5 kHz, 8 K	
Brace *et al.* (1989): single oscillator	SAW (surface acoustic wave)	SAW	None, hot chip: $LiNbO_3$ substrate	No	Nitrogen	CP	$\Delta f = 340$ Hz sscm^{-1}, $\Delta f/f = 4.67 \times 10^{-6}$ sccm^{-1}		0–500 sccm		47.5°C	10–30 s
Bruschi *et al.* (2004a)	PTAT and copper cylinder into flow	Same	None, hot-chip (PCB)	No	Water	CP and CT pulse	*0.90 V (l h^{-1})$^{-0.5}$* *0.43 V (l h^{-1})$^{-0.5}$*	0–6 l h^{-1} 6–30 l h^{-1}	30 l h^{-1}		1.5 K and 5 K	
Bruschi *et al.* (2004b)	Polysilicon resistor	Thermopile (n-poly silicon/ p-poly silicon)	SiO cross membrane	Yes: thermopile	Gas	CP	*0.46 mV sccm$^{-0.5}$*	10–140 sccm at 90° angle	10–140 sccm			
Bu *et al.* (1996)	Pt/Ta resistor	Same	SiN and SiO membrane over air	Yes: resistor on bulk	Air	Not specified	$\Delta V/V = -30\%$ at 1.2 m s^{-1}	0–0.3 m s^{-1}	0–1.2 m s^{-1}			
Cain *et al.* (2000); Sheplak *et al.* (2002)	Ti/Pt heaters resistors on membrane	Same	Vacuum or N_2 below SiN membrane	No	Air	CT	11 mV Pa^{-1} (Cain)	9 μPa–1.7 Pa at overheat 1.0				<1 ms
Castaner *et al.* (1997)	Diffused resistor ($POCl_3$) and NiCr resistor	Diode string from bipolar transistors (Darlington stack)	None: hot chip/ Plastic epoxy between hot and cold chip	Yes	Water	CP and CT (CTD)	Drop: 40 to 32 mV Over 0–9 l min^{-1}		0–9 l min^{-1} in 11 mm diam pipe		~1K	
Chen and Lal (2001)	Polysilicon resistor	Same	SiN membrane over air	No	Air	Wheatstone bridge	0.36 V ml^{-1} s at 8 V bridge voltage		0.25–1.25 ml s^{-1}			
Chen and Liu (2003)	Free Ni hotwire	Same	Supports: PI/Permalloy	No	Air	CC and CT	CC: $\Delta V = -50$ mV at 20 m s^{-1} CT: $\Delta V = 1.5$ V at 20 m s^{-1}		0–20 m s^{-1}		OH = 0.3	88 μs
Chen *et al.* (2005)	Ni resistor	Same	Supports: PI/ Permalloy	No	Air	CT	5 mV m^{-1} s		0–20 m s^{-1}			
	Pt heater wire	Same	PI beam	No	Water			1–500 μl min^{-1}	1–2200 μl min^{-1}			

Dittmann *et al.* (2001)						CC Wheatstone bridge	*1 mV µl^{-1} min* 0.9 V at 2200 µl min^{-1}					*7 ms (N2), 16(H_2O) (99%)*
Dittmann *et al.* (2001)	Pt heater wire	Same	PI beam	No	Nitrogen	CC Wheatstone bridge	8 V at 500 sccm		0.1–500 sccm			
Dominguez *et al.* (1993)	Implanted resistor	Diode	Oxidized porous Si	Yes: diode on bulk	Gas	CP (not explained)	50.9 µW K^{-1} (m s^{-1})$^{-0.5}$	50.9 µW K^{-1}(m s^{-1})$^{-0.5}$	0.1–>2.5 m s^{-1}			*300 ms (99%)*
Dominguez *et al.* (2005)	n^{+} Si diffused resistor	Diode string from bipolar transistors (Darlington stack)	Two hot chips – Pyrex support and PCB	Yes	Air/smoke	CP and CT (CTD) SD	Not specified (non-linear)		0–0.8 m s^{-1} and 0–5 m s^{-1} in 5 mm diam. side tube of 16 cm diam. pipe			
Ebefors *et al.* (1998)	p^{++} Polysilicon resistor	Same	Si prongs and PI grooves	No	Air	CP (Manual!)	0.1°C (l^{-1} min)		0–60 l min^{-1} (16 mm×16 mm)		5.77 mW	120 µs
Ebefors (2000)	p^{++} Polysilicon resistor	Same	Si prongs and PI grooves	No	Air	CT	10.4 V^{-2} (m s^{-1})$^{-0.5}$	2–15 m s^{-1}	2–15 m s^{-1}		80 K	30 µs
Esashi (1991)	Ti resistor	Same	SiO on air on glass	Yes: upstream Ti resistor	Gas	CC / Wheat stone	1.37 mW sccm$^{-0.5}$				260°C	
Fricke (1994)	AlGaAs resistor	GaAs/AlGaAs thermopile	AlGaAs membrane	Yes: diode on bulk	Air	CT	0.05–0.15 W m^{-1} s	0–25 m s^{-1}	0–25 m s^{-1}		7–27 K (2.7–10 mW)	
Friedberger *et al.* (2001)	SiC resistor	Same	Air under resistor	Yes: resistor on bulk	Air	CC[*l*]	2.5 mV m^{-1} s		0–3.5 m s^{-1}		240 mW	
Gajda and Ahmed (1995)	n-doped diffused resistors	Same	SiO cross membrane	No	Air	CT	26 mV at 11 m s^{-1}		0.25–11 m s^{-1}		250°C	*15 s (99%)*
Gamage *et al.* (2000)	Mono-crystalline Si	Same	SiO/air/SiN	No	Air	CC, pulse creation	*67 mV m^{-1} s*		2–21 m s^{-1}			80–100 ms (pulse measurement time)
Huang and Tong (1989)	MOST + diffused resistor	Diode	None, hot chip	No	Nitrogen, water	CT	N_2: 500 mV at 50 cm s^{-1} H_2O: 50 mV at 15 cm s^{-1}		N_2: 0–50 cm s^{-1} H_2O: 0–15 cm s^{-1}		15K	40 s
Huang *et al.* (1995)	Polysilicon resistor	Same	SiN on vacuum (300 mT)	No	Air	CT	70 mV Pa^{-1}					72 µs with cav.
Huang *et al.* (1996)	Polysilicon resistor	Same	SiN on vacuum (300 mT)	No	Air	CT	10 mV Pa^{-1}					72 µs with cav. 5 µs w/o cav.
Hung *et al.* (2000)	Pt resistor	Same	SiN membrane with holes	No	Air	CV and CC	CV: 46 µA(m s^{-1})$^{-0.5}$ CC: 27 mV(m s^{-1})$^{-0.5}$	1–11 m s^{-1}			CV: 4 V CC: 23 mA	
Jiang *et al.* (1994b)	Polysilicon resistor	Same	Polysilicon and Si prongs	No	Air	CT	250 mV at 30 m s^{-1}		30 m s^{-1}		2–3 mW	2 µs
Jiang *et al.* (1995)	Polysilicon resistor	Same	SiN on vacuum	No	Air	CT	*13 mV/V Pa$^{1/3}$*		0–0.7 Pa			
Jiang *et al.* (1997)	Polysilicon resistor	Same	SiN on vacuum	No	Air	CT	100 mV Pa^{-1} at OH = 1.1 ~ *2.5 V^2/Pa$^{1/3}$*		0.8–1 Pa			
Joshi (1989)	Thin-film resistor	SAW	None, hot chip: Y-Z $LiNbO_3$ substrate	No	Nitrogen	CP	2.38 × 10^{-6} sccm^{-1}		0–1000 sccm	0.04 sccm	469 mW	*50 s (95%)*
Joshi and Jin (1990)	Thin-film resistor	SAW	None, hot chip: $LiNbO_3$ substrate	No	Water	CP	42.2 kHz at 0.8 ml min^{-1}		0–0.8 ml min^{-1}	0.06 µl min^{-1}	103 mW (~9 K)	*70 s (90%)*

(*Continued*)

Table 2 (Continued)

Reference	*Heater*[a]	*Sensor*	*Insulation*[b]	*Fluid temp Compensation*	*Fluid type*	*Mode*[c]	*Sensitivity*[d,e,f]	*Sensitivity range*[g]	*Measured range*[h]	*Resolution*	*ΔT/I/V/P*[i]	*Time constant*[j]
Joshi (1994)	Thin-film resistor	SAW	None, hot chip: $LiNbO_3$ substrate	No	Nitrogen	CP	140 kHz at 1000 sccm		0–1000 sccm			
Kälvesten *et al.* (1995)	Polysilicon resistor	Diode	PI on Si membrane	Yes: diode on bulk	Air	CT	1.22 mW $Pa^{-0.5}$	0–4 Pa				7 ms/25 μs in CT
Kang and Park (2000)	Pt resistor	Same	Ceramic substrate	Yes: Pt resistor at distance	Air	CT	0.14 V m^{-1} s	5–12 m s^{-1}	0–12 m s^{-1}			
Kersjes *et al.* (1993)	Polysilicon resistor	Diode	0.2 μm Si Membrane	Yes: diode on bulk	Oil, air	CC	0.4 μs l^{-1} min	0–10 l min^{-1}	0–0.8 m s^{-1}	0.4 μs l^{-1} min	Oil: 8–27 K Air: 16–66 K	
Kersjes and Mokwa (1995)	Polysilicon resistor	Diode	Si membrane with SiO trench	Yes: diode on bulk	oil, water, base, acid	CV and CT	Acid: 0.024 mV^{-1} $(m\,s^{-1})^{1/3}$ Oil: 0.01 mV^{-1} $(m\,s^{-1})^{1/3}$	0–100 l h^{-1}	0–100 l h^{-1}			10 ms
Kersjes *et al.* (1996)	Polysilicon resistor	Diode	Si membrane with SiO trench	Yes: diode on separate membrane	Liquid not specified	CC	ΔPW = 25 μs at 0.5 m s^{-1}	0–0.5 m s^{-1}	0–0.5 m s^{-1}		5 K	
Kim *et al.* (2004)	Pt resistor	Pt resistor	Si membrane (30 μm)	No	Air	CT	0.32 V at 10 m s^{-1}		0–10 m s^{-1}	0.5 m s^{-1}		
Kim and Lee (2006)	Au resistor	Same	SiN suspended channel	No	Air	CT	200–800 mV Pa^{-1}		0–0.83 Pa			
Kuttner *et al.* (1991); Kohl *et al.* (1994)	Ge resistor	Same	Glass substrate	Yes: Ge resistor upstream	Kuttner: blood, Kohl: water, ethanol, hydrogen, helium, Nitrogen, argon	CT (CTD!!)	*ΔP/P = 20% (mm^{-1} s)*	0.05–0.25 mm s^{-1}	0.05–0.8 mm s^{-1} in 1.8 mm diam tube		0.3–2 K	
Kwok *et al.* (1996)	Polysilicon resistor	Capacitance: ε(T) of $SrTiO_2$	SiO membrane	No	Nitrogen	CP	118 Hz m^{-1} s	0–5 m s^{-1}	0–20 m s^{-1}		250 mW	
Legtenberg *et al.* (1991)	Polysilicon resistor	Polysilicon piezo resistor	SiN bridge	No	Air	Resonant	$T < T_b^k$: +4 kHz $sccm^{-1}$ $T > T_b$: 7 kHz $sccm^{-1}$	0–30 sccm				
Li *et al.* (2005)	Polysilicon resistor	Same	Dielectric membrane on air cavity	No	Water	CC	0.19 K at 20 nl^{-1}			1.3 n ls^{-1} ~4.2 mm s^{-1}	0.18 mW	
Lin *et al.* (1996)	Polysilicon resistor	Capacitance: ε(T) of Ta_2O_5	SiO cantilever	No	Nitrogen	CP	4.25 mV m^{-1} s	0–4 m s^{-1}	0–20 m s^{-1}			
Löfdahl *et al.* (1989b)	Implanted resistor	Diode	PI groove	Yes: diode on bulk	Air	CT	0.67 V^2 $(m\,s^{-1})^{-0.5}$					50 ms
Löfdahl *et al.* (2003)	Free Al hotwire	Same	Supports: Si prongs	No	Air	CT	100–300 mV Pa^{-1}		0–2.5 Pa		OH = 1.5	
Ma *et al.* (1995)	Au/Si resistor	Same	Glass substrate	No	Air	Full Wheatstone bridge	*5.8 mV cm^{-1} s*				2 mA	~μs
MacHattie (1979)	Transistor	Same	None, hot chip	No	Air	CP and CT	9–53% at 2 m s^{-1} 4–27% at 4 m s^{-1}		0–4 m s^{-1}			
Mailly *et al.* (2001)	Pt resistor	Same	Cross SiN Membrane	No	Air	CT	4.8 mV $(m\,s^{-1})^{0.45}$ mW^{-1}		0–20 m s^{-1}		18.3 mW	6 ms

Mastrangelo and Muller (1988)	Polysilicon resistor	Same	PSG	Yes: eight resistors on bulk around	Air	CT	35 mV at 100 sccm		0–100 sccm		8 mW (250°C)	25 ms[l]
Mizuno *et al.* (2004)	Carbon resistor	Same	Parylene suspended channel	No	Water	CC	380 μV nl^{-1} min	3–12 nl min^{-1}	3–12 nl min^{-1}		1.7 μA (28 μW)	
Moser *et al.* (1991)	Polysilicon resistor	Al/polysilicon and Al/p-Si thermopiles	SiO/SiO/SiN beam	Yes: thermopile cold junction on bulk	Nitrogen	CP	CMOS: 1.78 V W^{-1} (m s^{-1})$^{-0.5}$ Bipolar: 0.26 V W^{-1} (m s^{-1})$^{-0.5}$	1–25 m s^{-1}	0–25 m s^{-1}			
Neda *et al.* (1996)	Polysilicon resistor	Same	SiO membrane	No	Air	CT	22 mV (m s^{-1})$^{-0.5}$		0.05–35 m s^{-1}		6 mW (122°C)	*0.14 ms (90%)*
Pagonis *et al.* (2004)	Polysilicon resistor	Al/Si thermopiles	Porous Si membrane (10 μm) over air cavity (10 μm)	Yes: Al/Si thermopile	Nitrogen	CT	14 mV (m s^{-1})$^{-0.5}$ W^{-1}	2–10 m s^{-1}	0–10 m s^{-1}			
Papageorgiou *et al.* (2001)	Implanted resistor	Same	Dielectric membrane on air cavity	No	Water	CC[l]	*−0.062 K at 0.032 m s^{-1}*		0–0.032 m s^{-1}	150 pl s^{-1}	<2 K	200 μs in water
Perichon *et al.* (2000)	Polysilicon resistor	Al/Si thermopiles	Oxidized porous Si	Yes: polysilicon resistor on bulk and thermopile cold junction on bulk	Intention: blood		No measurements presented					<1 s
Rebière *et al.* (1994)	Thick-film Pt screen-printed resistor	SAW	None, hot chip	Not specified	Gas	Not specified	SAW:7 kHz l^{-1} min SH-APM: 30 kHz l^{-1} min	SAW: 0.2–0.8 l min^{-1} SH-APM 0–0.8 l min^{-1}	0.8 l min^{-1}	10 μl s^{-1} at 0.9 W	<100°C SAW: 246 W SH-APM: 26 W	
Rouhanizadeh *et al.* (2006)	Polysilicon resistor	Same	SiN membrane on air	No	Blood	CT	*0.018 V^2 Pa$^{-1/3}$*	*2–12 Pa*	*1–5 ml min^{-1}, 2–12 Pa*		"Low overheat ratio"	
Schmid (2002)	Mo	Same[l]	LTCC substrate	No	Fuel	CT	2.7 W at 10 m s^{-1}		0–10 m s^{-1}		120 K	*0.4 ms*
Stemme (1986)	Implanted resistor	Diode	PI groove	Yes: diode on bulk	Air	CT	9.8 mW (m s^{-1})$^{-0.5}$				100 K	
Stemme (1988)	Implanted resistor	Diode	PI groove	Yes: diode on bulk	Air	CC, pulse modulation	ΔPW = 3% m^{-1} s	2–30 m s^{-1}	2–30 m s^{-1}		110 mW	
Stephan and Zanini (1991)	Ni resistor	Same	Si bridges	Yes: resistor on separate bridge	Air	CTD	*5.8 V (m s^{-1})$^{-0.5}$*	0–52 m s^{-1}	0–60 m s^{-1}	3.5%	180°C	
Sundeen and Buchanan (2001)	Ni-ZrO_2	Same	None, hot chip	No	Air		*−0.67°C (cm s^{-1})$^{-0.5}$*				~20 s	*30–40°C*
Svedin *et al.* (2003a)	Polysilicon resistor	Same	PI-filled groove	No	Air	CT	2.8 V^2 (m s^{-1})$^{0.31}$		0–47 l min^{-1} in 40 mm diameter tube		120 K	
Tabata (1986); Tabata *et al.* (1987)	Pt	Same	oxidized porous Si (20 μm) membrane	Yes: Pt resistor on bulk	Air and oil	CT	Oil: 420 mV at 30 cm s^{-1} *Air: 7 mV m^{-1} s* Water 1.3 V at 7 l min^{-1}	oil: 0–0.3 m s^{-1}, air: 0–7 m s^{-1}	oil: 0–0.3 m s^{-1} air: 0–14 m s^{-1} water: 0–7 l min^{-1}		oil: 7 K (2.6 mW) air: 10 K	oil: <100 ms air: 10 ms (0–10 m s^{-1})
Tai and Muller (1988)	Polysilicon resistor	Same	PSG bumps	No	Air	CC	30 mV at 1.5 m s^{-1}		0–2 m s^{-1}		2 mW	~3 s (3–10 s)
Toda *et al.* (1996)	NiCr	Ni	None, hot chip	Yes: separate chip with Ni resistor	Air	CTD	*17 mW(l min^{-1})$^{-0.5}$*		0–5 L min^{-1} in 6 mm diam tube		20 K (approx.)	

(*Continued*)

Table 2 (Continued)

Reference	*Heater*[a]	*Sensor*	*Insulation*[b]	*Fluid temp Compensation*	*Fluid type*	*Mode*[c]	*Sensitivity*[d,e,f]	*Sensitivity range*[g]	*Measured range*[h]	*Resolution*	*ΔT/I/V/P*[i]	*Time constant*[j]
Tung *et al.* (2004)	Polysilicon resistor	Same	SiN on vacuum	No	Air	CT	2.85 V Pa^{-1}		0.05 – 0.72 Pa			
van Baar *et al.* (2001)	Pt (Cr)	Same	SiN support beams	Yes: similar resistors	Ar, CO_2, nitrogen	CC	Not measured		0 – 2 m s^{-1}			
van der Wiel *et al.* (1993)	Implanted resistor (patterned for uniform heating)	Diode string from bipolar transistors	Si membrane (15 μm thick)	Yes: Diode string on rim of sensor	Water (DI)	CC	>10 mV m^{-1} s	0 – 3 m s^{-1}	0.005 – 2.65 m s^{-1} – 5 – 3000 l h^{-1} (20 mm diam tube)			
van Putten and Middelhoek (1974)	Four polysilicon resistors	Same	None: hot chip (50 μm thick)	No	Air	Wheatstone bridge	2 μV m^{-1} s					
van Putten (1983)	Eight polysilicon resistors	Same	None: hot chip	No	Air	Double Wheatstone bridge	200 mV at 200 m min^{-1}		35 – 200 m min^{-1}			<100 ms
Verhoeven and Huijsing (1994)	Implanted resistor	Al/p-Si Thermopile	Si membrane	Yes: Thermopile	Gas	CTD	*ΣΔ output: 121pulses(m s^{-1})$^{-0.5}$*		0 – 50 m s^{-1}		*20 mW*	5 ms
Wu *et al.* (2001)	Polysilicon resistor	Same	SiN suspended channel	Maybe	Air	CC	360 μV nl^{-1} min		0 – 120 nl min^{-1}		70 μA	
Xie *et al.* (2003)	Cr/Au resistor	Same	Parylene membrane on cavity	Yes: resistor on bulk	Air	CC	55 μV μl^{-1} min	0 – 9 μl min^{-1}		0.2 μl min^{-1}	4 mA (6.3 K)	1 ms
Xie *et al.* (2003)	Cr/Au resistor	Same	Parylene membrane on cavity	Yes: resistor on bulk	Water	CC	12.2 μV nl^{-1} min	0 – 90 nl min^{-1}				
Xu *et al.* (2005a)	Polysilicon resistor	Same	SiN on vacuum	No	Water	CT	ΔV/V = 0.14 at 5.3 Pa		0 – 5.3 Pa		37 K	
Yoon and Wise (1988)	Metal / Polysilicon resistor	Au/Cr resistors	SiO/SiN/SiO membrane	No	Air	Wheatstone bridge[l]	34 mV/V at 5 m s^{-1}		0 – 8 m s^{-1}		14 mW (40 K)	
Yoon and Wise (1992)	Cr/Au resistor	Polysilicon/Au thermopile	SiO/SiN sandwich on air	No	Gas	CT	20 mV (m s^{-1})$^{-0.5}$	0 – 4 m s^{-1}	0 – 4 m s^{-1}		60 K	
Yoshino *et al.* (2001)	Pt resistor	Same	SiO membrane with air slit on air	Yes: Pt resistor on bulk	Air	CT	*1.6 W m^{-1} $Pa^{-0.5}$*		0 – 0.9 Pa		60 K	

[a]Implanted resistor: using ion implantation.
[b]Silicon dioxide and silicon nitride are abbreviated to SiO and SiN, respectively, irrespective of stoichiometry.
[c]Mode indicates the operating mode in which the sensor has been evaluated. CC, constant current; CP, constant power; CT, constant temperature; CTD, constant temperature difference; CV, constant voltage.
[d]Italicized sensitivities indicate that the figure has been estimated from plots.
[e]In cases where the output signal is not described by a sensitivity factor, only the output at the largest measured flow rate is given.
[f]In cases of different parameters for similar devices (e.g., temperature difference), the highest response was given.
[g]Sensitivity range: the sensitivity indicated is only valid within this range, if absent, the sensitivity is valid for the entire measured range.
[h]Italicized values indicate that the figure was estimated from plots.
[i]ΔT/I/V/P: the temperature, current, voltage or power at which the sensor is operating; the indicated value depends on the operating mode. If the temperature is indicated in degree celcius, it implies the absolute temperature at which the sensor is driven; otherwise, itrefers to the temperature difference with the fluid. Values between brackets are approximate values of a non- constant parameter.
[j]Time constants differ often between heating and cooling. The smallest time constant is given. In case a percentage is given, the value specifies the rise time needed to reach this percentage of the steady-state value.
[k]T_b: Temperature at which the resonant buckles.
[l]Assumed, not specified in the article.

Thermal insulation is arguably the most important aspect of thermal sensors. For an all-silicon sensor that does not have any precautions for thermal insulation, the high thermal conductivity of silicon will result in a hot-chip configuration where the entire silicon die is heated up to the same temperature, effectively reducing the resolution to the chip size and resulting in a relatively large power consumption (Bruschi *et al.* 2004a, Dominguez *et al.* 2005, Huang and Tong 1989, MacHattie 1979, Toda *et al.* 1996, Tong and Huang 1987). Placing the heater on silicon, silicon dioxide, and silicon nitride membranes is a common insulation method (e.g., Kersjes *et al.* 1993, Tabata 1986, Yoon and Wise 1988). **Figure 4(a)** and **4(b)** illustrates two basic methods of achieving such membranes: back-side etching and front-side etching. While back-side etching requires an extra mask step, it does not require etch openings in the membrane, which can allow natural convection in the cavity below to affect the measurements (Gajda and Ahmed 1995, Hung *et al.* 2000). A slightly more complex process, using membrane transfer bonding on cavities at low pressure (e.g., Breuer *et al.* 1999) or resealing of a cavity created by a sacrificial layer using low-pressure chemical vapor deposition (e.g., Huang *et al.* 1995), overcomes this problem and can further reduce conduction downward. Thermal insulation can also be obtained with an insulating substrate, such as glass (Kohl *et al.* 1994, Kuttner *et al.* 1991, Ma *et al.* 1995, Platzer 1976, Esashi 1991) or low-temperature cofired ceramics (LTCC) (Schmid 2002), or by having a thick thermally insulating layer, either below or around the heated area, made of, for example, polyimide (Dittmann *et al.* 2001, Ebefors *et al.* 1999, Kälvesten *et al.* 1995, Löfdahl *et al.* 1989a, 1992, Stemme 1986) or oxidized porous silicon (Dominguez *et al.* 1993, Pagonis *et al.* 2004, Perichon *et al.* 2000, Tabata 1986, Tabata *et al.* 1987) (see **Figure 4(d)**, **4(e)**, and **4(i)**). As in conventional hot-wire anemometers, thermal insulation, or at least control over thermal conduction, can be obtained by prongs that suspend the hot element (Chen and Liu 2003, Chen *et al.* 2003, 2002, Ebefors *et al.* 1999, Haasl *et al.* 2005, Jiang *et al.*, 1994a, b, Löfdahl *et al.* 2003). In microchannels, an equivalent to this is a heater that crosses the channel edge to edge, as illustrated in **Figure 4(i)**. Achieving insulation in microchannels is a challenge as the channel walls can conduct away a large portion of the heat. **Figure 4(j)** shows the principle presented by Wu *et al.* (2001) where a polysilicon heater is integrated into a suspended silicon nitride channel.

Since thermal anemometers are inherently sensitive to fluid temperature changes, many devices include a separate sensor that measures the fluid temperature. For hot-chip configurations, this can be an identical but unheated sensor at a different point in the flow, often upstream, (Esashi 1991, Kersjes *et al.* 1996, Kohl *et al.* 1994, Kuttner *et al.* 1991, Toda *et al.* 1996), while for sensors with the heated element isolated from the rest of the chip, it can be a sensing element such as a resistor or a diode on the bulk area (e.g., Stemme 1988, Tabata 1986). A different approach for measuring the fluid temperature was used by Bruschi *et al.* (2004b) with the sensor illustrated in **Figure 5**. A polysilicon heating resistor

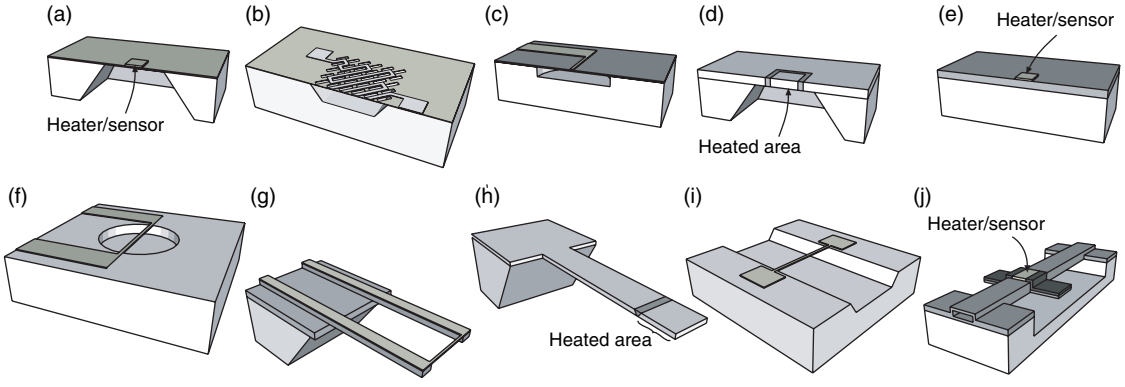


Figure 4 Cross sections of thermal anemometers to illustrate thermal insulation methods: (a) back-side-etched thermally insulating membrane; (b) front-side-etched thermally insulating membrane (Hung *et al.* 2000); (c) membrane transfer bonded to the cavity or the resealed sacrificially etched cavity (Breuer *et al.* 1999); (d) thermally conducting membrane with thermally insulating grooves; (e) thick thermally insulating layer on the substrate or thermally insulating substrate; (f) open cavity; (g) prong suspension (Jiang *et al.* 1994a); (h) insulating groove suspension; (i) cross-channel bridge (top half of the channel not shown), with or without thermally insulating reinforcement; and (j) insulating channel walls (Wu *et al.* 2001).

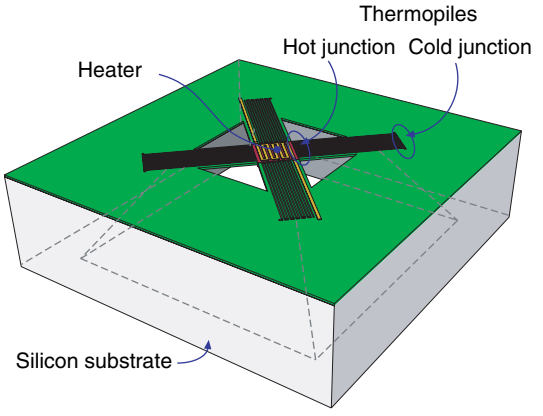


Figure 5 Thermal anemometer used in pulse mode to obtain flow temperature compensation. (Source: Bruschi P *et al.* 2005).

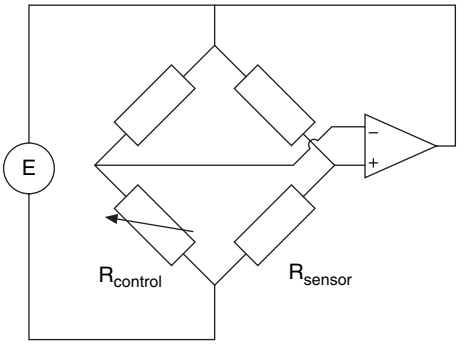


Figure 6 Constant temperature (CT) circuit. The temperature of the sensing resistor is controlled by $R_{control}$. The bridge voltage E is a measure of the flow.

is positioned on a cross-shaped membrane that has n/p-Si thermopiles on its arms measuring the temperature difference between the membrane and the bulk of the chip. Due to the low thermal mass of the heater, operating it in pulsed mode enables the temperature difference between the fluid and the bulk of the chip during the off state of the pulse to be measured.

2.07.2.3.2 Thermal anemometers: modes of operation

Thermal anemometers can be run in several different modes. Constant voltage (CV) and constant current (CC) require the least control. For sensors where a resistor is used both for heating and sensing, the resistance, and thus the temperature, can be easily calculated from the current or the voltage. A common implementation for these modes is to use a Wheatstone bridge. With this set-up, a fluid temperature sensor can also be easily introduced to provide temperature compensation. Both methods are approximations of a slightly more complex approach: the constant power (CP) control of the heater. All three methods suffer, however, from the time constant of the system being limited by the thermal time constant of the heated element.

A solution to this is to run the sensor in CT mode. By using a feedback circuit to keep the temperature of the sensor constant and by measuring the feedback voltage, the time constant of the system is only limited by the response time of the circuit (**Figure 6** for a standard analog CT circuit). Another advantage, specific for liquid flowmeters, is that there is no risk of overheating when the sensor is removed from the liquid during operation. In conventional hot-wire anemometry, CT is the most common measurement method, while a wire operated at a low current (to avoid self-heating) in CC mode is often used for fluid temperature measurements. **Figure 7** shows the typical response curve of a thermal anemometer run in CT mode, $V = \sqrt{A + B\sqrt{U}}$, which illustrates the general characteristic of thermal flow sensors, namely that the sensitivity is highest at low flow velocities and decreases at higher velocities.

In cases where the fluid temperature can vary much (e.g., Castaner *et al.* 1997, Steurer and Kohl 1998) or where the sensor's ambient temperature is only slightly above the ambient (e.g., Kuttner *et al.* 1991), it is desirable to control the heating element's temperature at a CTD with the ambient. However, this puts more demands on the circuitry, and, as Lammerink *et al.* (2000) pointed out, in all cases cannot generate an output signal that is independent

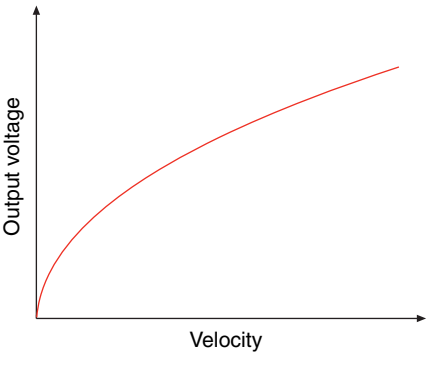


Figure 7 Typical response curve of a thermal anemometer run in constant temperature (CT) mode.

of the fluid temperature. The requirement for a CTD circuit to provide a fully ambient-temperature-independent signal is that the response of the ambient and heated element temperature sensors is linear or at least the same in the ambient temperature range as in the heated element's temperature range. The solution presented by Lammerink *et al.* (2000) is an operating-mode dubbed temperature-balance anemometry (TBA) using two hot-element calorimetric flow sensors as described further below.

It is also possible to combine several heated elements to achieve higher sensitivity. By placing four rectangular hot-element sensors in a square, a full Wheatstone bridge is formed (Betzner *et al.* 1996, Ma *et al.* 1995, van Putten 1983). The flow will cool down the elements perpendicular to it more than those that are parallel to it, resulting in a bridge imbalance. A second effect used, as described by van Putten (1983), is that, for a chip placed in the flow, the thermal boundary layer is thinner at the leading edge, resulting in a higher temperature gradient and hence a higher heat transfer coefficient. Van Putten, moreover, designed a double Wheatstone bridge to increase the differential-mode sensitivity and common-mode suppression. The second bridge is placed inside the first bridge and is electrically linked at a 90° angle (**Figure 8**). The temperature gradient due to the thermal boundary layer thickness difference was isolated in a later version of the double-bridge sensor. Van Putten (1988) and van Putten *et al.* (1997) split up the bridges and used heating from the inner bridge to keep the average temperature of the outer bridge constant. In this way the outer bridge, driven in CC mode, is kept at aCV, resulting in constant current, constant voltage (CCCV) mode. The bridge voltage of the outer bridge is shown to be dependent on the heat dissipation difference between the upstream and downstream resistors. Besides resulting in a direction-dependent signal, this operating method provides a high rejection of the common-mode dissipation variations caused by varying flow and ambient temperature. The sensor is still sensitive to ambient temperature variations, but the dependence is reduced to a linear effect.

Stemme (1988) designed an integrated digital feedback circuit that oscillates the hot element's temperature between two set temperatures. A CC source heats the element until it reaches the top temperature, then switches off until the element has cooled down to the bottom temperature. The output signal consists of the lengths of the current pulses in the supply voltage. The ratio between the heating and the cooling time is a function of the flow rate. Yoon

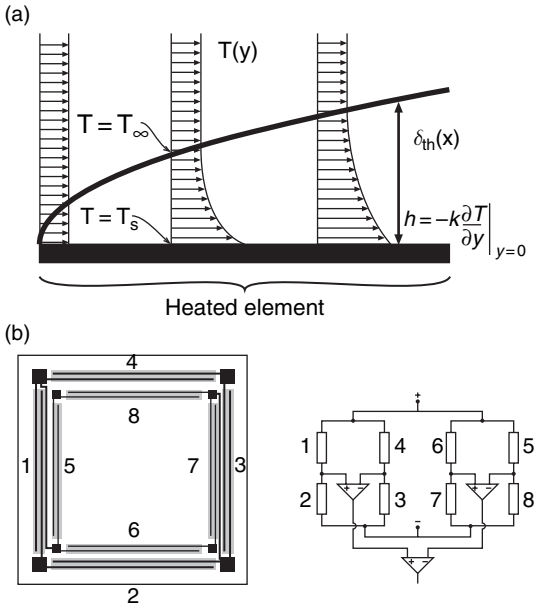


Figure 8 (a) Near-wall flow profile evolution for a chip placed in the flow. (b) Double Wheatstone bridge configuration by van Putten (1983) (Source: Van Putten A 1983 An integrated silicon double bridge anemometer. *Sens. Actuators* **4(3)**, 387–96. © 1983 Elsevier.) chip layout and circuit. Note that the very high resistivity polysilicon heaters (gray) are contact with aluminum leads (black) such that they are wide and short to reduce the resistance. For a flow *U* coming from the left, the resistors 1, 3, 5, and 7 are cooled down more than the others, resulting in a differential output from both bridges.

and Wise (1992) used this operating mode as a self-test mode for a sensor that, under normal conditions, is run in CT mode. The pulse time is dependent on both the thermal time constant and the flow rate. A second flow-rate sensor on the same chip, operated in CT mode, measures the latter. Comparing signals of both modes allows the thermal time constant to be monitored. Surface film depositions, such as oil and carbon from a dirty environment, can in this way be monitored with a stated sensitivity of below 1 μm. At these film thicknesses, velocity errors are at the same level as the resolution of the sensor. Stenberg *et al.* (1988) presented a two-element calorimetric sensor for the same purpose of measuring surface film thickness together with flow (see below).

Alternating current (AC) driving modes are uncommon for single-element thermal anemometers and used only in some special cases where they have specific advantages. One is the anemometer based on a single thermopile developed by Al Khalfioui *et al.* (2003), shown in **Figure 9**. The thermopile consists

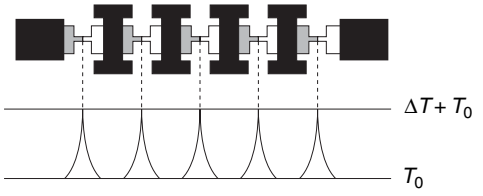


Figure 9 Linear thermopile with temperature profile. (Source: Al Khalfioui M, Michez A, Giani A, Boyer A, Foucaran A 2003 Anemometer based on Seebeck effect. *Sens. Actuators A Phys.* **107(1)**, 36–41. © 2003 Elevier.)

of a small number (e.g., 5) of $Bi_2Te_3-Sb_2Te_3$(P)/$Bi_2Te_3-Bi_2Se_3$(N) thermocouples placed in a line on a polyimide substrate. Each cold junction is covered with a Ni/Ag heat radiator to keep it at room temperature. The hot junctions are heated using AC, thus allowing the direct current (DC) thermoelectric voltage to be read out independent of the driving current.

On the border between micromachining and conventional technology, flow sensors based on acoustic waves have been presented by different groups (Ahmad 1985, Brace *et al.* 1989, Joshi 1989, Joshi and Jin 1990, Rebiere *et al.* 1994). A piezoelectric substrate (e.g., $LiNbO_3$) is heated and a pair of interdigitated transducers is used to generate an acoustic oscillation. The flow-induced cooling of the entire chip is detected as a frequency shift in the oscillation. Heating can be generated by thin- or thick-film resistors in between and around the transducers and also by the acoustic waves themselves (Brace *et al.* 1989). Care needs to be taken in the spacing of the heating meanders to avoid acoustic reflection (Joshi 1994). The large size of the sensors used (ranging from 0.2 to 2.7 cm^2) causes large time constants ranging from 10 to 90 s. Joshi (1994) suggested that time constants can be drastically reduced using ultrasonic Lamb wave oscillators on silicon nitride membranes as presented by Wenzel and White (1988). Most of the presented devices use SAW modes with the exception of the one presented by Rebiere *et al.* (1994), who evaluated a device using a shear horizontal acoustic wave mode traveling through the bulk and facilitating the placement of the ultrasonic transducer contacts outside of the flow.

Another type of resonance-based flow sensor makes use of temperature-dependent stress in membranes and beams. A membrane (Bouwstra *et al.* 1989) or a bridge (Bouwstra *et al.* 1990) is heated with a polysilicon resistor to an elevated temperature with a DC signal and is excited into vibration by superimposing an AC signal at its resonance frequency. The stress in the resonating structure is dependent on its temperature and affects the resonance frequency. As the sensor is cooled by airflow, the change in the resonance frequency can be measured by piezoelectric resistors on the edge of the resonator. Compared to the membrane resonator, the bridge resonator has a higher sensitivity since heat conduction to the support structure is reduced and both sides of the structure are in contact with the flow. The resonance frequency was found to change the most near the buckling stress of the beam and had a larger absolute sensitivity to flow in the buckled state than in the unbuckled state (Legtenberg *et al.* 1991).

Using the temperature dependence of the dielectric constant of high-k dielectrics such as Ta_2O_5 and $SrTiO_3$, Kwok *et al.* (1996) and Lin *et al.* (1996) constructed thermocapacitive metal oxide semiconductor (MOS) flow sensors. The sensing capacitor placed on a thermally insulating membrane or beam is heated with a polysilicon resistor. The capacitance is measured by having the capacitor form the floating gate of a MOS transistor (Lin *et al.* 1996) or by using a dedicated capacitance readout circuit (Kwok *et al.* 1996).

Lee *et al.* (2002) developed a vortex-based flow sensor. By means of a nozzle placed in front of a V-shaped obstruction (**Figure 10**), a flapping mode can be generated. The frequency of the planar liquid jet flapping between the two exits from the channel formed by the obstruction is a function of the flow velocity and the geometry. A number to characterize

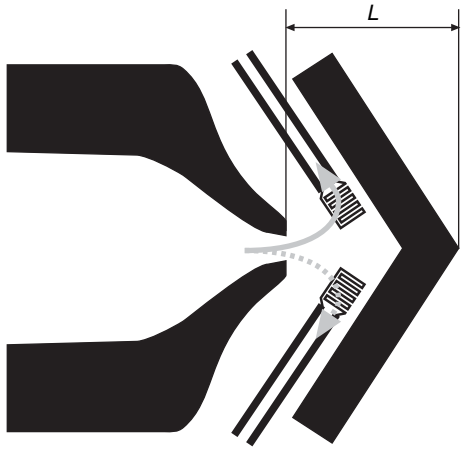


Figure 10 Top view of nozzle, obstruction, and sensing resistors; arrows indicate the flapping mode. (Source: Lee G-B, Kuo T-Y, Wu W-Y 2002. A novel micromachined flow sensor using periodic flapping motion of a planar jet impinging on a V-shaped plate. *Exp. Therm. Fluid Sci.* **26(5)**, 435–44. © 2002 Elsevier.)

this behavior is the Strouhal number, defined as $St = fL/U$, where U is the average flow velocity, f the oscillation frequency, and L the characteristic length scale (defined in **Figure 10**). The Strouhal number for this geometry was shown to be a constant over the measured flow range (0.16–4 mm s^{-1}) and two tested distances ($L = 1.44$ and 2.79 mm), the result being a linear relationship between the flapping frequency and the average flow velocity. The flapping frequency, ranging from 0.01 to 0.31 Hz, was measured by a pair of resistors in the CC mode. A peculiarity with the measurement method is that it was possible to create a flapping flow on a V-shaped plate at the low Reynolds numbers (0.2–5.4) associated with this setup. The limitations of this method are the high pressure loss and the low frequencies involved; at 0.083 Hz mm^{-1} s the fluid will flow 12 mm in one flap period. The latter might be improved by miniaturizing the device, and if successful, eliminate the need for low-frequency electronics.

2.07.2.3.3 Calorimetric flow sensors (alternative names: thermotransfer, thermal dilution)

This class of thermal flows sensors measures flows by using at least two elements: a heater and a downstream sensor. The sensor measures the temperature change of the fluid by the heater. The thermal capacity of the fluid is an important parameter, hence the name calorimetric. As mentioned above, thermal anemometers can use distinct heater and sensor elements, but in that case the sensor element measures the temperature of the same thermal entity, measuring the heat loss without the heat passing through the fluid to get to the sensor. Ideal calorimetric flow sensors have heater and sensor in perfect thermal isolation from each other except for the fluid. In reality, however, a large gray zone exists, and the conduction of heat from the heater to the sensor via the substrate produces an unwanted signal. **Table 3** provides an overview of micromachined calorimetric flow sensors.

Nearly always a second sensor is placed upstream. This makes the flow sensor bidirectional and provides a differential signal, so that the effect of the conduction through the substrate can be removed. An offset is still created in cases where the temperature behavior of the upstream and downstream sensors differs. For resistive sensors, for example, this results in a difference of resistance. By placing a pair of sensors orthogonally to the first pair, a 2D flow sensor is obtained (Fürjes *et al.* 2004, Kim *et al.* 2004, Mayer *et al.* 1996, Park *et al.* 2003a, Robadey *et al.* 1995).

A subclass of calorimetric flow sensors uses multiple hot sensors: heaters that double as sensors measure the difference in heat loss between the upstream and the downstream heater (Lammerink *et al.* 2000, Lyons *et al.* 1998, Nagata *et al.* 2001, van Oudheusden *et al.* 1989, Stenberg *et al.* 1988). What differentiates these sensors from the full Wheatstone bridge sensors described above is that the difference in temperature between the upstream and the downstream heater is a result of the flow-induced heat transfer between them, one heating the other through the fluid, and not from the difference in hydrodynamic boundary layer thickness. With this method, besides flow velocity, other parameters, such as, for example, in the case of Stenberg *et al.* (1988), the thickness of fouling biofilms growing on the sensor surface, can be measured.

2.07.2.3.3.(i) Calorimetric flow sensors: modes of operation As with thermal anemometers, the heater can be run in different operating modes.CC, CV, and CP are easy to implement but have disadvantages. Lammerink *et al.* (1993) calculated the behavior of a three-wire calorimetric flow sensor in a microchannel (cross-channel bridge, see **Figure 4(i)**), operated in CP mode. The temperature is assumed to be constant along the wire and to decrease linearly in the direction perpendicular to the flow and the wire. This results in the addition of a term $-\kappa T/l_z^2$ to the steady-state ($\partial T/\partial t = 0$) 1D heat transfer equation along the length axis of the channel:

$$U\frac{\partial T}{\partial x} = \kappa\frac{\partial^2 T}{\partial x^2} - \kappa\frac{T}{l_z^2}$$

with l_z being the height of the channel. With boundary conditions set to a heating power P for the heater width $2L$ and $\lim_{x\to\infty} T = 0$, the solution of this equation becomes $T = T_0$ for $-L<x<L$, $T = T_0\exp(\lambda_1(x+L))$ for $x>L$ (downstream), and $T = T_0\exp(\lambda_2(x-L))$ for $x<L$ (upstream), with

$$\lambda_{1,2} = \frac{1}{2\kappa}\left(U \pm \sqrt{U^2 + 4\kappa^2/l_z^2}\right)$$

and $T_0 = Pf(k, \lambda_{1,2}, \text{geometry})$, where f is a function of the geometry (height and width of the channel and

Table 3 Overview of calorimetric flow sensors

Reference	*Heater*	*Sensor*	*Insulation*[a]	*Fluid temp. Compensation*	*Fluid type*[b]	*Mode*	*Sensitivity*[c,d,e]	*Sensitivity range*[f]	*Measured range*	*Resolution*	*ΔT/I/P*	*Time*[g]
Ashauer *et al.* (1999)	Polysilicon resistor	Al/Si thermopiles	SiO/SiN membrane	Yes: resistor on bulk	Water, oil, IPA	CP	*1.7 V mm^{-1} s*	0.1 – 2.5 mm s^{-1} in a 0.4 × 0.6 mm^2 flow channel	−10 – 10 mm s^{-1}	0.1 mm s^{-1}	5 mW	
Ashauer *et al.* (2001)	Polysilicon resistor	Al/Si thermopiles	SiO/SiN membrane	Yes: resistor on bulk	Water, oil	Not specified	0.66 V ml^{-1} h differential		±0.25 ml h^{-1}	10 µl h^{-1}		2 ms
Billat *et al.* (2002)	Silicon resistor	Silicon resistor	Hotwire configuration	No (external temp. must be kept constant!)	Air, SF_6	CP (free convection measurement)						
Brace *et al.* (1989): Dual-oscillator	SAW	SAW	None, hot chip: LiNbO3 substrate	No	Nitrogen	CP	*−2.1 kHz (cm s^{-1})$^{-0.5}$*	*6 – 500 sccm*	0 – 500 sccm		47.5°C	10 – 30 s
Bracio *et al.* (2000)	Pt resistor	Ge thermistor	Membrane on air (material not specified)	Yes: two Ge thermistors on bulk		CT	No measurements		0.1 – 50 m s^{-1}	1.3 ml h^{-1}	20 K	1 ms in 50 mm dia. channel
Bruschi *et al.* (2005)	Two polysilicon resistors	Al/Si thermopiles (20 n+polysilicon/ Al thermocouples)	SiO membranes	No	Nitrogen	ΔT = 0 method	0.35 mV at 200 sccm over thermopiles		0 – 200 sccm		6 V	
Buchner *et al.* (2006)	W90Ti10 resistor or polysilicon resistor	$W_{90}Ti_{10}$/Polysilicon thermopile	SiN membrane on polymer filling	No	Water	CT	*3 V µl^{-1} s*	0 – 2 µl s^{-1}	0 – 20 µl s^{-1}		44.1 K	
Dillner *et al.* (1997)	NiCrSi resistor	$Bi_{0.87}Sb_{0.13}$ thermopile	SiN/SiO/SiN stress comp membrane	Yes – cold junction on bulk	Dry air	CP	5.9 V W^{-1} m^{-1} s		0 – 1660 sccm		1 mW	
Ernst *et al.* (2002)	Cr resistor	Ge thermistor (amorphous)	SiN membrane	Yes: two Ge thermistors on bulk	Water	CP (isocaloric) and CT	CT: 2.197 V^2 (mg s^{-1})$^{-0.5}$ CP: 1.07 K W^{-1} µl^{-1} h	CT: 0 – 7 µl s^{-1} CP: 0.1 – 90 µl h^{-1}	100 nl h^{-1} – 3 µl min^{-1}		CT: 5 K CP: 5 mW	
Frederick *et al.* (1985)	NiCr resistor	$LiTaO_3$ substrate with two NiCr electrodes on either side and one on the bottom: pyroelectric	None ($LiTaO_3$ wafer)	No (but stated to be "relatively independent of fluid temperature")	N_2,O_2	AC voltage	20 mV at 500 m min^{-1}		0.5 – 200 m min^{-1}			
Fürjes *et al.* (2004)	Pt resistor	Pt resistor	SiN membrane	No	Air	CC	2 mV at 4 m s^{-1}		0 – 4 m s^{-1}			
Glaninger *et al.* (2000)	Pt resistor or NiCr resistor	Ge thermistor (amorphous)	SiN membrane	Yes: two Ge thermistors on bulk	Air	CTD (CP for characterization)	ΔP = 24 mW at 50 ml s^{-1} in 0.54 mm 2 channel		0.01 – 200 m s^{-1} in 50 mm diam tube 0.6 ml h^{-1} – 150 l h^{-1} in 0.54 mm^2 cross-sectional tube		25 K	*CT: 20 ms* *CP: 1.6 ms*
Gongora-Rubio *et al.* (1999)	Ru-based resistor	NTC thermistor (b ≈ 2000)	None (Green Tape is thermally conductive)	Yes: separate unheated bridge	Nitrogen	CP	ΔT=18 K at 4 slm for 60 mA		0 – 4 slm			1 s

Häberli *et al.* (1997)	Polysilicon resistor	n-polysilicon/ p-polysilicon thermopile (20)	SiO membrane on air	Yes: thermopile cold on bulk	Air	Not specified	38 mV at 38 m s^{-1}		0 – 38 m s^{-1}			
Johnson and Higashi (1987)	$Ni_{80}Fe_{20}$ (permalloy) resistors	$Ni_{80}Fe_{20}$ (permalloy) resistors	SiN membrane on air	Yes: resistor on bulk (upstream)	Air, He, CO_2	CC	0.037 V g^{-1} min air	0 – 1 g min^{-1}	0 – 1.2 g min^{-1} (i.e. For air at STP = 1 l min^{-1})			3 ms
Kaltsas and Nassio poulou (1999)	p-polysilicon resistor	Al/p-polysilicon thermopile	40 μm thick porous silicon layer	Yes: thermopile cold on bulk	Nitrogen	CC	6 mV m^{-1} s W^{-1}	0 – 0.4 m s^{-1}		4.1 mm s^{-1} ~ 10 sccm		1.5 ms
Kaltsas *et al.* (2002)	p-polysilicon resistor	Al/p-polysilicon thermopile	40 μm thick porous silicon layer	Yes: thermopile cold on bulk	N_2, O_2, Ar	CC and CP (evaluated)	N_2: 1.045 mV (m s^{-1})$^{-0.5}$ O2: 1.105 mV (m s^{-1})$^{-0.5}$ Ar: 0.766 mV (m s^{-1})$^{-0.5}$		0 – 4 m s^{-1}			1.5 ms
Kim *et al.* (2004)	Pt resistor	Pt resistor	Si membrane (30 μm)	No	Air	CT – only flow direction is calorimetric!	0.32 V at 10 m s^{-1}		0 – 10 m s^{-1}	0.5 m s^{-1}		
Koch *et al.* (1999)	Ti resistor	Ti resistor	SiN/SiO grid in the liquid	No	Ethanol	No measure-ments	No measure-ments					
Lai *et al.* (1997)	Diffused resistor	Diffused resistor	None	Yes: separate chip on back side of PCB	Gas	CTD	1.37 at 3 m s^{-1}		−4 to 4 m s^{-1}		44.5 K	
Lammerink *et al.* (1993)	Cr/Au resistor	Cr/Au resistor	SiN bridge	No	Water, air, IPA	CV	13 μV K^{-1}, IPA: 8×10^3 K m^{-1} s water: 2×10^3 K m^{-1} s air: 3×10^3 K m^{-1} s	IPA: 0 – 1.4 mm s^{-1} water: 0 – 2 mm s^{-1} air: 0 – 2 m s^{-1}	0 – 2 m s^{-1}		3 V	
Makinwa *et al.* (2001)	Polysilicon resistors	Al/p+Si (diffused) thermopile	None (Bipolar process)	Yes: off-chip diode	Air	CTD	*Not specified*		0.1 – 25 m s^{-1}	0.5 m s^{-1} (±3%), 3°	Normally 15 K	
Makinwa and Huijsing (2002b)	Polysilicon resistors	Al/p+Si (diffused) thermopile	None	Yes: off-chip diode in case of CTD	Air	CTD and CP (TBA)	*Not specified*		2 – 18 m s^{-1}	±4%, ±2°		
Mayer *et al.* (1995): Air-gap	Polysilicon resistor	Al/polysilicon thermopiles	SiO membrane on air	Yes: thermopile cold on bulk	Nitrogen	CV	0.25 – 1.2 V m^{-1} s W^{-1}		0.1 – 4 m s^{-1}			
Mayer *et al.* (1995): Bridge	Polysilicon resistor	Al/polysilicon thermopiles	SiO membrane on air	Yes: thermopile cold on bulk	Nitrogen	CV	0.039 – 0.56 V m^{-1} s W^{-1}		0.1 – 4 m s^{-1}			
Moser *et al.* (1992)	Polysilicon resistor	Polysilicon resistor	SiO/SiO/SiN bridge	No	Nitrogen	CP	Type A: 2.5 mV m^{-1} s linear, 180 μV K^{-1} at 23 m s^{-1} (nonlinear) Type B: 2.5 mV m^{-1} s linear	Type A: 0 – 5 m s^{-1}, Type B: 1 – 3 m s^{-1}	Type A:0 – 23 m s^{-1} Type B:0 – 25 m s^{-1}	25 mm s^{-1}	Type A: 10 mW	*4.4 ms*

(*Continued*)

Table 3 (Continued)

Reference	*Heater*	*Sensor*	*Insulation*[a]	*Fluid temp. Compensation*	*Fluid type*[b]	*Mode*	*Sensitivity*[c,d,e]	*Sensitivity range*[f]	*Measured range*	*Resolution*	*ΔT/I/P*	*Time*[g]
Moser and Baltes (1993); Häberli *et al.* (1997)	Polysilicon resistor	n-polysilicon/ p-polysilicon thermopile (15) – Al at junctions to avoid p-n diodes	SiO membrane on air	Yes: thermopile cold on bulk	Nitrogen	CP (equal mag. series resistor on bulk)	3 μm CMOS: 0.36 mV sccm^{-1} mW^{-1} 1.2 μm CMOS: 0.5 mV mW^{-1} sccm^{-1}, Häberli:35 mV at 37 m s^{-1}	3 μm CMOS: 10 – 100 sccm, 1.2 μm CMOS: 0 – 81 sccm in 0.5 mm dia. channel	0 – 1000 sccm in 0.5 mm dia. channel, Häberli: 0 – 38 m s^{-1} in open air (6 – 8 mm opening)			1.45 ms
Nagata *et al.* (2001)	Pt resistor	Pt resistor	SiN membrane	No	Gas[j]	CV	1.85 V at 4.75 m s^{-1}		0 – 0.8 m s^{-1}		1.5 V	
Nguyen and Kiehnscherf (1995)	Polysilicon resistor	Polysilicon resistor	Si membrane (10 – 30 μm) on air	No	Water, Nitrogen	CT and CP	*water, CT: 0.26 V (ml min^{-1})$^{-0.5}$* CP: nonlinear and not monotonous N_2 CT: 2.9 V at 300 ml min		water:0 – 210 ml min^{-1} nitrogen:0 – 300 ml min^{-1}		CT: 30 K	1 – 4 ms
Nguyen *et al.* (2000)	Polysilicon resistor	Al/polysilicon thermopiles	SiN/ZnO membrane on air	Yes: thermopile cold on bulk (out of flow!)	Air	CP	No data					
Norlin *et al.* (1998)	Polysilicon resistor	Al/polysilicon thermopiles	Quartz substrate	Yes: separate on-chip temperature sensor	Water	Not specified	80 mV (amplified) at 150 μl min^{-1}		0 – 200 μl min^{-1} in 1 × 0.5 mm cross-section channel			
Oda *et al.* (2003)	Pt resistor	Pt/p^{+} doped polysilicon thermopile	SiO membrane on air	Yes: resistor on bulk	Air	CT	*0.8 mV (l h^{-1})$^{-0.5}$*		0.6 l h^{-1} – 12000 l h^{-1}		3.86 V	
Ohnstein *et al.* (1990) (electronics by Chavan *et al.* (2003)	Pt resistor	Pt resistor	SiN membrane on air	Yes: two resistors on bulk	Air	CT	70 mV at 9000 m min^{-1}		0.9 – 9000 m– min^{-1}			
Pagonis *et al.* (2004)	p-polysilicon resistor	Al/p-polysilicon thermopile	10 μm thick porous Si on 10 μm thick air cavity	Yes: thermopile cold on bulk	Nitrogen	CT	14 mV (m s^{-1})$^{-0.5}$ W^{-1}		0 – 2000 sccm (0 – 9.43 m s^{-1})			
Park *et al.* (2003a)	Pt resistor	Pt resistor	Si membrane	No – but not required either	Air	Not specified	Measures only angle		5 and 10 m s^{-1}			*10 s*
Polla *et al.* (1983)	Polysilicon resistor or P-implanted resistor	ZnO pyroelectric layer	None	No	Air	Not specified	184 mV at 2 m s^{-1} *8.8 mV m^{-1} s*	2 m s^{-1}		SNR ≈ 200	at 10.4 circuit gain and 36°C	1 at ΔT = 3°C, 40 s at ΔT = 10°C for v = 2 m s^{-1}

Qiu *et al.* (1996)	Ni resistor	Ni resistors	SiN membrane on air	Yes: two resistors on bulk	Air (60%RH)	CT[l]	700 mV at 2.7 m s^{-1}		0 – 3 m s^{-1}		55 K	
Rahnamai and Zemel (1981)	NiCr resistor	$LiTaO_3$ substrate with two NiCr electrodes on either side and one on the bottom: pyroelectric	None ($LiTaO_3$ wafer)	No (but stated to be ''relatively independent of fluid temperature'')	Air	AC voltage	10 mV^2 cm^{-1} min	0 – 120 cm min^{-1}	0 – 2000 cm min^{-1}		3 V	
Rasmussen *et al.* (2001): polysilicon sensors	Polysilicon resistor	Polysilicon resistor	SiO membrane on air	No	Liquid	Not specified			0.01 mm s^{-1}			
Rasmussen *et al.* (2001): thermopile sensors	Polysilicon resistor	Metal(Al[l])/ polysilicon thermopiles	SiO membrane on air	No	Liquid	Not specified	2.2 V m^{-1} s	0 – 10 mm s^{-1} in 100 × 100 μm channel				3 ms (response time)
Robadey *et al.* (1995): double bridge	Polysilicon resistor	Al/n-polysilicon thermopiles	SiO membrane on air	Yes: thermopile cold on bulk	Nitrogen	CP	0.19 V m^{-1} s W^{-1}	0 – 240 sccm	−400 to 400 sccm in a 0.2 × 8 mm^2 channel	0.52 mm s^{-1}		
Robadey *et al.* (1995): membrane (front mounting – rear mounting not analyzed)	Polysilicon resistor	n-polysilicon/p-polysilicon thermopile – Al at junctions to avoid p–n diodes	SiO membrane on air	No	Nitrogen	CP	1.9 V m^{-1} s W^{-1}	0 – 200 sccm	−200 to 200 sccm in a 0.2 × 5 mm^2 channel	5.2 mm s^{-1}		
Rodrigues and Furlan (2003)	Polysilicon resistor	Polysilicon resistor	SiN membrane	No	Liquid	No measure-ments	No measure-ments					
Sabate *et al.* (2004)	Ni resistor	Ni resistor	SiN/oxynitride membrane on air	No	Air	CT	different resistor pairs: 40 mV slm^{-1} 2.1 mV slm^{-1} 0.25 mV slm^{-1}	0 – 0.1 slm 0.1 – 2.0 slm 2.0 – 8.0 slm in 7 mm^2 cross-section channel			150°C	
Shin and Besser (2006)	Al resistor	Al resistors	Pyrex substrate	No	Nitrogen	CC	0.12Ω sccm^{-1} and 0.50 sccm^{-1} (different resistor configurations)	0 – 18 sccm in 610 μm x 500 μm channel		±6.6% at 200 mW (reproducibility test)	200 mW (42 K)	55.3 ms, response time = 70 ms
Stenberg *et al.* (1988)	Implanted resistors	Diode	Polyimide-filled grooves, epoxy on back	No	Water	CC		0.2 – 2 m s^{-1}		±10% no film, ±30% with film		
Steurer and Kohl (1998)	Ge thermistor (amor phous)	Ge thermistor (amorphous)	Hotwire configuration (SiN support)	Yes: resistors on bulk	Nitrogen, helium, hydrogen	CTD	Not specified		10 – 500 sccm	0.5%		<10 ms

(*Continued*)

Table 3 (Continued)

Reference	*Heater*	*Sensor*	*Insulation*[a]	*Fluid temp. Compensation*	*Fluid type*[b]	*Mode*	*Sensitivity*[c,d,e]	*Sensitivity range*[f]	*Measured range*	*Resolution*	*ΔT/I/P*	*Time*[g]
Svedin *et al.* (2003a)	Polysilicon resistor	Polysilicon resistor	PI-filled groove	No	Air	Wheatstone bridge (CC)	0.7 mV V^{-1} at 5 l min^{-1}		0 – 20 l min^{-1}			
Tan *et al.* (2006)	Cr/Au resistor	Cr/Au resistor	Polyimide substrate	No	Nitrogen	CC	ΔR/R=8.5% at 2.75 l min^{-1}		0 – 2.75 l min^{-1} in 3 mm dia. tube			
Tanase *et al.* (2002)	Polysilicon resistor	Al/Si thermopiles	polysilicon membrane on SiO	Yes: thermopile cold on bulk, and (planned) separate temp. sensor	Blood	Not specified	14 mV at 21 cm s^{-1}		0 – 14 cm s^{-1}		2 K	
Trautweiler *et al.* (1996)	Diffused resistor[h]	Al/Si thermopile	15 μm Si membrane on glass	No	Air	CC[l]	13.2 mV at 5 slm in 6 × 1 mm^2 channel		−5 to 5 slm		(176 mW/ 13 K)	
van Baar *et al.* (2001)	Pt resistor	Pt resistor	SiN beam	Not required due to AC measure-ment, alt. upstream probe	Gases: CO_2, N_2	Not specified	Not specified		0 – 5 m s^{-1}			
Weiping *et al.* (2005)	NiCr resistor	Ni resistor	None: SiN/SiO on Si substrate	No	*Water*	CV	0.55°C at 350 μl min^{-1} in single 130 μm dia. capillary 0.6°C at 700 μl min^{-1} in double capillary		10 – 700 μl min^{-1}		~60°C at 18 V	
Yang and Soeberg (1992)	Diode	Diode	Si membrane (3 μm) on Glass substrate	Yes: upstream reference diode	Water	CP	32 V ml^{-1} min (amplified)	0 – 0.2 ml min^{-1}	0 – 10 ml min^{-1}			0.2 s at 0.2 ml min^{-1}
Yoon and Wise (1992)	Cr/Au resistor	Polysilicon/Au thermopile	SiO/SiN sandwich on air	No	Gas	N/A	Only flow direction sensor is calorimetric, the flow rate sensor is single element					
Yu *et al.* (1993)	NiCr resistor	$LiTaO_3$ substrate with two NiCr electrodes on either side and one on the bottom: pyroelectric	None ($LiTaO_3$ wafer)	No (but stated to be ''relatively independent of fluid temperature'')	Air	AC voltage	Reynolds number sensitivity: 8×10^{-5} at f = 1.55 Hz					

[a]Silicon dioxide and silicon nitride are abbreviated to SiO and SiN, respectively, irrespective of stoichiometry.
[b]Italicized fluid type indicates that the fluid was not specified in the reference.
[c]Italicized sensitivities indicate that the figure has been estimated from plots.
[d]In cases where the output signal is not described by a sensitivity factor, only the output at the largest measured flow rate is given.
[e]In cases of different parameters for similar devices (e.g., temperature difference), the highest response was given.
[f]Sensitivity range: the sensitivity indicated is only valid within this range, if absent, the sensitivity is valid for the entire measured range. Italicized values indicate that the figure was estimated from plots.
[g]Time constants differ often between heating and cooling. The smallest time constant is given. In case a percentage is given, the value specifies the rise time needed to reach this percentage of the steady-state value. Italicized values indicate that the figure was estimated from plots.
[h]Assumed, not specified in the arich.

width of the heater), the thermal conductivity, k, and $\lambda_{1,2}$. In **Figure 11**, the results are plotted for some typical values of U.

Figure 12 illustrates how, at low flow velocities, the temperature difference between the downstream and the upstream sensors, ΔT, increases linearly. However, at increasing velocities, the heater is cooled so much that ΔT decreases again.

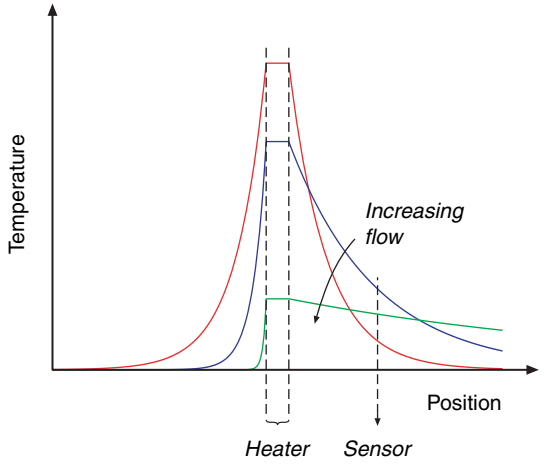


Figure 11 Temperature profile near the heater for three different flow rates. The vertical line indicates the position of the downstream sensor.

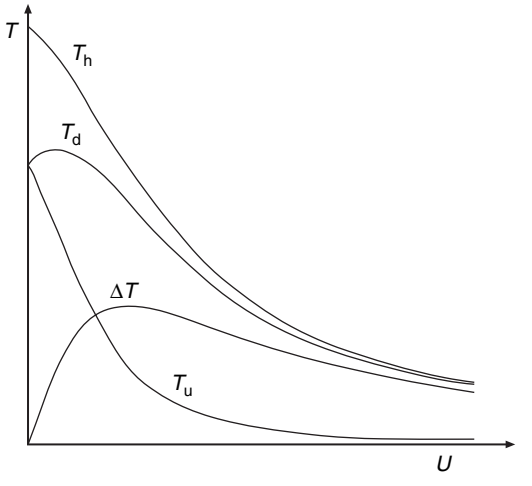


Figure 12 Sensor temperatures as a function of the flow velocity U. T_h is the heater temperature, T_u and T_d the upstream and downstream sensor temperatures, respectively, and ΔT the temperature difference between the downstream and the upstream sensor. (Source: Lammerink T, Tas N, Elwenspoek M, Fluitman J 1993 Micro-liquid flow sensor. *Sens. Actuators A (Phys.)* **A37–A38**, 45–50. © 1993 Elsevier.)

CT control of the heater overcomes this problem and provides a monotonously increasing ΔT. As with thermal anemometers, it is also possible to control the temperature of the heater to maintain a CTD with the fluid.

The aforementioned TBA operating mode (Lammerink *et al.* 2000, Makinwa and Huijsing 2002a) is based on two heated elements at a small distance from each other kept at the same temperature by a CP controller. The total power to the two heated elements remains constant and the ratio of the power distribution between the two elements is a measure for the flow. Because of this, the restraint on the response of the temperature sensors is loosened, so that the two sensors only have to have the same response, unlike CTD where the response of the two sensors has to be the same at both the ambient and the heated temperature. This gives the freedom to use sensing methods that are highly sensitive and reproducible but nonlinear, such as metal–semiconductor thermocouples. An AC-driven variant was presented by Lammerink *et al.* (2001) that consisted of two heating elements driven at 90° phase shift (creating a pair of 180° phase-shifted thermal signals) and a third element between them to sense the temperature variations using a phase-lock amplifier. The advantages of this system lie in the avoidance of DC offset and $1/f$ noise.

Although most often the heater and the sensor type do not directly require a specific operating mode, a specific subclass of sensors based on the pyroelectric effect does. Pyroelectric materials form a subclass of piezoelectric materials that distinguish themselves by generating a surface charge when undergoing temperature changes. **Figure 13** shows

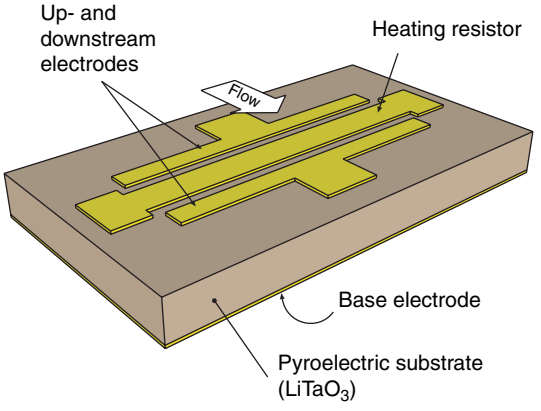


Figure 13 Flow sensor based on the pyroelectric effect. (Source: Zemel and Rahnamai H 1980).

a sketch of the device invented by Zemel and Rahnamai (1980) in which the electrodes on a lithium tantalate ($LiTaO_3$) substrate enable readout of the local charges upstream and downstream of a heating resistor. By applying a constant-amplitude low-frequency voltage (1–3 Hz) on the heating resistor, a signal at the double frequency is generated at both electrodes. The downstream electrode will experience a larger temperature fluctuation and so generate a larger voltage fluctuation. Advantages of the method are its low power consumption and large measurement range (0.5 ml min^{-1}–20 l min^{-1} was reported in Hsieh *et al.* (1991)). The voltage is measured differentially over the two capacitors. A thin-film pyroelectric flow sensor was presented by Polla *et al.* (1983). Sputter-deposited ZnO was used, and the charge buildup was measured in DC mode. The charge decay was found to be 85 h. A disadvantage of both variants, however, is the large time constant, which ranges from seconds to tens of seconds. A thorough overview of pyroelectric flowmeters is given in Hsieh and Zemel (1995), Hsieh *et al.* (1995a, b).

SAW sensors, which are described in Section 2.07.2.3.1, have also been used in calorimetric mode. Brace *et al.* (1989) used two SAW chips, thermally and acoustically isolated from each other, operating the upstream oscillator at a high CP and the downstream oscillator at low power.

Recently, a new application area for calorimetric flow sensors has emerged. Instead of avoiding natural convection effects, the heater is heated up to a high temperature (150–700°C) to create fluid flow away from the heater. If placed in an enclosed cavity, the direction of the flow is determined only by the acceleration force acting on it, and the airflow becomes a measure for the acceleration and the inclination. The principle was first patented by Plöchinger (1994), and micromachined variants followed soon after (Billat *et al.* 2001, Chaehoi *et al.* 2005, van Honschoten *et al.* 2000, Leung *et al.* 1997, Lin and Jones 2005, Luo *et al.* 2001, Milanovic *et al.* 1998). With the exception of the microflown-based sensor described by van Honschoten *et al.* (2000), all are based on the three-element caloric principle: one heater in the center and one temperature sensor on either side in a bridge formation over a cavity (**Figure 14**) The step to two-axis designs is not big and has been demonstrated both for inclinometers (Billat *et al.* 2001) and accelerometers (Leung *et al.* 1998). Luo *et al.* (2002) have presented an elaborate analysis of the parameters affecting the linearity, sensitivity, and frequency response of convective accelerometers.

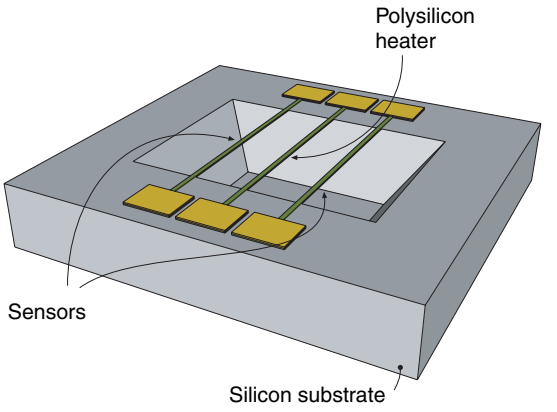


Figure 14 Thermal accelerometer. (Source: Leung *et al.* 1998).

2.07.2.3.3.(ii) The microflown The microflown, conceptualized by de Bree *et al.* (1995), is a flow sensor designed to measure particle velocity. This quantity, defined as the flow volume rate per unit area, is used in acoustic theory together with acoustic pressure to describe acoustic waves. Since normal microphones measure acoustic pressure, the combination with a microflown allows the measurement of acoustic intensity, the product of both. Combining a pressure microphone with a microflown provides a low-cost alternative to the conventional method of measuring flow intensity using two phase-matched microphones, as the phase-matching circuitry demands are high and, hence, expensive (Elwenspoek 1999). Conversely, two microflowns can be placed at a distance from each other and have their signals combined to calculate the acoustic pressure (de Bree *et al.* 1996). The operating principle can be applied to different calorimetric set-ups: a heater with one or two sensors or a two hot-sensor principle (de Bree *et al.* 1995, van Honschoten *et al.* 2005). In contrast to pressure microphones, the particle-velocity measurement is direction sensitive and does not have a low-frequency cutoff. de Bree *et al.* (1999) used the common-mode and differential signals from a microflown operated at CP to create a large-range direction-sensitive flow sensor. The magnitude of the common-mode signal is used as it shows the typical flow dependency of a CP-operated thermal anemometer ($P/(A+B\sqrt{U})$) but is insensitive to the flow direction. The differential-mode signal behaves like a typical nonuniform calorimetric flow sensor (see **Figure 12**), in which the sign as well as the highly sensitive low-flow signal can be used to complement the common-mode signal.

2.07.2.3.4 Time-of-flight flow sensors

An alternative operating method for a heater/sensor setup is to measure the time it takes for a heat pulse to travel from the heater to the sensor. Like calorimetric flow sensors, this sensor type can also be made direction sensitive. As the distance between the heater and the sensor is known, the calculation of the flow speed might at first sight seem easy. However, on a microscale, the thermal diffusivity of the fluid plays a major role and causes the measured signal to deviate from the ideal relationship $t_{peak} = x/U$ between the peak of the pulse measured at a distance x and the flow U.

For the simplified case of a line heat source in a uniform flow, Marshall (1958) derived an analytical solution describing the thermal response to a heat pulse. For this he started from the combined diffusion and convection heat transport equation with a heat generation term $Q/\rho c$, where Q is the generated heat (in W m^{-1}), ρ the density of the medium, and c_p the specific heat capacity of the medium:

$$\frac{\partial T}{\partial t} + U\nabla T = \kappa\nabla^2 T + \frac{Q}{\rho c_p}$$

The general solution at a point at a distance x from the heat pulse source for some different velocities are plotted in Figure 15(a). At low velocities, the thermal pulse significantly broadens before reaching the sensing point. The time at which the temperature peaks at a point at a distance x from the heat pulse source is expressed as (Figure 15(b)):

$$t_{peak} = \frac{\sqrt{4\kappa^2 + U^2x^2} - 2\kappa}{U^2}$$

Rewriting this as a function of the flow velocity results in the following equation:

$$U = \frac{\sqrt{x^2 - 4\kappa t_{peak}}}{t_{peak}}$$

This implies that in order for the ideal relationship $U = x/t_{peak}$ to be valid, $4\kappa t_{peak}$ needs to be much smaller than x^2, which is equivalent to the requirement that $U \gg 4\kappa/x$. Table 4 shows the velocities at some typical distances for which $U = 10 \times 4\kappa/x$ and illustrates the low threshold to nonideal behavior.

Further nonidealities are introduced as micromachined thermal flow sensors are often wall-mounted or consist of wires spanning a microchannel, in which cases the uniform flow assumption no longer holds. Table 5 provides an overview of micromachined thermal time-of-flight sensors.

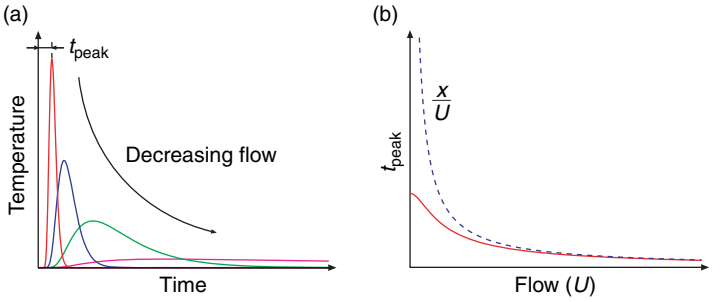


Figure 15 (a) Temperature signal at the downstream sensor for different flow velocities and (b) heat pulse peak time versus flow at the downstream sensor.

Table 4 Threshold for nonideal behavior of thermal time-of-flight sensors for some fluids

	Thermal diffusivity at 273 K ($m^2 s^{-1}$) based on values Lide (2006)	*Velocity above which deviation is less than 10% for heater-sensor distance of*		
		100 μm	*500 μm*	*2000 μm*
Water	1.3×10^{-7}	6 cm s^{-1}	1 cm s^{-1}	0.3 cm s^{-1}
Nitrogen	2.0×10^{-5}	8 m s^{-1}	2 m s^{-1}	0.4 m s^{-1}
Helium	1.6×10^{-1}	60 km s^{-1}	10 km s^{-1}	3 km s^{-1}

Table 5 Overview of thermal time-of-flight sensors

Reference	*Heater*	*Sensor*	*Insulation*	*Fluid temp Compensation*	*Fluid type*	*Mode*	*Sensitivity*	*Measured range*	*time cst cool*
Ashauer *et al.* (1998, 1999)	Polysilicon resistor	Al/Si thermopiles	SiO/SiN membrane	Yes: resistor on bulk	Water, oil, Isopropanol			$11.2-140\,mm\,s^{-1}$ in a $0.4\times0.6\,mm^2$ flow channel	
Bedö *et al.* (2000)	Pt resistor	Pt resistor	SiN membrane	Not required due to AC measurement, alt. upstream probe	Gas, water	AC, CT	Gas: $\Delta P = 12\,mW$ at $7\,m\,s^{-1}$ in $2\times2\,mm^2$ channel Water: $\Delta P = 120\,mW$ at $1.2\,m\,s^{-1}$ in $2\times2\,mm^2$ channel	$0-8\,m\,s^{-1}$	
Lambert (1986)	Pb resistor	PbTe/Pb thermocouple pair	Glass substrate (alt. Mylar or Polyimide)	Yes: same sensor at 90° from the measuring sensor	Air	AC phase	Phase difference: 46° at $250\,g\,s^{-1}$	$10-250\,g\,s^{-1}$ in 6 cm diam tube	"rapid"
Lammerink *et al.* (1995)	Pt resistor	Pt resistor	Free-hanging between Si and glass sandwich	No	He/N_2 mixture, varying concentrations			$0-20\,ml\,min^{-1}$ in a 1000 μm wide flow channel	
Lee *et al.* (2002)	Cr resistor	Same	Quartz substrate	No (not required)	Water	CC/phase detection	*0.083 Hz mm^{-1} s*	$0.15-4\,mm\,s^{-1}$ ($2\,nl\,s^{-1}$)	
Norlin *et al.* (1998)	Polysilicon resistor	Al/polysilicon thermopiles	Quartz substrate	Yes: separate on-chip temperature sensor	Water		*Only mentioned as operating mode, not implemented.*		
Rodrigues and Furlan (2003)	Polysilicon resistor	Polysilicon resistor	SiN membrane	No	Liquid		*Only mentioned as operating mode, not implemented.*		
Shoji and Esashi (1994)	Ti/Pt resistor	Ti/Pt resistor	SiO/SiN membrane	No	Gas (air[1])	CC pulse		$0-5\,m\,s^{-1}$	Heater: 24 ms
Sultan *et al.* (1993)	Pt resistor	Pt resistors	15 μm Polyimide on Si wafer – trenches can be etched if needed	Yes: thermal feedback thermistor (not specified where it should be)		AC phase and magnitude	$\Delta V = 2\,V$ at $140\,g\,s^{-1}$ (amplified?)		
van Kuijk *et al.* (1995)	Pt resistor	Pt resistor	Hoya glass substrate	No	He/N_2 mixture, varying concentrations			$0.1-100\,mm\,s^{-1}$ in a 300 μm-wide flow channel	
Yang and Soeberg (1992)	Diode	Diode	Si membrane (3 μm) on glass substrate	Yes: upstream resistor	Lamp oil	AC	Phase difference: 500° at $0.1\,ml\,min^{-1}$ – 80° at $10\,ml\,min^{-1}$		

van Kuijk *et al.* (1995) and Lammerink *et al.* (1995) used Marshall's results to develop a flow sensor that calculates both the flow velocity and the diffusivity by combining a thermal anemometer with a thermal time-of-flight sensor. With these measured values as inputs to an artificial neural network, the flow rate and the composition of a nitrogen/helium mixture can be determined. The measurements were performed using silicon nitride-supported Pt resistors with 80 µm spacing spanning a 600 µm × 1000 µm channel.

Figure 16(a) shows the time-of-flight sensor presented by Shoji and Esashi (1994) using a comb-like structure with four cantilever structures extending from an 11-mm-wide base. The heating and sensing elements are Pt resistors thermally isolated from the bulk by a SiO_2/Si_3N_4 membrane and supported by a p^+-doped silicon frame. The resistors themselves are also supported to avoid bending of the frame due to thermal expansion. Designed to measure airflows up to 5 m s^{-1}, it measures at three different distances (1, 4.2, and 10.2 mm).

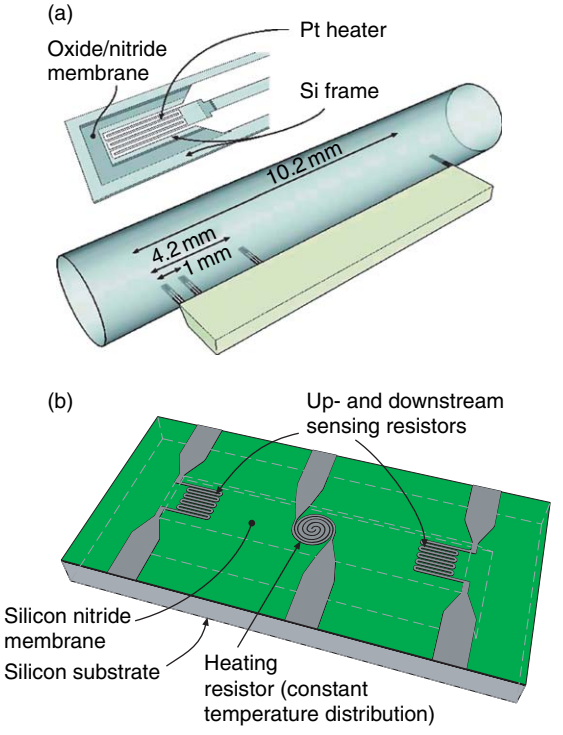


Figure 16 Thermal time-of-flight sensors: (a) chip with heater and sensor on cantilevers. (Source: Shoji and Esashi 1994). (b) Heater with constant temperature (CT) layout to avoid stress in the supporting membrane. (Source: Bedö G, *et al.* 2000).

Yang and Soeberg (1992) constructed a sensor by placing three diodes spaced 1500 µm from each other. The center diode is heated with a square wave signal, and the phase difference with the first harmonic of the measured temperature signal at the downstream diode is measured. The temperature signal at the downstream diode is measured relative to the reference upstream diode. This protects the amplifier inputs by compensating for fluid temperature variations. To be able to measure phase differences larger than 360°, the amplitude of the first harmonic is also taken into account. They were able to measure and theoretically model the flow of lamp oil in a 300 µm×1500 µm channel in the range of 10 down to 0.1 ml min^{-1}, with a resolution of 0.5 µl min^{-1} in the range 0.1–0.5 ml min^{-1}. The heater was driven at frequencies from 2 to 5 Hz, the sensitivity increasing with higher frequencies.

Phase-based time-of-flight sensors have also been presented by Bedö *et al.* (2000), Lambert (1986), and Lambert and Harrington (1986). Bedö and coworkers solve the 1D combined diffusion and convection heat transport equation for a sinusoidal temperature variation at a point in the flow. Similar to the findings of van Kuijk *et al.* (1995), a detection limit for flow velocity can be formulated at the point where conductive heat transfer becomes larger than convective heat transfer. The angular frequency below which the conductive heat transfer starts dominating is calculated for nitrogen and water in the described set-up and is found to be an issue only for nitrogen since the critical frequencies for water and nitrogen with a 2 mm heater/sensor distance are 10 mHz and 1.5 Hz, respectively. The average flow velocity above which the phase at the sensing resistor would become insensitive to the flow was also calculated for air and water and was found to agree qualitatively with the measurements. The flow sensor was tested in water and nitrogen at 1 and 4 Hz up to flow velocities of 8 m s^{-1}. To reduce thermal stress that could break the membrane upon which the heater is placed, the resistor was designed to have a CT distribution (see **Figure 16(b)**, the upstream resistor can be used for temperature compensation, but was not necessary.) Lambert (1986) used thermopiles as detectors on a polyimide film on silicon placing them 146 µm from the heater. The shorter heater/sensor distance results in a higher excitation frequency, 154 Hz (with air), and thus a faster response time.

Ashauer *et al.* (1999) have elegantly shown how a basic design of silicon nitride membrane with a polysilicon resistor and thermopile sensors on each side of

the resistor can be used in all three modes by altering the distance between the heater and the hot junctions of the thermopiles. If the hot junctions are placed very close to the heating resistor, the heater and the sensor will act as a single thermal entity, and the sensor can be used as a thermal anemometer. By placing the hot junctions at a larger distance, the device can be used in calorimetric (for low flow velocities) and time-of-flight mode (for high flow velocities). The calorimetric mode could be used from 0.125 up to 8 mm s^{-1}, after which the time-of-flight mode provided valid measurement values up to 128 mm s^{-1}.

2.07.2.3.5 Multiaxis thermal flow sensors

Flow is a 3D phenomenon, and, although in specific cases such as laminar flow through a tube the entire flow profile in the volume of interest can be derived from one single-axis measurement, in many cases one is interested in all three components of the flow. The batch fabrication methods used in microsystems lend themselves well for the integration of multiple flow sensors on one chip, providing an advantage over conventional methods, where the fabrication complexity and hence the cost for conventional flow sensors increase dramatically as multiple axes are to be measured. As an example, conventional hot wire probes roughly double in price for each additional axis.

Calorimetric flow sensors are the most straightforward devices for extending into a second dimension because placing a pair of sensors orthogonal to the first pair usually does not pose any structural problems. Caution needs to be exerted, since thermal crosstalk between sensors measuring in different directions can reduce the sensitivity (Kim *et al.* 2004). Hot-chip 2D sensors have been proven to work (Liu *et al.* 1997, Makinwa and Huijsing 2002a, b, van Oudheusden and Huijsing 1990), and have also been commercialized in wind sensors because of their ease of fabrication (Mierij Meteo 2006) (**Figure 17**).

Most types of calorimetric sensors have been shown to be relevant for adaptation to 2D measurements. To achieve thermal insulation, the entire structure with heater and sensors can be placed on a membrane (Kim *et al.* 2004, Park *et al.* 2003b, Robadey *et al.* 1995), or a cross-shaped half-open channel can be made to enable sensor bridges to be used (Fürjes *et al.* 2004). Hsieh *et al.* (1991, 1995a) designed a 2D pyroelectric flow sensor where one of the sensing electrodes was shared. The drawback of this reduced construction was that the two components needed to be sensed sequentially. Kim *et al.* (2004) and Yoon and Wise (1992) obtained a direction-sensitive sensor by using a thermal anemometer to measure the flow rate and a calorimetric flow sensor only for direction measurement.

Instead of having specific 2D calorimetric flow sensors that share the heater, it is most often possible to place a complete 90° rotated sensor on the same chip, next to the original one. Achieving 3D measurements is more complex for calorimetric sensors, as it is difficult to fabricate a sensor–heater–sensor structure in the direction out of the wafer plane. A two-element calorimetric sensor is somewhat easier and has been presented by de Bree (2003). Sensitivity in the third dimension is achieved by a pair of wires above each other, one on each side of the wafer, as illustrated in **Figure 18**.

For thermal anemometers to function in multiaxis applications, an asymmetry needs to be present in the design of the heating/sensing element. In practice, this implies that the sensor is a wire and cannot be, for example, a small circular meandering resistor.

Two-dimensional in-plane sensors have been designed by Breuer *et al.* (1999), Kälvesten *et al.* (1996), Xu *et al.* (2005a), and Yoshino *et al.* (2001) to measure wall shear stress. With the latter two, the wires are placed on a zig-zag line to get a two-component measure of the shear stress along this line (**Figure 19**).

Out-of-plane anemometers have been achieved by using folding structures. Ebefors *et al.* (1998) filled KOH-etched V-grooves with polyimide and used the shrinkage caused by curing to fold up 30 μm-thick silicon prongs of a hot wire. Each V-groove causes a limited angular deflection, depending on the curing parameters. By combining several grooves and a temporary bending stop (in place only during fabrication), the total deflection can be set to exactly 90° (**Figure 20(a)**). Designed to measure the small eddies in turbulent flows, the wires are 500 μm long and have a $2 \times 3\,\mu m^2$ cross-section. The design of the sensor allows it to be used in the same way as a conventional hot wire. Chen *et al.* (2003) fabricated a wall-mounted 3D hot-wire probe. Illustrated in **Figure 20**, the surface-micromachined prongs with the wires are released from the substrate by etching a sacrificial copper layer and bent up using an external magnetic field. The 2.7-μm-thick polyimide prongs are coated with a permalloy thin film to obtain this effect. After bending, the nickel hinges are strengthened by electroless plating. The entire process is performed at low temperature, providing freedom in the choice of substrate material. Pure time-of-flight flow sensors do not lend themselves to flow-direction sensing and have as yet not been presented in the literature.

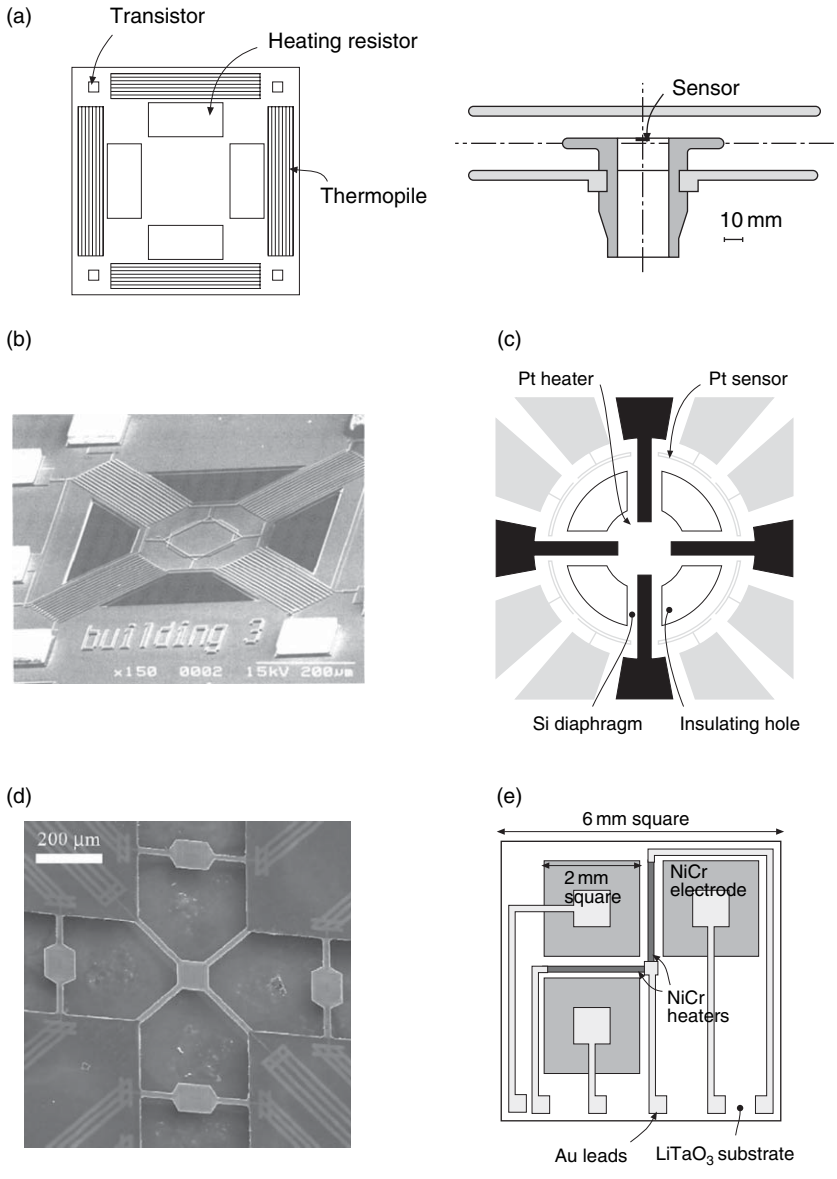


Figure 17 (a) Hot-chip flow sensor and its wind-meter packaging. (Source: van Oudheusden B, Huijsing J 1990 An electronic wind-meter based on a silicon flow sensor. *Sens. Actuators A (Phys.)* **A22(1–3)**, 420–4. © 1988 Elsevier.) (b) Calorimetric flow sensor with thermopile readout. (Source: Robadey J, Paul O, Baltes H 1995 Two-dimensional integrated gas flow sensors by CMOS IC technology. *J. Micromech. Microeng.* **5(3)**, 243–50.) (c) Calorimetric flow sensor with resistive sensors on a membrane. (Source: Kim S, Nam T, Park S 2004 Measurement of flow direction and velocity using a micromachined flow sensor. *Sens. Actuators A Phys.* **114(2–3)**, 312–18. © 2004 Elsevier.) (d) Calorimetric flow sensor with bridge-type resistive sensors. (Source: Fürjes P, Légrádi G, Dücso C, Aszódi A, Bársony I 2004 Thermal characterisation of a direction dependent flow sensor. *Sens. Actuators A Phys.* **115(2–3)**, 417–23. © 2004 Elsevier.) (e) Pyroelectric flow sensor with shared electrode. (Source: Hsieh H, Spetz A, Zemel J 1995a Pyroelectric anemometry: Vector and swirl measurements. *Sens. Actuators A (Phys.)* **A49(3)**, 141–7. © 1995 Elsevier.)

2.07.3 Mechanical Flow Sensors

Mechanical flow sensors are perhaps the most intuitive and straightforward class of flow sensors. As the wind can be seen blowing through the trees, it is the mechanical effect of fluid flows on the surroundings that provides an indication of the direction and the magnitude of the flow.

The force that a body in a fluid flow experiences can be broken down into two components, one in the

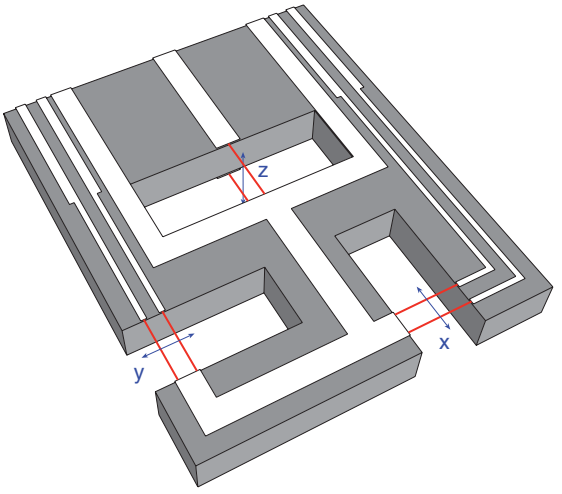


Figure 18 Three-dimensional microflown. (Source: de Bree H-E 2003).

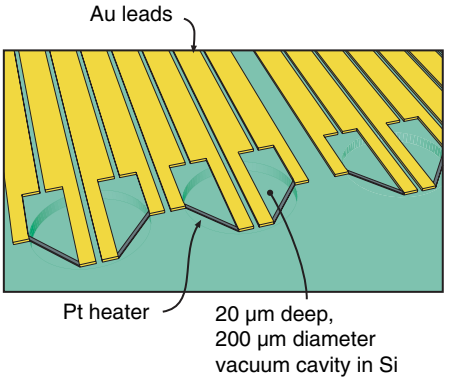


Figure 19 Row of two-component shear-stress sensors. (Source: Breuer *et al.* 1999).

direction of the flow, the drag force F_D, and one perpendicular to the flow, the lift force, F_L (**Figure 21**). For turbulent flows, these two forces can be described using the following equations:

$$F_D = C_D \frac{\rho U_\infty^2}{2} A$$

$$F_L = C_L \frac{\rho U_\infty^2}{2} A$$

where C_D and C_L are the drag and lift coefficients, dimensionless values dependent on the shape of the body, viscosity of the medium, and the roughness of the body's surface; ρ is the fluid's density; U_∞ the free-stream fluid velocity, and A the area of the body. For laminar flows, C_D becomes linearly proportional with the inverse of the flow velocity (Nakayama 1999). This results in a linear dependence of the drag force on the flow velocity.

Drag can be broken up into two components: induced drag and parasitic drag. Induced drag is caused by the production of lift while parasitic drag is defined as the remaining part. Parasitic drag in turn can also be broken up into two components: pressure drag and viscous drag. Pressure drag is caused by the pressure difference between the upstream and downstream parts of the body while viscous drag is caused by the skin friction of the fluid over the surface of the body.

Although both the lift force and the drag force can be used for obtaining a measure of the fluid velocity, the great majority of micromachined mechanical flow sensors use the latter. The definition of pressure drag also highlights the vague line between the class of differential-pressure sensors and that of pressure drag flow sensors. Differential-pressure sensors, described below, can be seen as drag flow sensors that measure the pressure difference caused by the drag of the orifice. In this chapter, the distinction is made based on the mechanical structure used for measuring the pressure drag. These devices using membranes to measure the pressure difference are classified as differential-pressure sensors, all others are considered drag flow sensors.

Optical readout of mechanical sensors provides a fully nonelectrical interface to the flow, which can be desirable in explosive or flammable environments. Mechanical flow sensors can be divided into three groups, based on the type of moving element used.

2.07.3.1 Drag-force Sensors

2.07.3.1.1 Cantilever flow sensors

The largest group of mechanical flow sensors are those that measure the deflection of a cantilever under the influence of the flow-induced drag forces. Gass *et al.* (1993) and Nishimoto *et al.* (1994) demonstrated the applicability of cantilever flow sensors in liquid. The liquid is guided past the cantilever, and the flow regime is laminar, which results in a linear deflection-flow dependence. The deflection is measured using piezoresistive strain sensors. Cantilever flow sensors with piezoresistive readout in air have been used by Su *et al.* (1996), van der Wiel *et al.* (1995), and Zhang *et al.* (1997).

To achieve maximum deflection while retaining mechanical stability, the cantilever is weakened at the clamp point where the sensing resistors are positioned. This can be done by thinning the beam in the out-of-plane direction (Gass *et al.* 1993) or by making it narrower in the wafer plane (Mireles 2005, Su *et al.* 2002, Zhang *et al.* 1997).

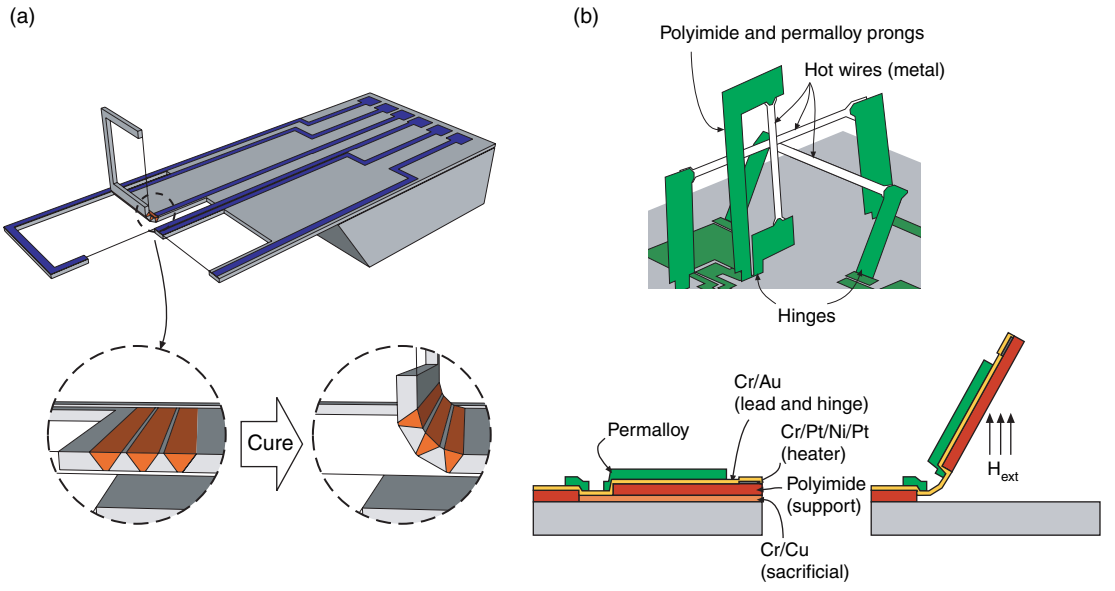


Figure 20 Three-dimensional hot-wire sensors, the out-of-plane structure obtained by (a) curing shrinkage of polyimide (Ebefors *et al.* 1998, © 1998 IEEE.); (b) plastic deformation of nickel hinges with an external magnetic field (H_{ext}). (Sources: Chen J, Liu C 2003 Development and characterization of surface micromachined, out-of-plane hotwire anemometer. *J. Microelectromech. Syst.* **12(6)**, 979–88, © 2003 IEEE, see also Chen *et al.* 2003.).

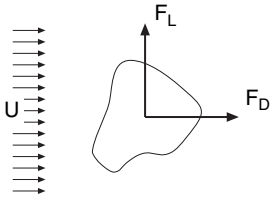


Figure 21 Drag and lift forces on a body in a fluid.

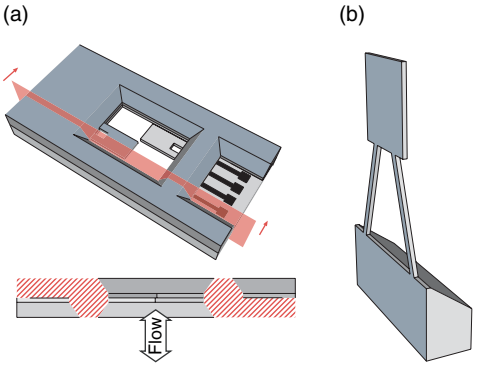


Figure 22 (a) Dual-cantilever bidirectional gas flow sensor. (Source: van der Wiel *et al.* 1995.) (b) Cantilever with weak support beams for high sensitivity. (Source: Su *et al.* 2002.)

A bidirectional mechanical gas flow sensor featuring sensitivity independent of the direction of the flow in a pipe has been fabricated by van der Wiel *et al.* (1995) (**Figure 22(a)**). Two identical chips with a cantilever are bonded to each other, resulting in a differential signal from the piezoresistors that is independent of the flow direction. An important design parameter of the cantilevers was the resonance frequency as it limits the measurement range as the cantilever starts resonating at high flow rates. The sensitivity of the Wheatstone bridge lay between 0.04 and 0.12 $\mathrm{mV\,V^{-1}(mg\,s^{-1})^{-2}}$ in air. Su *et al.* (1996, 2002) could present more sensitive airflow sensors with sensitivities $(\Delta R/R)/U_\infty^2$ ranging from 0.6 to 4.5×10^{-5} $\mathrm{(m\,s^{-1})^{-2}}$ by extending the support beams (see **Figure 22(b)**).

The cantilevers fabricated by Zhang *et al.* (1997) measure both pressure drag and viscous drag. Up to a flow regime of 10 $\mu\mathrm{l\,min^{-1}}$, the viscous drag forces are dominant, and the force on the cantilevers is linearly dependent on the flow rate. Above 100 $\mu\mathrm{l\,min^{-1}}$, the pressure drag forces dominate and the force on the cantilever is quadratically dependent on the flow. The transition region is nonlinear.

Chun *et al.* (1998) demonstrated the feasibility of using a waveguide cantilever as a flow sensor. The optical coupling the waveguide on the cantilever and a fixed waveguide is a measure for the flow rate over the cantilever. Mireles (2005) suggested a drag flow sensor based on the transfer of a surface-micromachined cantilever mirror structure to an optical fiber (**Figure 23(a)**). The drag-induced bending of the

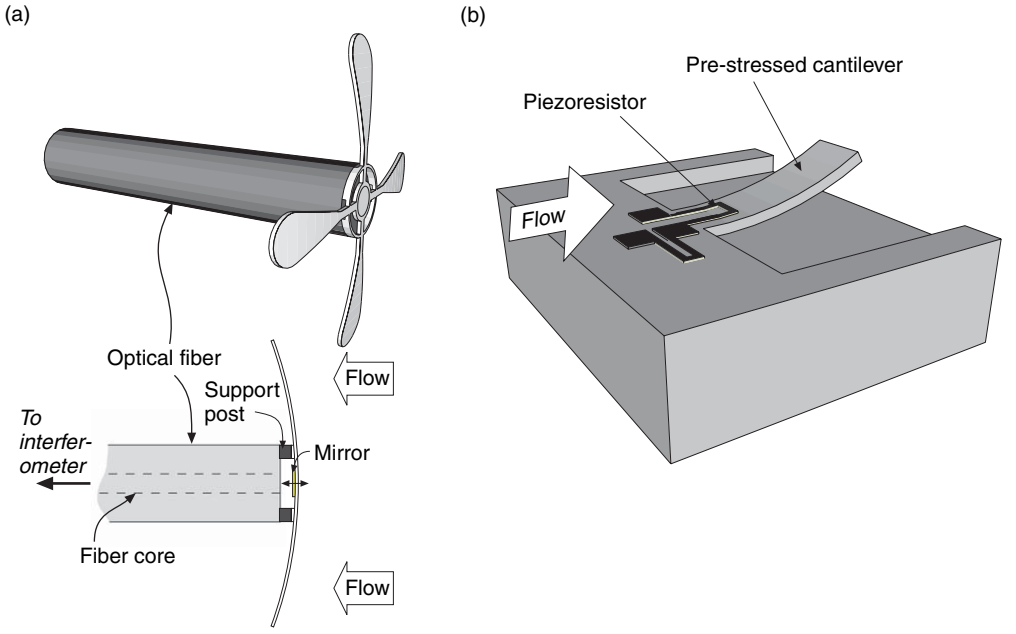


Figure 23 (a) Drag flow sensor with optical readout. (Source: Mireles 2005.) (b) Prestressed cantilever flow sensor. (Source: Kim *et al.* 2003.)

four cantilevers cause the mirror to move away from the fiber, a movement that can be measured interferometrically.

Go *et al.* (2001) and Kim *et al.* (2003) presented prestressed cantilever flow sensors. The flow direction for these is in the length direction of the cantilevers, and the deflection decreases as the flow increases (see **Figure 23(b)**).

While most sensing cantilevers are designed to have resonant frequencies above 10 kHz to avoid environmental excitation, Go *et al.* (2001), Su *et al.* (1996), and Zhang *et al.* (1997) deliberately designed arrays of cantilevers with a natural frequency around 1 kHz. At free-stream flow velocities of 3–6 m s^{-1}, the cantilevers vibrate at this frequency. Although the main objective of the structure is to generate cooling, the magnitude of the vibration turned out to increase with increasing flow and was measurable with the integrated piezoresistors. Radhakrishnan and Lal (2005) (**Figure 24**) used electroplating to form an out-of-plane cantilever that exerts a twisting stress on the substrate, which is measured with piezoresistors.

2.07.3.1.1.(i) Hair-type flow sensors A subclass of the cantilever drag flow sensors are those modeled after animal sensory hairs. Ozaki *et al.* (2000) fabricated a 2D airflow sensor that consisted of a wire manually glued to a cross-shaped beam

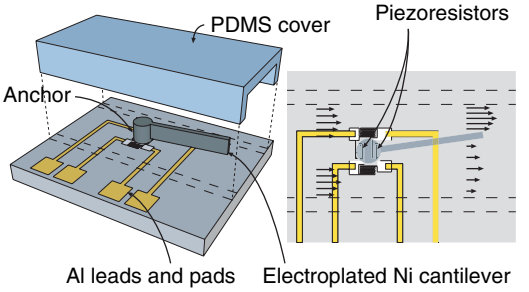


Figure 24 Electroplated beam, twisting when exposed to flow. (Source: Radhakrishnan and Lal 2005.)

structure with integrated piezoresistors (**Figure 25**). Fan *et al.* (2002b) were able to eliminate the manual assembly step by using plastic deformation magnetic assembly (PDMA). With PDMA (Zou *et al.* 2001), a magnetic field is used to fold up a metal cantilever coated with a permalloy. The hinge is plastically deformed causing the cantilever to stay in position after it has been removed from the magnetic field. Water flow is measured by measuring the deflection of a support beam that carries the vertical cantilever. Arrays of such beams oriented in different directions can be used to obtain a full picture of the flow near the surface (Chen *et al.* 2003).

The behavior of cricket sensory hairs has been imitated by van Baar *et al.* (2005) (see **Figure 25**)

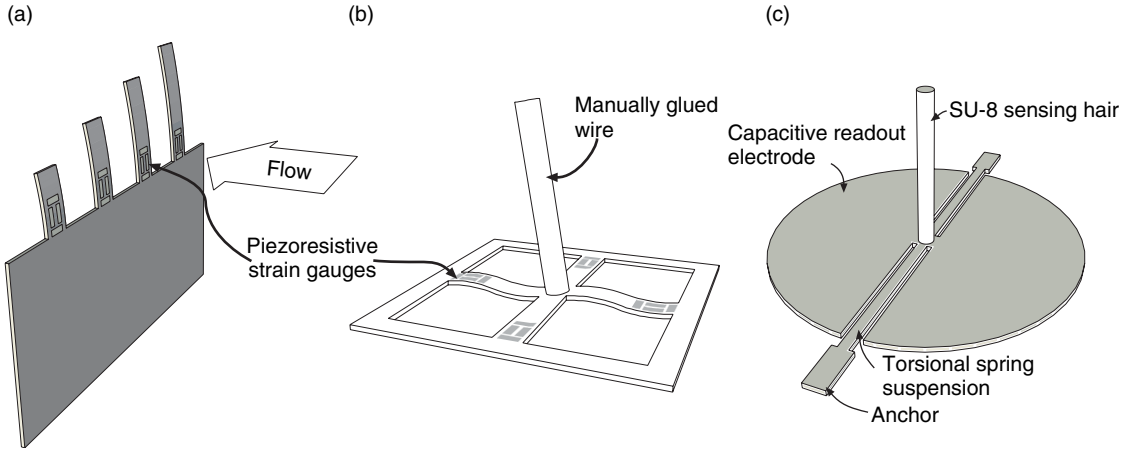


Figure 25 Three variations of hair-type flow sensors: (a) single-axis cantilever beams. (Source: Ozaki Y, Ohyama T, Yasuda T, Shimoyama I 2000 An air flow sensor modeled on wind receptor hairs of insects. *Proc. IEEE 13th Annu. Int. Conf. Micro Electro Mechanical Systems*, pp. 531–6. © 2000 IEEE.) (b) Dual-axis with piezoresistive readout. (Source: Ozaki Y, Ohyama T, Yasuda T, Shimoyama I 2000 An air flow sensor modeled on wind receptor hairs of insects. *Proc. IEEE 13th Annu. Int. Conf. Micro Electro Mechanical Systems*, pp. 531–6. © 2000 IEEE.) (c) Single-axis with capacitive readout. (Source: van Baar J, Dijkstra M, Wiegerink R, Lammerink T, De Boer R, Krijnen G 2005 Arrays of cricket-inspired sensory hairs with capacitive motion detection. *Proc. IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS)*, pp. 646–9. © 2005 IEEE.)

who made hairs out of either silicon-rich nitride (van Baar *et al.* 2003, 2005) or SU-8 (van Baar *et al.* 2005, Dijkstra *et al.* 2005), supported by a disc suspended with a gimbal structure. The SU-8 hairs are produced using a surface micromachining process, while the silicon-rich nitride hairs are fabricated using a deep reactive ion-etched silicon wafer as a sacrificial mold. The rotation of the supporting disc is measured capacitively. Single-axis suspended discs have shown a directional sensitivity to flow, and the process allows for dual-axis sensors as well.

Similar hairs have been fabricated by Engel *et al.* (2005) using polyurethane. First, polyurethane is mixed with a conductive filler and molded on a glass substrate with electrodes to form a pressure-conductive rubber pattern (Shida and Yuji 1996). Two types of fillers were used: carbon black and multiwalled carbon nanotubes (MWCNTs), the latter yielding a larger sensitivity. On these force-sensitive resistors, polyurethane hairs are placed using a lost-wax process. The hairs are extremely flexible, making them very robust. A drawback of the method is the large device-to-device variation, requiring the averaging of several hairs to obtain reliable results.

Although many of these solutions can be fabricated in array format, electrical readout becomes a problem as the number of leads out to the array edge rapidly increases. Brücker *et al.* (2005) provided a solution by using an optical readout to measure the deflection of each hair in a sensory array. The hairs are made by a lost-mold process where holes are laser-drilled in a wax sheet to form a mold for polydimethylsiloxane (PDMS) rubber (Schmitz *et al.* 2005). Pattern recognition is employed to identify the deflection of each individual hair and provides a full picture of the liquid flow over the surface. Drawbacks that optical access to the top surface is required, and that vibrations might affect measurements.

A sensor that combines some characteristics of hair-type sensors with the shear sensors described below is shown in **Figure 26** (Dao *et al.* 2002, 2003).

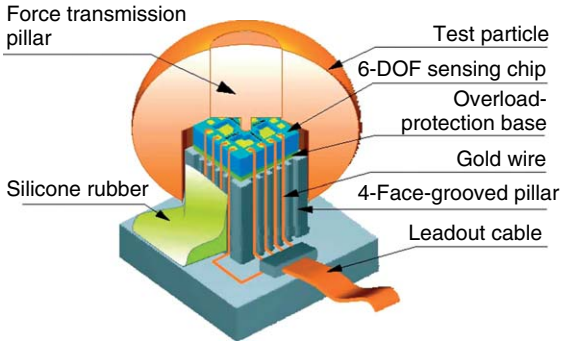


Figure 26 Shear-stress and vortex sensor with six degrees of freedom. (Source: Dao D, Toriyama T, Sugiyama S, Nguyen A, Wells J 2003 A MEMS-based microsensor to measure all six components of force and moment on a near-wall particle in turbulent flow. *TRANSDUCERS '03. 12th Int. Conf. Solid-State Sensors, Actuators and Microsystems. Dig. Tech. Pap.*, Vol. 1, pp. 504–7. © 2003 IEEE.)

It consists of a 8–9-mm-diameter polyethylene-sensing particle mounted on a micromachined force sensor with six degrees of freedom: three translational and three rotation. Polyethylene was chosen because its density lies very close to that of water, resulting in minimum buoyancy. The force sensor uses 18 piezoresistors, out of which 2 are shear piezoresistors, a four-terminal piezoresistor that senses the shear force instead of the normal force on a piezoresistor (Pfann and Thurston 1961). The sensor is placed such that the top half of the sphere protrudes from the surface and measures both the drag and the shear forces in a multiphase liquid flow.

2.07.3.2 Lift-force Flow Sensors

As opposed to the drag flow-based sensors, lift force-based sensing are less common. The only nonturbine sensor was developed by Svedin *et al.* (1998a, b, 2003a, b) who used a pair of airfoil plates attached to a flexible polysilicon beam. By placing this sensor at an angle in the flow, a differential lift force is generated (**Figure 27**). Several improvements were presented on the original design. Use of a Wheatstone bridge to measure the strain difference between the upstream and the downstream plate allowed reduction of common-mode effects such as accelerations and temperature variations (Svedin *et al.* 1998a). By placing strain sensors on detector beams at the edges of the plates (see **Figure 27**), the natural frequency of the device was increased by a factor 3, while the sensitivity was retained or improved. The higher natural frequency implies an improved time response, a parameter that, for mechanical flow sensors, is usually a trade-off against sensitivity. Like all mechanical flow sensors, the lift-force sensor has a low sensitivity at low flow rates. Svedin *et al.* (2003a) added a heating resistor at the center of each wing to exploit the high low-flow sensitivity of thermal sensors. The resistors were thermally isolated from the rest of the wings by polyimide-filled KOH-etched grooves. The resistors were evaluated separately, in CT mode, and in conjunction with each other, as a two-element calorimetric sensor. The thermal anemometry measurements provided a higher flow sensitivity, while the calorimetry-based measurements could indicate the flow direction, but with a lower sensitivity and a nonuniform flow response. The measurements in **Figure 27(c)** illustrate the complementary use of the high sensitivity ranges.

2.07.3.2.1 Turbine flow sensors

A common method of measuring flow in the macroworld is using turbines. Because of stiction and friction problems they are much less popular in microelectromechanical systems (MEMS). Rapoport and Reed (1991) fabricated a surface-micromachined rotor. A microscope with stroboscopic lighting was used to measure the rotation rate of the rotor for water and nitrogen flow. To be able to measure water flow, a glass plate was placed over half the rotor to generate the shear stress required to rotate the rotor. Friction between the rotor and the axle was a serious problem, so a minimum required flow rate before the rotor would start rotating. The intended application of the sensor was blood-flow measurements, but no measurements were presented, and significant deposition of red blood cells on the sensor was observed during attempts.

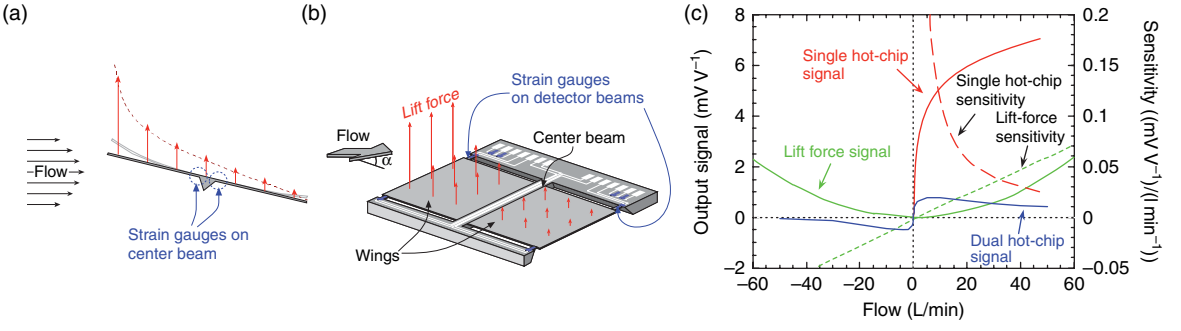


Figure 27 Lift-force sensor: (a) principle; (b) two piezoresistive readout methods, at the base of the wing and at the edges; and (c) schematic of the signals from the thermal and lift flow sensors. (Sources: Svedin N, Kalvesten E, Stemme G 2003. A lift force sensor with integrated hot-chips for wide range flow measurements. *Sens. Actuators A (Phys.)* **A109(1–2)**, 120–30; Svedin N, Kälvesten E, Stemme G 2003. A new edge-detected lift force flow sensor. *J. Microelectromech. Syst.* **12(3)**, 344–54; Svedin N, Stemme E, Stemme G 2003. A static turbine flow meter with a micromachined silicon torque sensor. *J. Microelectromech. Syst.* **12(6)**, 937–46. © 2003 Elsevier, courtesy of Niklas Svedin.)

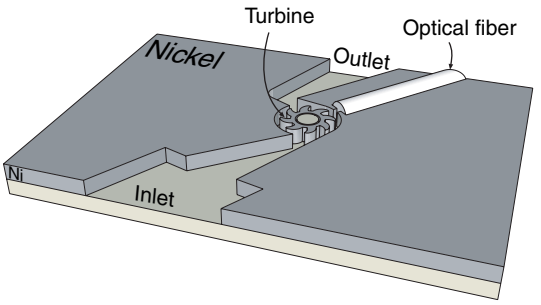


Figure 28 LIGA (*Litografi Galvanisierung Abformung*, a combination of deep-etch X-ray lithography, electroforming, and molding techniques)-fabricated turbine with optical readout. (Source: Himmelhaus *et al.* 1992.)

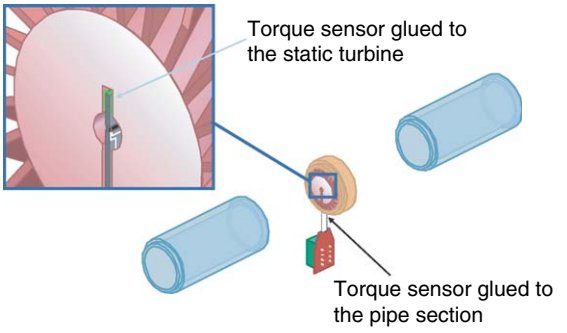


Figure 29 Static turbine flow sensor. (Source: Svedin *et al.* 2003c), courtesy of Niklas Svedin.

Himmelhaus *et al.* (1992) used LIGA (Litografi Galvanisierung Abformung, a combination of deep-etch X-ray lithography, electroforming, and molding techniques) (Becker *et al.* 1986) to fabricate a 250-μm-diameter, 200-μm-high turbine rotor (**Figure 28**). An optical fiber was used both to illuminate and to read out the reflection off the blades. An oscillating signal was obtained as the turbine rotated. This approach also suffered from reliability issues due to friction. The microturbine designed by Holmes *et al.* (2004), intended to generate a flow-dependent power, had similar problems. Despite using ball bearings to facilitate the rotation of the SU-8 rotor, laser-machined to achieve slanted blades, the generated power was only 1% of that predicted by computational fluid dynamics software.

The third optical readout method of turbine rotation used particle image velocimetry (PIV) algorithms to identify the movement of the turbine blades. Han *et al.* (2004) used infrared light to see through the caps of an all-silicon bulk-micromachined microturbine (Frechette *et al.* 2000).

A solution to the friction problems encountered with micromachined turbines has been presented by Svedin *et al.* (2003c). The torque that a fixed turbine wheel exerts on its axle is used as a measure for the flow rate past the blades of the turbine (**Figure 29**). The torque was measured by means of piezoresistors placed at the center of rotation where the silicon chip had been weakened by deep reactive ion etching. The signal of the sensor was shown to be quadratically dependent on the volume flow. Varying the angle and the length of the turbine blades resulted in a trade-off between pressure drop and sensitivity: short turbine blades at a sharp angle to the flow provide not only the highest sensitivity but also the largest pressure drop.

2.07.3.3 Skin-friction Sensors

The third class of mechanical flow sensors is based on the skin-friction effect, also known as wall shear stress. A fluid flowing over a surface exerts a shear stress τ that can be expressed as follows:

$$\tau_0 = \mu \left. \frac{\partial U_x}{\partial y} \right|_{y=0}$$

where μ is the dynamic viscosity of the fluid, x the direction of the flow, and y the direction normal to the surface. Although measuring the wall shear stress can give information on the flow of the fluid near the surface, wall shear stress of itself is also a topic of research in fluid dynamics.

Mechanical skin-friction sensors are sometimes also called direct sensors, as they directly measure the skin friction in contrast to, for example, thermal sensors, which measure a secondary effect. A general trade-off in mechanical skin-friction sensors is sensitivity versus resolution: resolution increases with decreasing measurement surface but results in decreasing force.

Often, the deflection of a spring-suspended element, called the floating element, is measured. Schmidt *et al.* (1987) introduced one of the first micromachined direct skin-friction sensors. It consists of a floating element suspended by four tethers, and the deflection of the floating element is linearly dependent on the shear force upon it. This design has been presented in several variants with different materials and readout methods (Ng *et al.* 1991, Padmanabhan *et al.* 1995, Pan *et al.* 1995a, Schmidt *et al.* 1987). **Figure 30** shows five variants of this type. The device in **Figure 30(b)** consists of a polyimide floating element with an embedded chromium electrode on a silicon surface. The 30-μm-thick polyimide is spun in several steps and is etched using an oxygen plasma,

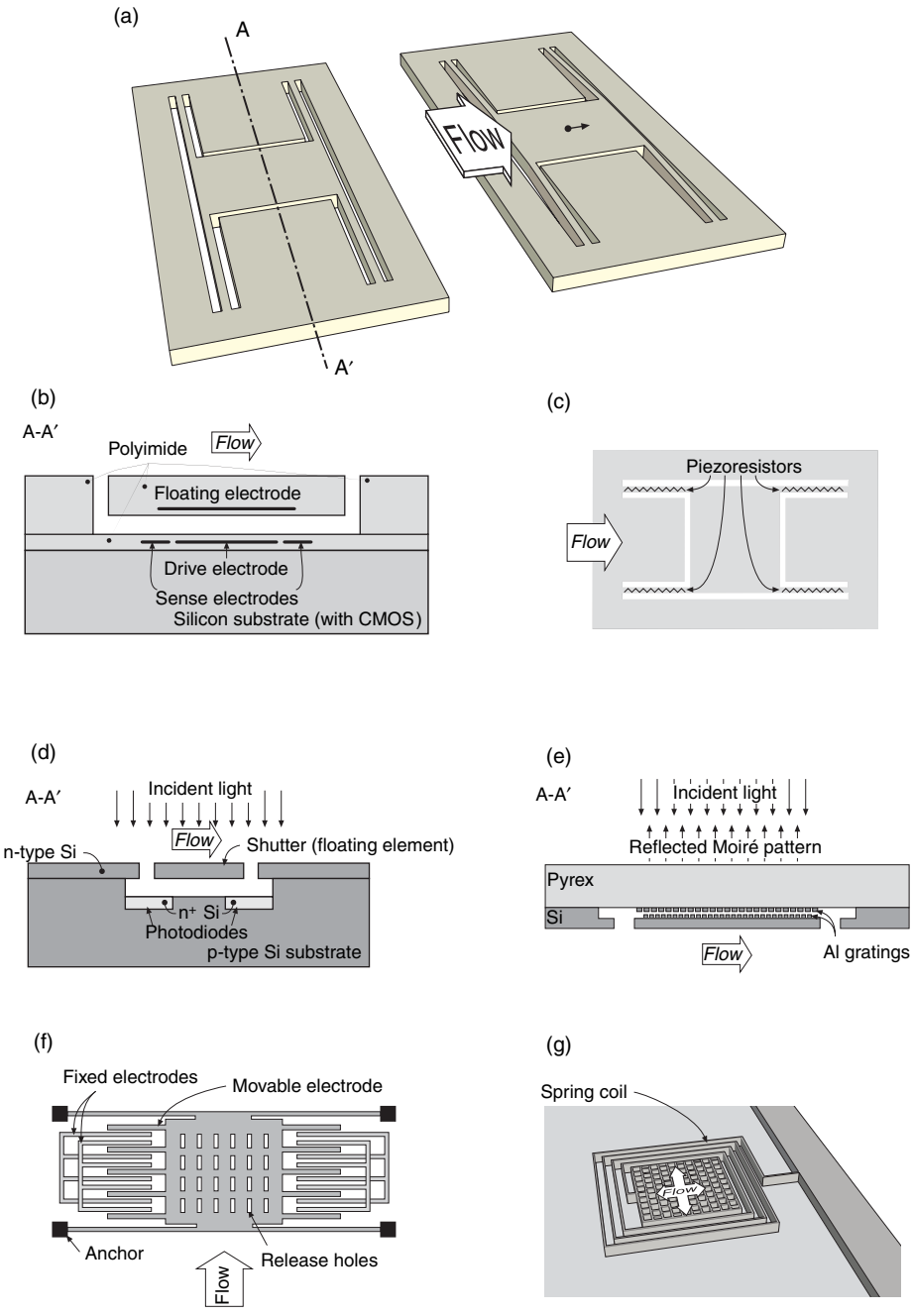


Figure 30 Readout methods for floating-element skin-friction sensors: (a) four-tether skin-friction sensor and its most common sensing mode. (b) Capacitive using in-plane capacitors. (Source: Schmidt *et al.* 1987.) (c) Piezoresistive, note the differing sensing mode. (Source: Ng K-Y, Shajii J, Schmidt M A 1991 A liquid shear-stress sensor fabricated using wafer bonding technology. *Solid-State Sensors and Actuators 1991 Digest of Technical Papers, TRANSDUCERS '91*, pp. 931–4. © 1991 IEEE.) (d) Optical using integrated photodiodes. (Source: Padmanabhan A, Goldberg H, Breuer K, Schmidt M 1995 A silicon micromachined floating element shear-stress sensor with optical position sensing by photodiodes. *8th Int. Conf. Solid-State Sensors and Actuators and Eurosensors IX. Dig. Tech. Pap.*, Stockholm, Sweden, Vol. 2, pp. 436–9. © 1995 IEEE.) (e) Optical using moiré interference. (Source: Horowitz *et al.* 2004.) (f) Capacitive using out-of-plane capacitors. (Source: Pan T, Hyman D, Mehregany M, Reshotko E, Willis B 1995b Characterization of microfabricated shear stress sensors. *ICIASF '95 Record. Int. Congr. Instrumentation in Aerospace Simulation Facilities*, pp. 6–1. © 1995 IEEE.) (g) Spiral spring suspended with stress sensitive diode readout. (Source: Haronian *et al.* 1999.)

and the 3-μm capacitor gap is created using aluminum as a sacrificial layer. The displacement, orthogonal to the tethers, is measured using on-chip differential capacitor readout electronics. The size of the element, 0.5 mm × 0.5 mm, is restricted by the stress in the polyimide/metal layer that would cause the floating element to bend more than 1 μm. The device suffered from sensitivity drift as moisture variations caused the polyimide to swell. The high input impedance made it sensitive to electromagnetic noise interference and the air–dielectric interface accumulated charges resulting in drift (Sheplak *et al.* 2004). The same group at MIT presented a variant for measuring skin friction in liquid (see **Figure 30(c)**). Designed to measure much higher shear stresses (1–100 kPa instead of 1 Pa), the floating element is only 120 μm × 160 μm and is suspended by tethers of 10 μm × 30 μm. The moving parts are 5 μm thick and made of monocrystalline silicon bonded to a support wafer to form a silicon-on-insulator (SOI) structure. The sensitivity direction is parallel to the tethers, as two of the tethers are used as piezoresistors in a half bridge. Barlian *et al.* (2006) presented the same basic design for liquid skin-friction measurements, but used side-implanted piezoresistors to achieve higher shear-stress sensitivity and to reduce out-of-plane sensitivity. A further evolution of MIT sensor is shown in **Figure 30(d)** (Padmanabhan *et al.* 1995). For measuring with a resolution of below 1 Pa in air, the piezoresistive readout was not sensitive enough, and the high impedance of the capacitive sensors caused too much drift in the wind tunnel environment. The solution presented by Padmanabhan and coworkers uses two photodiodes to measure the displacement of the floating element. The photodiodes are integrated in the carrier substrate prior to the bonding of the device layer that will form the floating element. By measuring the differential current under illumination by a 7 mW pulsed laser at 1 kHz, shear stress can be measured very precisely. Limitations of this setup lay in the optical setup as vibrations or expansion of the tunnel would cause the light source to move and affect the measurements. A solution to these problems was presented by Horowitz *et al.* (2004). By placing slightly differing aluminum gratings on the floating element and its glass substrate, the displacement of the floating element can be measured by observing the moiré fringes. An advantage of this setup is that the packaged sensor can be mounted on a camera lens (see **Figure 30(e)**), eliminating vibration and thermal expansion effects of the wind tunnel.

An alternative for improving the capacitive readout shear stress sensor was presented by Pan *et al.* (1995b). The use of polysilicon avoids possible problems that polyimide can cause due to viscoelastic effects and absorption of humidity. **Figure 30(f)** also illustrates the use of folded beam suspensions that provide a lower stiffness and fully linear spring behavior. Pan and coworkers also demonstrated the use of force feedback on the floating element to measure the shear stress. For this, a variant of a commercial analog devices accelerometer (ADXL series) was evaluated for its applicability as a shear-stress sensor. Shear stresses up to 5 Pa could be measured before the force required for the feedback became too large to compensate.

Haronian (1999) observed the resonance of a floating element excited at its natural frequency by the flow. The floating element is suspended with a spiral spring (see **Figure 30(g)**), making it sensitive to shear stress in both in-plane directions. At the suspension point of the beam a pair of stress-sensitive diodes is used to differentially measure the stress created by the resonating mass. The amplitude and the frequency of the resonance were observed to increase with increasing flow rates, the latter nearly linearly for flows ranging from 5 to 8 ml s^{-1} in a channel of undisclosed dimensions.

A single-beam shear-stress sensor was presented by Zhe *et al.* (2005). Though electrodes are present for running in force-feedback mode, only the open-loop mode has been demonstrated. Tseng and Lin (2003) fabricated and tested a 2D shear-stress sensor where the floating element is a rectangular SU-8 pillar. The 200-μm-wide sides of the 400-μm-high beam are gold-coated and the bending of the beam is read out using two optical fibers. The cleaved fiber ends are also coated with a metal to form a semitransparent mirror. The cleaved interface and the gold-coated beam form the two reflectors for a Fabry–Perot interferometer.

Figure 31 shows the principle for an extremely small floating-element skin-friction sensor as conceived by Desai and Haque (2004). By designing the sensor so that it needs to be placed perpendicular with respect to the surface, only the floating element is exposed to the flow. This minimizes the interference with the flow and increases the spatial resolution drastically. Fabricated on an SOI wafer with a 20-μm device layer, the floating element's surface is 20 μm × 4 μm. Shear stress on the floating element causes the structure to twist around a central point resulting in measurable capacitance changes at

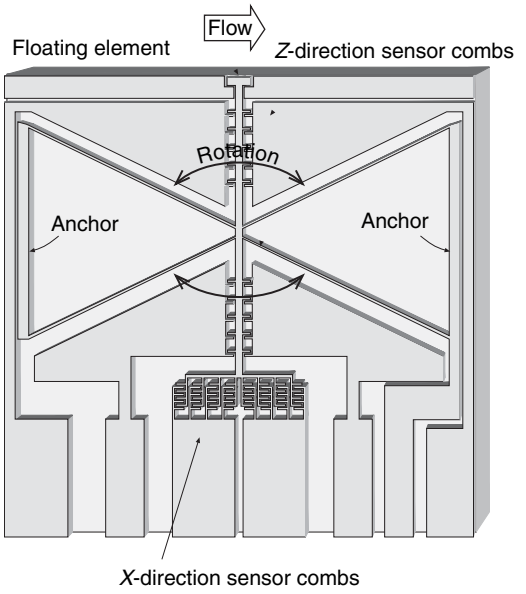


Figure 31 Capacitive floating-element skin-friction sensor by Desai and Haque (2004); the X-shaped spring structure allows rotation only around its center point.

six points. The bottom two capacitors, C_{x1} and C_{x2}, are measured differentially to measure the shear in the x-direction. The sum of the other four capacitances C_{z1-4} provides a measure for the shear in the z-direction. $C_{x1}-C_{x2}$ is not affect by the common mode that a z-shear causes, and the C_z capacitors are placed symmetrically around the point of rotation so that their sum is not affected by shear in the x-direction.

A cantilever-based skin-friction sensor has been introduced by von Papen *et al.* (2002, 2004), illustrated in **Figure 32**. Variants have been designed based on the shape and the number of the attachment points of a low and wide cantilever sticking up from the surface of the wall where the skin friction is to be measured. The variants differ in how they bend when placed in a flow. In variant A, the cantilever is solid, which causes it to bend in the plane perpendicular to both the flow and the wall. Variants B–D bend mainly in the plane perpendicular to the flow but parallel to the wall. For a fixed cantilever area, the sensitivity of the variants increases, variant A being least sensitive and D the most sensitive. The bending of all variants is measured using piezoresistors, strategically placed at the point of maximum bending. More information on micromachined shear-stress sensors can be found in Löfdahl and Gad-el Hak (1999b), Naughton and Sheplak (2002), and Sheplak *et al.* (2004).

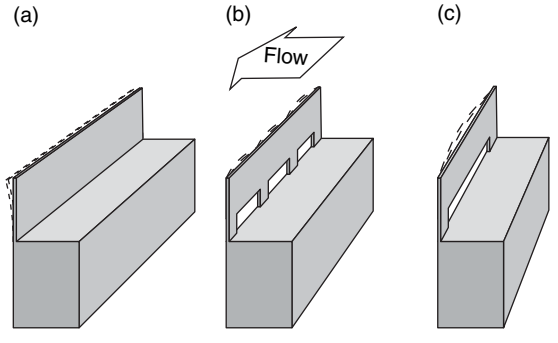


Figure 32 Cantilever-based skin-friction sensor. The bending mode of the cantilever is defined by the clamping points. (Source: von Papen T, Steffes H, Ngo H, Obermeier E 2002 A micro surface fence probe for the application in flow reversal areas. *Sens. Actuators A Phys.* **97–98**, 264–70. © 2002 Elsevier.)

2.07.4 Differential-pressure Flow Sensors

A fluid flowing through a channel undergoes a pressure drop that is dependent on the velocity of the fluid. By measuring the drop, one can measure the flow. In contrast to conventional differential-pressure flowmeters, where the flow is turbulent and proportional to the square root of the pressure drop (Bentley 1983a), the volume flow (Q_v) through microchannels is laminar and linearly dependent on the pressure difference, Δp, over the channel. For a no-slip boundary condition, this dependence is as follows (Boillat *et al.* 1995)

$$Q_v = \frac{1}{C}\frac{AD^2\Delta p}{2\mu l} \qquad [1]$$

where A is the channel cross-sectional area, D the equivalent channel hydraulic diameter, equal to $4A$ per wetted perimeter, l the length of the channel, μ the dynamic viscosity, and C a dimensionless friction factor. Caution should be exercised, however, since the Knudsen number reaches values above 0.01 in channel heights less than 8 μm for air at standard ambient pressure and temperature. Below these heights, the no-slip boundary condition and thus eqn [1] do not apply (Arkilic *et al.* 1997, Jang and Wereley 2004). Micromachined differential-pressure flow sensors can be divided into two groups based on whether the pressure difference is measured relatively or as a difference between two absolute pressures.

2.07.4.1 Relative Pressure Measurement

Cho *et al.* (1992) (**Figure 33(a)**) have presented a highly sensitive gas flow sensor whose fabrication is based on a dissolved wafer process. The channel structure and the pressure membrane are defined by p^{++} doped regions and bonded to a glass support before the rest of the wafer is KOH-etched using the p^{++} region as an etch stop. The pressure difference is read out capacitively. The pressure difference is generated by forcing the flow through a channel 60 µm wide, 3.5 µm high, and 8.4 mm long. The sensor is designed to measure process gases in semiconductor manufacturing. The same principle has been used for gas flow sensing by Furuberg *et al.* (2003), where the pressure difference is measured using piezoresistors (see **Figure 33(b)**).

An alternative approach for creating the pressure difference is to force the flow through an orifice in the pressure-sensing membrane as illustrated in **Figure 33(c)** (Nishimoto *et al.* 1994, Richter *et al.* 1999). In this case, the diameter of the orifice can easily become larger than the thickness of the membrane, in which case the flow through it will no longer be laminar, and the pressure difference will become quadratically dependent on the flow (Richter *et al.* 1999). Both Nishimoto *et al.* (1994) and Richter *et al.* (1999) use piezoresistors on the edge of the membrane for readout. An interesting detail is that Richter and coworkers used a commercially available pressure sensor (Honeywell 24PC) and created the orifice with a laser drill.

A differential-pressure sensor that does not require the flow to pass through it has been presented by Berberig *et al.* (1997) (see **Figure 34**). It is a direct miniaturization of the conventional Pitot tube whose operating principle is based on Bernoulli's law:

$$p_{\text{static}} + \rho g h + \frac{\rho U^2}{2} = cst$$

where p_{static} is the static pressure, ρ the fluid's density, h the height, and U the flow velocity. The pressure

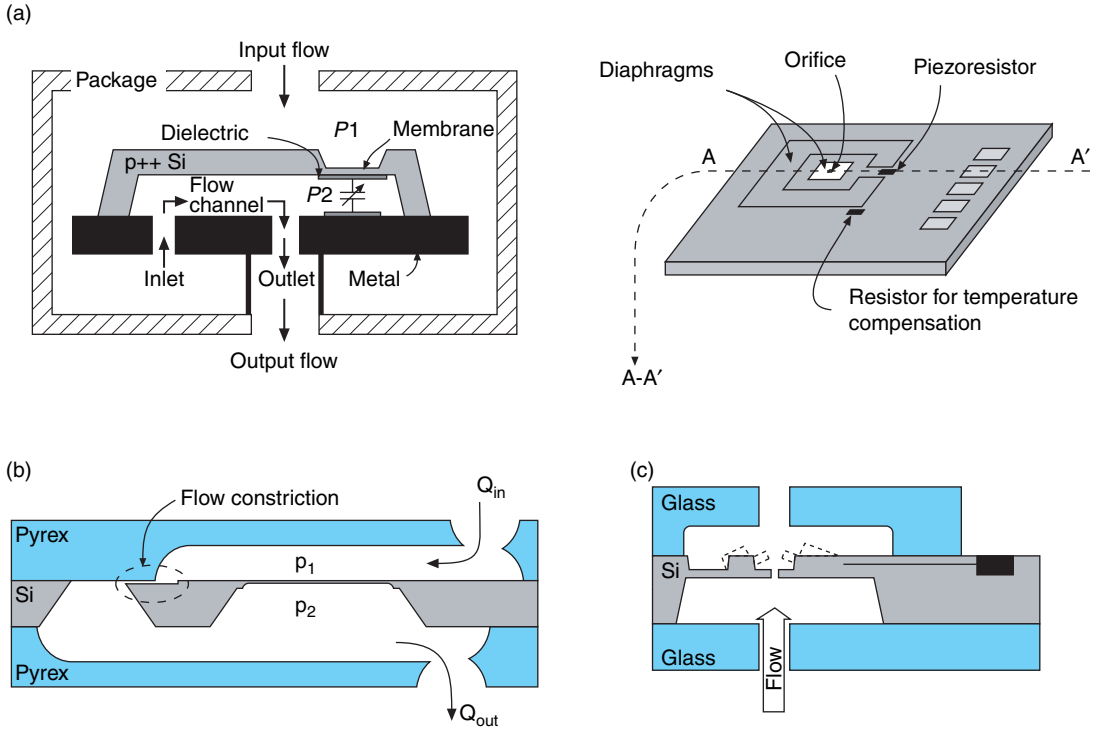


Figure 33 Cross-section of differential-pressure flowmeters: (a) using capacitive readout. (Source: Cho S T, Najafi K, Lowman C E, Wise K D 1992 An ultrasensitive silicon pressure-based microflow sensor. *IEEE Trans. Electron. Devices* **39(4)**, 825–35. © 2003 IEEE.) (b) Using piezoresistive readout. (Source: Furuberg L, Wang D, Vogl A, Merveille C, Hansen S I, Rogne H, Solli L 2003 New pressure sensor, optical mirrors and a flow rate sensor manufactured by bulk micromachining. *Proc. Sensor 2003*, Nuremberg, Germany, p. A6.3. Courtesy: SINTEF, Norway.) (c) Using piezoresistive readout on a post-processed orifice in the diaphragm of a pressure sensor to act as a flow sensor. (Source: Nishimoto T, Shoji S, Esashi M 1994 Buried piezoresistive sensors by means of MeV ion implantation. *Sens. Actuators A (Phys.)* **A43(1–3)**, 249–53. © 1994 Elsevier.)

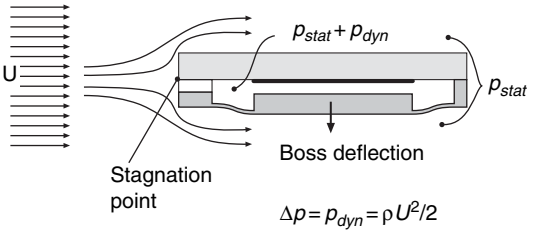


Figure 34 Micromachined Pitot tube flow sensor. (Source: Berberig O, Nottmeyer K, Mizuno J, Kanai Y, Kobayashi T 1997 Prandtl micro flow sensor (PMFS): A novel silicon diaphragm capacitive sensor for flow velocity measurement. *Int. Conf. Solid-State Sensors and Actuators, Proceedings*, Vol. 1, pp. 155–8. © 1997 IEEE.)

difference between the stagnation point at the front of the chip (where $U = 0$) and at the side of the chip (where $U = U_\infty$) becomes as follows:

$$\Delta p = \frac{\rho U^2}{2}$$

(the difference in height between the two points can be ignored). Since the cavity inside the chip is connected to the stagnation point, this pressure difference acts on the boss resulting in a measurable capacitance change. The deflection of the membrane is proportional to the pressure difference over it, and for small deflections, the capacitance changes proportionally to the deflection, resulting in a quadratic flow dependence of the capacitance: $|\Delta C| \sim U^2$. The sensitivity of this device increases linearly with flow and goes to zero as the flow is reduced.

A special type of differential-pressure sensor introduced by Sood and Ghosh (2004) is based on the interplay between the Bernoulli principle and the Seebeck effect. **Figure 35** illustrates how a sensing chip is placed at an angle to the flow to create a flow acceleration over the chip surface. Due to the adiabatic expansion of the gas flow between the stagnation point, at the leading edge of the chip and the far edge of the chip, a pressure difference, and hence a temperature difference, is generated. By using the Seebeck effect of the chip material, a voltage difference can be measured between the two ends of the chip. This voltage difference is shown to be a function of the square of the difference in Mach numbers, defined as the ratio between the fluid velocity, U, and the sound velocity in the fluid, c: $Ma = U/c$. Measuring in argon and nitrogen, two flow regimes were evaluated: incompressible flow $Ma^2 < 0.05$ and compressible flow $0.07 < Ma^2 < 0.20$. In both regimes, the voltage signal was linearly dependent on Ma^2, with a higher sensitivity in the lower range: approximately 300 μV in *n*-Ge at $Ma^2 = 0.05$ and 650 μV at $Ma^2 = 0.20$.

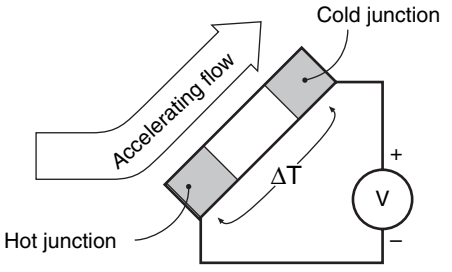


Figure 35 Chip placement in the flow. The shaded areas are the electrical connections. (Source: Sood and Ghosh 2004.)

2.07.4.2 Absolute Pressure Measurement

The pressure difference over a flow constriction can also be measured separately using two absolute pressure measurements. Flow sensors using pressure measurements relative to atmospheric pressure are also included in this category. The principle is illustrated in **Figure 36**. Piezoresistive (Boillat *et al.* 1995, Oosterbroek *et al.* 1999, Van Der Schoot *et al.* 2001) as well as capacitive (Oosterbroek *et al.* 1997) readout pressure sensors have been used. The pressure-sensing membranes in the mentioned devices are defined by wet etching, and the chips are anodically bonded to a glass substrate. An advantage of this type of sensor is that only the back side of the pressure sensor is exposed to the fluid while the electrodes and piezoresistors are placed out of harm's way.

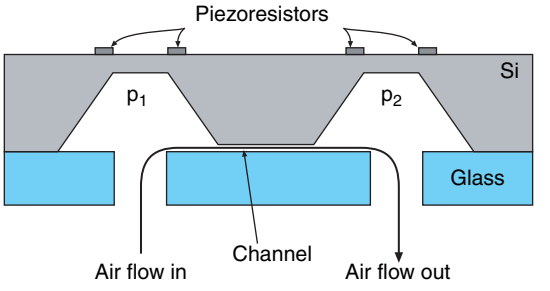


Figure 36 Cross-section of differential-pressure flow sensor based on two pressure measurements relative to atmospheric pressure. (Source: Boillat M, van der Wiel A, Hoogerwerf A, de Rooij N 1995 A differential pressure liquid flow sensor for flow regulation and dosing systems. *Proceedings of the Micro Electro Mechanical Systems, 1995, MEMS '95*. IEEE, pp. 350. © 1995 IEEE.)

The sensor presented by Takahata *et al.* (2004) consists of two capacitive absolute pressure-sensing chips connected to each other with a coil. The coil, the so-called stentenna, is created from a flat structure, fabricated using microelectrodischarge machining. The application is blood-flow measurement, where the stentenna, with a pressure-sensing chip at each end, is inflated in the blood vessel using an angioplasty balloon. The inductive coil, split into two, and the two capacitive sensors are coupled as indicated in **Figure 37**, forming two LC tanks. As the capacitors vary relative to each other due to flow-induced pressure difference, the impedance variation of the circuit can be measured using an external coil. The main advantage of this system is that it contains no active elements and can be powered from outside the body. The device was limited by its quality factor and the low coupling between the internal and the external coil. Continuing on the concept, DeHennis and Wise (2006) from the same group developed a fully integrated differential-pressure sensor with integrated BiCMOS circuitry. The post-BiCMOS process steps include a transfer bond of the circuitry to a glass substrate and a thinning down of both the glass and the silicon substrates to obtain a total device thickness of less than 200 μm. Two epoxy-glued surface-mount capacitors are the only nonintegrated sensor parts. The gold antenna coil is postprocessed on top of the BiCMOS and thinned down to a membrane during the last silicon-thinning step. It is compliant to be able to wrap around the surface of an artery or a stent and not interfere with the blood flow. The circuitry is powered using an RF signal and transmits information back by amplitude-modulating this signal, also known as backscatter modulation. The capacitances, placed at 1 cm distance from each other, are read out using capacitance-to-frequency conversion.

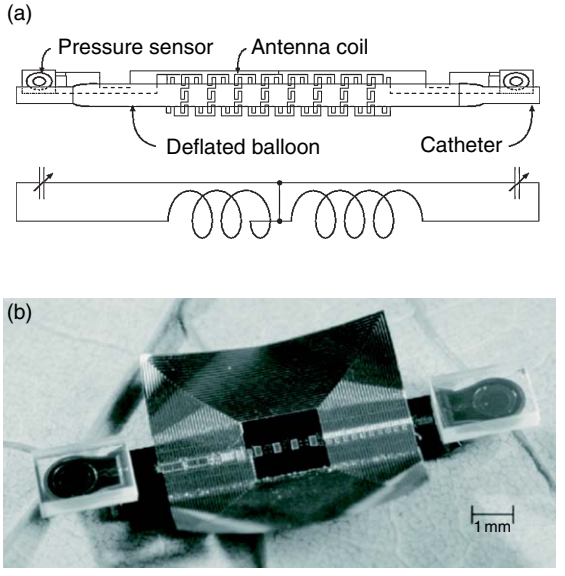


Figure 37 (a) Circuit of the double LC circuit where the variable capacitors are absolute pressure sensors (Source: Takahata K, DeHennis A, Wise K, Gianchandani Y 2004 A wireless microsensor for monitoring flow and pressure in a blood vessel utilizing a dual-inductor antenna stent and two pressure sensors. *17th IEEE Int. Conf. Micro Electro Mechanical Systems,* Maastricht, The Netherlands, pp. 216–19. © 2004 IEEE.) (b) Photograph of a differential-pressure sensor with integrated BiCMOS circuitry and flexible antenna. (Source: DeHennis A, Wise K 2006. A fully integrated multisite pressure sensor for wireless arterial flow characterization. *J. Microelectromech. Syst.* **15(3)**, 678–85. © 2006 IEEE.)

In general, pressure-based flow sensors have the advantage of not needing to heat up the fluid, which enables them to be used in applications that involve fluids that decompose or ignite at elevated temperatures. Also, using capacitive readout, they are able to measure very small flows (Cho *et al.* 1992), whereas thermal-based flow sensors are not practical in very low-flow applications because the gas velocity in those applications is too low to generate any significant convective cooling (Cho *et al.* 1992, Jang and Wereley 2004). A disadvantage of most pressure-based flow sensors is the generated pressure loss, which limits their use to applications that can tolerate this.

2.07.5 Optical Flow Sensors

Micromachined flow sensors that directly measure in the optical domain are few. In order to obtain a signal from the flow, the signal needs to be reflected off an interface that moves along with the flow. Hence, direct optical flow sensors can be used only with multiphase flows.

Flows containing or seeded with many small particles, such as bubbles or beads, can be measured with laser Doppler anemometry (LDA) or PIV. Review articles on these methods can be found in Briers (2001) and Adrian (2005), respectively. The velocity of flows in channels that contain large-area interfaces such as slug flow can also be measured optically by means of the refractive index of the liquid at the edge of the channel (Kraus *et al.* 2004). These methods have all been applied to flows in microchannels, most often with conventional optics. A particular class of optical flow sensors are the cytometers, which count the rate of particles or biological cells

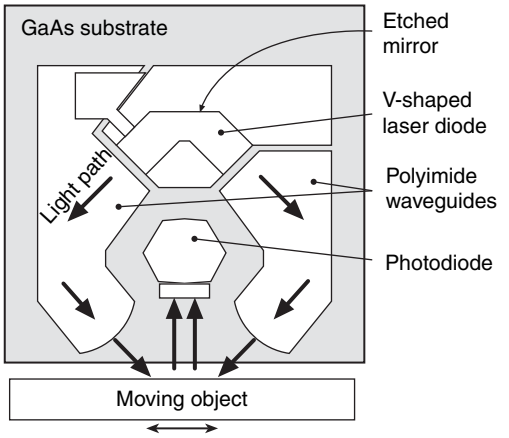


Figure 38 Monolithic laser Doppler anemometry (LDA) chip for near-surface velocity sensing. (Source: Ito T, Sawada R, Higurashi E 1999 Integrated microlaser Doppler velocimeter. *J. Lightwave Technol.* **17(1)**, 30–4. © 1999 IEEE.)

passing the sensor. Although they could be used as flow sensors, they will not be discussed here.

A fully integrated, micromachined laser Doppler anemometer has been developed by Higurashi *et al.* (2003) and Ito *et al.* (1999). The complete LDA system consists of a V-shaped laser diode, fluorinated polyimide waveguides, and a photodiode and was fabricated monolithically on a GaAs substrate (**Figure 38**). The coherent light coming out of the two waveguides reflects off particles moving close to the surface of the sensor, and the difference in Doppler shift can be observed by the photodiode. This sensor provides a nonintrusive measure of the flow velocity of red blood cells just below the skin surface (see also Section 2.07.11.2).

A second application of micromachined LDA was designed by Fourguette *et al.* (2001, 2003). Described in detail in Section 2.07.11, these sensors provide a signal that is linearly dependent on the flow velocity divided by the distance from the wall. In general, these optical methods have a linear response and high-frequency response. Aside from the seed particles, if used, the optical sensors are also nonintrusive.

2.07.6 Ultrasonic Flow Sensors

Ultrasonic transducers can be used in several ways to act as flow sensors (Vass 1997, Hauptman *et al.* 2002). Ultrasonic transit time flowmeters (**Figure 39(a)**) measure the flow-induced time-of-flight difference

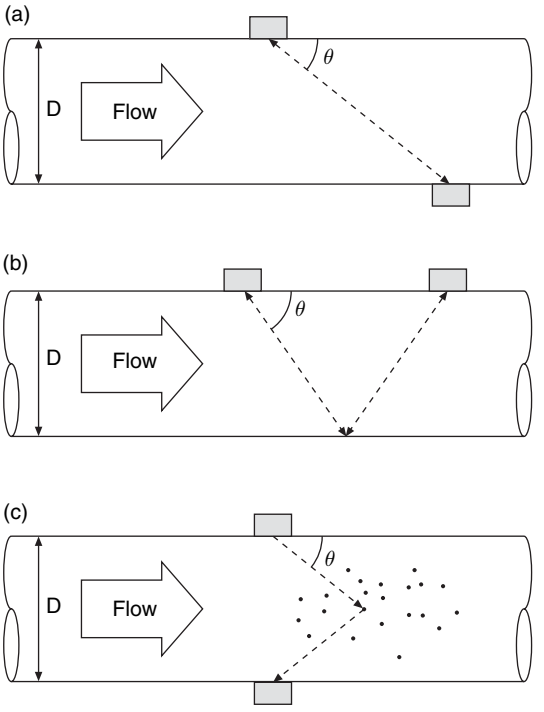


Figure 39 Ultrasonic measurement modes: (a) transit-time flowmeter and (b) Doppler flowmeter.

between an ultrasonic signal traveling upstream and one traveling downstream. This time difference is given as follows:

$$\Delta t = \frac{2D\cot\theta}{c^2} U_{\text{avg}}$$

where D is the pipe diameter, θ the angle of the signal path, c the speed of sound in the medium, and U_{avg} the average flow velocity along the signal path (Bentley 1983b). To increase the path length, detectors can be placed such that the signal is reflected a certain number of times off the pipe wall before it reaches the detector. In multiphase flows (see **Figure 39(b)**), where air bubbles or particles can function as sonic reflectors, the Doppler effect induces a frequency shift in the reflected wave. This shift is as follows (Bentley 1983b):

$$\Delta f = \frac{2f}{c}(\cos\theta) U_{\text{avg}}$$

where f is the frequency of the ultrasonic wave. An alternative to the Doppler effect in multiphase flows is to use the cross-correlation of random attenuation fluctuations measured by two pairs of transducers. The time shift between these fluctuations is directly dependent on the flow velocity.

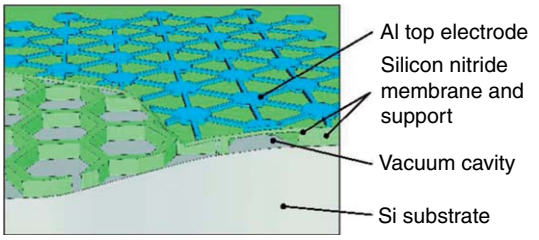


Figure 40 Partial cross-section of an ultrasonic transducer based on a mesh of membranes. (Source: Cittadine 2000.)

The principles employed by conventional ultrasonic transducers can directly be miniaturized by using MEMS technology. Piezoelectric and magnetostrictive transducers (Quandt 1998) can be fabricated using thin-film deposition technology, while sacrificial etching enables producing cavities for capacitive transducers (Cittadine 2000, Jin *et al.* 1998, Ladabaum *et al.* 1998). In air, capacitive ultrasonic MEMS sensors have an acoustic impedance that matches the air significantly better than piezoelectric sensors do; an example is shown in **Figure 40** (Cittadine 2000).

SAW sensors can detect convective cooling as a change in the frequency of a SAW delay line oscillator. They have been described in more detail in Section 2.07.2.

2.07.7 Coriolis Flow Sensors

The principle of Coriolis flow sensors is the apparent force that a fluid experiences when it is moving in a rotating frame of reference. This force is directly proportional to the rotation rate of the frame (ω) and the mass flow rate of the fluid ($\partial m/\partial t$):

$$F_c = 2\frac{\partial m}{\partial t}\omega$$

This relationship implies that, by measuring the effect of the force, Coriolis flow sensors directly measure mass flow, independent of the type of fluid and of whether or not it is single phase. In addition, provided the spring is linear, Coriolis mass flow sensors have a linear response. All electrical connections are outside the flow. One disadvantage is that because the fluid is forced through the tubing, the measurement tubing has to be of the same size as the measured tube to prevent a pressure drop over the instrument.

The practical implementation of Coriolis mass flow sensors consists of bringing the tubing into a vibratory resonance mode that imposes a varying rotation onto the liquid. This is called the primary mode. The Coriolis force then excites another mode, preferably also a resonance mode, called the secondary mode. The amplitude of this mode is a measure for the flow. An important factor when designing Coriolis flow sensors is the trade-off between the relative positions of these two resonance frequencies. On the one hand, one wants the resonance frequencies of the primary and secondary modes to lie close to each other so that the Coriolis force excitation is amplified near the secondary resonance frequency. On the other hand, if the modes lie too close to each other, one risks exciting both modes by asymmetries due to structural imperfections. Although the parasitic secondary mode excitation would still lie at a 90° phase shift with the Coriolis-based excitation, this still generates an increased noise level.

Only a very few Coriolis mass flow sensors have been made using MEMS technology. The first MEMS Coriolis mass flow sensor was presented by Enoksson *et al.* (1997). It is bulk-micromachined, and KOH etching and silicon fusion bonding are used to create a double loop of hexagonal tubes. In one operation mode, illustrated in **Figure 41**, electrostatic actuation is used to drive the sensor into a primary bending mode resonance to achieve a maximum deflection. The secondary torsional mode that is activated by the Coriolis force is read out by detecting the angle of reflection of a laser beam

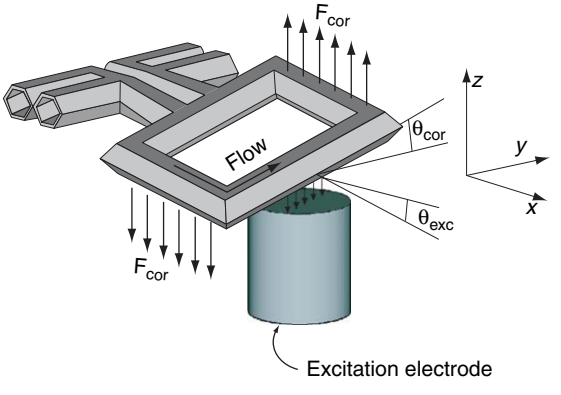


Figure 41 Half of a double-loop device, operated in bending mode, using an external electrode, generating a twisting mode due to mass flow through the tube. (Source: Enoksson P, Stemme G, Stemme E 1997. A silicon resonant sensor structure for Coriolis mass-flow measurements. *J. Microelectromech. Syst.* **6(2)**, 119–25. © 1997 IEEE.)

pointed at the loop. The diameter of the tubes, the distance between the top and the bottom wall of the hexagonal cross-section, is 800 µm. It has a measurement range of $\pm 0.5\,\mathrm{g\,s^{-1}}$ ($1800\,\mathrm{g\,h^{-1}}$) with a relative precision of 0.8%. The torsional mode can also be used as the primary mode, resulting in a secondary bending mode. In the presented sensor, in-phase primary modes result in out-of-phase secondary modes and vice versa (see **Figure 42**).

Another group at Integrated Sensing Systems (ISSYS), Inc., Ypsilanti, MI, USA, introduced a single-loop Coriolis mass flow sensor with surface-micromachined tubes transferred to a support wafer (Tadigadapa *et al.* 2002, Zhang *et al.* 2001) (see **Figure 43**). This process allows more freedom in the tube cross-section design, and cross-sections of 10 and 100 µm have been presented. Both the actuation and the readout are capacitive. The actuating mode is bending and the sensing mode is torsional.

The Industrial Technology Research Institute, Hsinchu Hsien, Taiwan, filed a patent (Fan *et al.* 2002a) describing a readout method for micromachined Coriolis mass flow sensors where the substrate and a mirror addition to the moving part of the loop compose a Fabry–Perot interferometer. This allows miniaturized optical readout, which was not possible with the original sensor presented by Enoksson and coworkers.

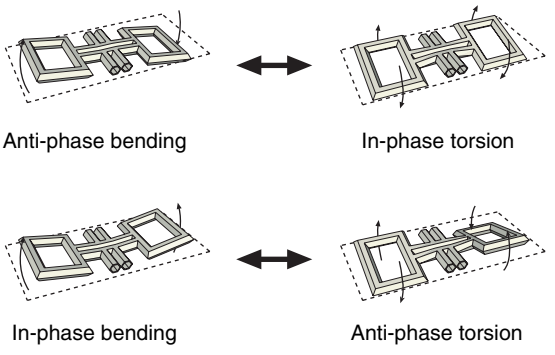


Figure 42 The four resonance modes of the Coriolis sensor presented by Enoksson *et al.* (1997) (Source: Enoksson P, Stemme G, Stemme E 1997 A silicon resonant sensor structure for Coriolis mass-flow measurements. *J. Microelectromech. Syst.* **6(2)**, 119–25. © 1997 IEEE.) Each mode can be used both as a primary as well as a secondary mode.

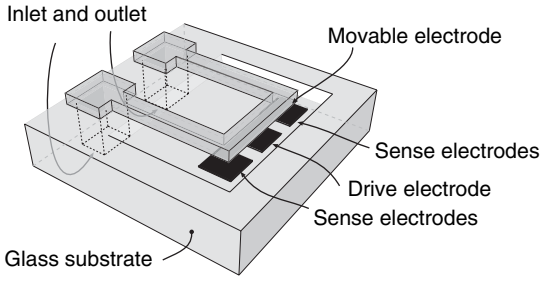


Figure 43 Single-loop Coriolis mass flow sensor with capacitive readout. (Source: Zhang *et al.* 2001.)

An important property of the Coriolis mass flow sensors is that they all can be used as density sensors, since the resonance frequency depends on the mass of the fluid in the tubes:

$$f_0 = 2\pi\sqrt{k_t/m_t + \rho V_t}$$

where k_t is the spring constant of the tube in the resonance mode, m_t the tube wall mass, V_t the volume of the fluid in the tube, and ρ the density of the fluid.

2.07.8 Direct Electrical Flowmeters

2.07.8.1 Electrohydrodynamic

The electrohydrodynamic (EHD) flowmeter is based on measuring the transition behavior of the current when a voltage is applied between two electrodes in a liquid flow. The transient current shows one or more cusps, and the time at which these occur can be correlated with the flow velocity. The dependence is determined by the type of charge carriers in the fluid: for deionized (DI) water at increasing flow rates, an increasing cusp time was noted, while the cusp time decreased for measurements in ethanol. The amount of cusps is defined by how many different types of charge carriers in the fluid are activated. To obtain the ideal situation where only one type of carriers is emitted, the voltage must be adjusted to the fluid.

Richter *et al.* (1991) constructed a micromachined EHD flowmeter by placing two grid electrodes across a channel at distances ranging from 10 to 60 µm from each other (**Figure 44**). The structure is identical to an EHD pump, and since the voltages used for generating the current are the same as those used for EHD pumps, precautions must be taken to avoid the generation of a flow due to the measurement. One solution presented in Richter *et al.* (1991) suggests the use of three electrodes, where the center electrode acts as an emitter, creating an equal in-flow and counterflow pressure.

For an orifice size of $6.25\,\mathrm{mm^2}$ and a voltage of 300 V, flow rates up to $1800\,\mathrm{\mu l\,min^{-1}}$ ($4.8\,\mathrm{mm\,s^{-1}}$) for DI water and $50\,\mathrm{\mu l\,min^{-1}}$ ($0.13\,\mathrm{mm\,s^{-1}}$) for ethanol

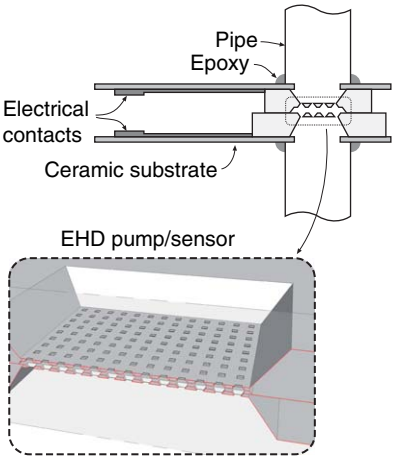


Figure 44 Schematic drawing of a packaged electrohydrodynamic (EHD) pump/flowmeter. (Source: Richter A, Plettner A, Hofmann K, Sandmaier H 1991 Electrohydrodynamic pumping and flow measurement. *Proceedings. IEEE Micro Electro Mechanical Systems. An Investigation of Micro Structures, Sensors, Actuators, Machines and Robots*, pp. 271–6. © 1991 IEEE.)

were measured. The measurements were taken at 1 Hz; the cusp times varied from 50 to 390 µs. These numbers are dependent on the area of the orifice and the distance between the grids.

2.07.8.2 Electrochemical

The time-of-flight principle used for thermal sensors can be straightforwardly applied to the electrochemical domain. Instead of a heater and a temperature sensor, a molecule or an ion generator and a sensor are used.

Wu and Sansen (2002) used the oxygen generated by electrolysis at the anode of an electrochemical cell, and measured the oxygen level with an identical cell 1.5 cm downstream (**Figure 45**). The cells consist of three thin-film electrodes: Pt work and counter electrodes and a Ag reference electrode. The sensing cell can be run in two modes: as an amperometric sensor or as an impedance sensor. In the latter case, only the work and counter electrodes are used.

The use of the sensing electrode in amperometric mode was possible when measuring the flow of phosphate-buffered saline (PBS, a commonly used buffer in biochemistry), where the Cl^- concentration is constant, forming a pseudo Ag/AgCl reference electrode. The amperometric mode can be used even under addition of limited amounts of electroactive species to the buffer. For liquids where the Cl^- concentration is not constant (e.g., tap water), the sensing

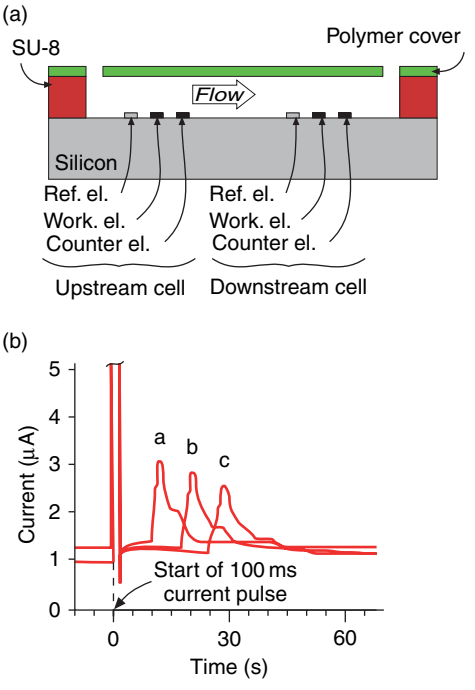


Figure 45 (a) Cross-section of the electrochemical time-of-flight flow sensor with integrated channel. The polymer cover is self-adhesive. (Source: Wu J, Sansen W 2002 Electrochemical time of flight flow sensor. *Sens. Actuators A (Phys.)* **A97–98**, 68–74. © 2002 Elsevier.) (b) Sketch of the current response from the oxygen sensor in amperometric mode. a, b, and c are peaks at 10, 6 and 4 µl min-1 in phosphate-buffered saline (PBS) buffer. The secondary peak on the right flank of each curve is due to oxygen diffusion in the direction orthogonal to the flow. (Source: Wu J, Sansen W 2002 Electrochemical time of flight flow sensor. *Sens. Actuators A (Phys.)* **A97–98**, 68–74. © 2002 Elsevier.)

cell can be used in impedance-sensing mode. A pulse of increased impedance can be sensed, which is speculated to be caused by oxygen microbubbles that are formed at the upstream cell.

The sensor is capable of measuring flow rates ranging from 1 to 15 µl min^{-1}. The range limits for electrochemical flow sensors are imposed by effects similar to those limiting thermal time-of-flight sensors. At flow rates below 1 µl min^{-1}, diffusion becomes too large a factor in comparison to convection, and the oxygen peak cannot be resolved. At flow rates above 15 µl min^{-1}, the current peak measured by the oxygen sensor lies too close in time to the current peak that is caused by the generating cell. This peak is picked up by the sensing cell due to the coupling in the conductive fluid (indicated by A in **Figure 45(b)**).

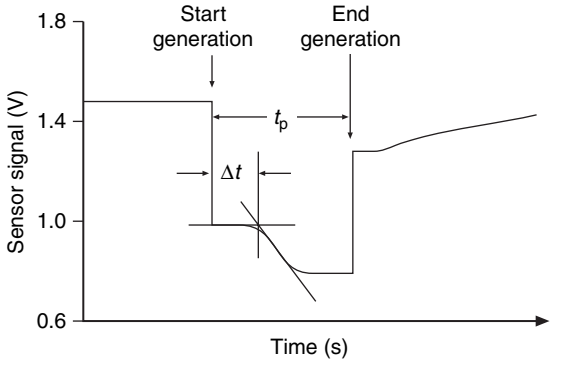


Figure 46 Typical output signal of an ion-sensitive field emission transistor (ISFET)-based flow sensor measuring the generation of OH^- ions with a constant current (CC) of 30 μA. The pulse length t_p lies between 10 and 30 s. Δt is the transit time needed by the OH^- ions to reach the ISFET gate region. (Source: Poghossian A, Berndsen L, Schoening M 2003a Chemical sensor as physical sensor: ISFET-based flow-velocity, flow-direction and diffusion-coefficient sensor. *Sens. Actuators B (Chem.)* **B95(1–3)**, 384–90. © 2003 Elsevier.)

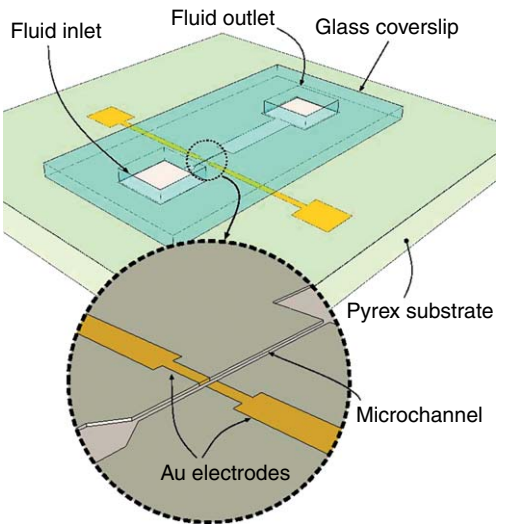


Figure 47 Impedance-based flow sensor. (Source: Ayliffe *et al*. 1999.)

Poghossian *et al.* (2003a) and Schoening *et al.* (2001) employ the H^+ and OH^- groups, respectively, generated during electrolysis of water as tracers to be detected by an ion-sensitive field emission transistor (ISFET). The ISFET used is a field emission transistor where Ta_2O_5 on the gate makes is sensitive to pH levels. The sensor is made direction sensitive by placing an ISFET on either side of the ion generator. Unbuffered 0.1 M KNO_3 water solution (pH ~ 7) was used to measure upon. Flow rates ranging from 0.1 to 0.6 ml min^{-1} were measured in a channel of 2 mm wide and 0.5 mm height. **Figure 46** shows a typical output signal of a sensor of this type. An advantage of using this technique is that it can be combined with other uses of ISFETs, such as sensing pH, penicillin concentration, diffusion coefficient, and temperature (Poghossian *et al.* 2003b).

Besides its use as a time-of-flight sensor, a correlation has been noted between the relaxation time and the flow rate, suggesting a possible application in flow sensing in a way analogous to calorimetric flow sensing.

2.07.8.3 Electrical Impedance

Besides measuring the time of flight of the electrical impedance effect of an electrochemical pulse discussed above (Wu and Sansen 2002), the direct measurement of the electrical impedance change of a liquid due to flow has also been reported; Ayliffe and Rabbitt (2003) describe an electrical impedance sensor integrated in microfluidic channels for bioMEMS applications. The channels are formed by a glass/SU-8/glass sandwich, while electroplated gold electrodes are integrated in the walls (**Figure 47**). A dependence of both the magnitude and the phase of the impedance in a physiologic saline solution is shown.

Observing the impedance using a network analyzer, a magnitude dependence with respect to the flow could be observed at frequencies below 200 kHz with a maximum effect at 350 Hz. The phase profile of the impedance also changed. At frequencies above 200 kHz, no significant flow dependence was observed. The magnitude increase could be reproduced in an electrochemical/convective simulation and was explained by the flow washing out the ion cloud that is formed between the electrodes. On applying a DC bias (50 mV) during the AC interrogation of the impedance, this ion cloud is larger, which increases the flow sensitivity. Flows ranging from 2.4 to 4.8 μl min^{-1} were measured.

Collins and Lee (2004) measured the admittance across a channel to determine the flow rate of a saline solution. With an electrode distance of 100 μm (**Figure 48**), as opposed to 5–10 μm in the set-up of Ayliffe *et al.* (1999), the capacitance is much smaller. Within the range of 10 Hz–5 kHz, an optimal measurement frequency, where the absolute value of the admittance was maximized, was found to lie at around 500 Hz. Besides the frequency dependence,

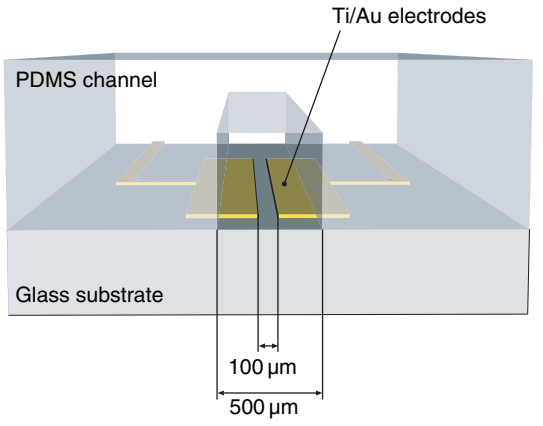


Figure 48 Admittance-based flow sensor integrated in polydimethylsiloxane (PDMS) microchannel. (Source: Collins and Lee 2004.)

an optimal excitation voltage (0.4 V) and a buffer concentration (0.2 M) were observed. Using an AC excitation voltage of 400 mV and measuring the current through a 1-kΩ series resistor, sensitivities ranged from 10^{-4} to 5.2×10^{-4} mA μl^{-1} min depending on the electrolyte or the buffer used. Flows from 35 down to 0.1 μl min^{-1} could be measured. Only the root mean square (RMS) value of the current was taken into account, and no information was given on the observed phase shift.

2.07.8.4 Electromagnetic – Using Faraday's Law

According to Faraday's law, when a conductor moves through a magnetic field, a voltage that is dependent on the relative velocity between the conductor and the field is produced in the conductor. Therefore, a conducting fluid, when flowing through a magnetic field oriented perpendicular to the flow, will induce a flow rate-dependent voltage in the direction perpendicular to the flow and the field (**Figure 49**). For channels with a circular cross section, the voltage between two electrodes placed opposite to each other in the channel wall can be expressed as $E = DBV$, where D is the diameter of the channel, B the magnetic field strength, and V the average flow rate (Thürlemann 1941, 1955).

Yoon *et al.* (2000) presented a microelectromagnetic flow sensor based on anisotropically etched silicon microchannels with metal electrodes (**Figure 50**). Flow rate measurements on a fluid with a conductivity of 100–200 mS cm^{-2} in the range of 9.1–62 ml min^{-1}

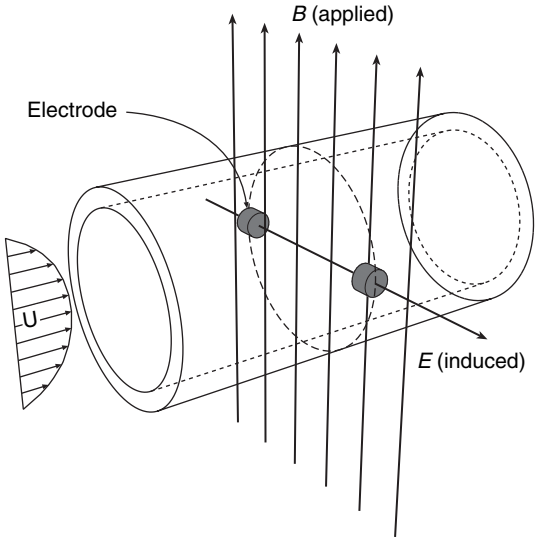


Figure 49 Schematic representation of the electromagnetic flowmeter.

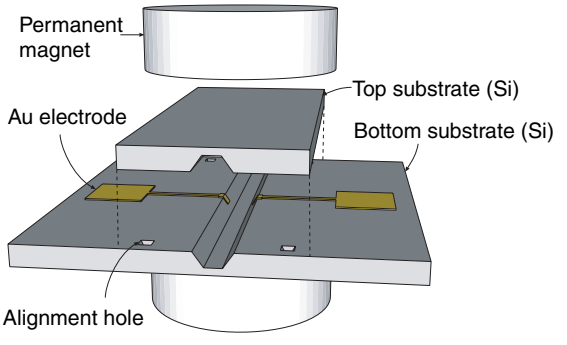


Figure 50 Exploded view of an electromagnetic flowmeter integrated in an anisotropically etched microchannel. (Source: Yoon *et al*. 2000.)

in a magnetic field with flux density of 250 mT resulted in induced voltages of 261 μV to 7.3 mV, respectively. Using microelectric discharge machining, Takahata and Gianchandani (2005) produced a sensor that can be implanted in blood vessels using balloon angioplasty. The implanted sensor consists of three parts: an antenna coil for excitation, a varactor diode to create on oscillator circuit, and a set of cuff electrodes (**Figure 51(a)**). The coil and the electrodes are fabricated in a planar form and threaded on the balloon. The expansion of the balloon plastically deforms the structures; this keeps them in place after the balloon is deflated and removed. The circuit is illustrated in **Figure 51(b)**. An external coil is used to interrogate the system. When a magnetic field is

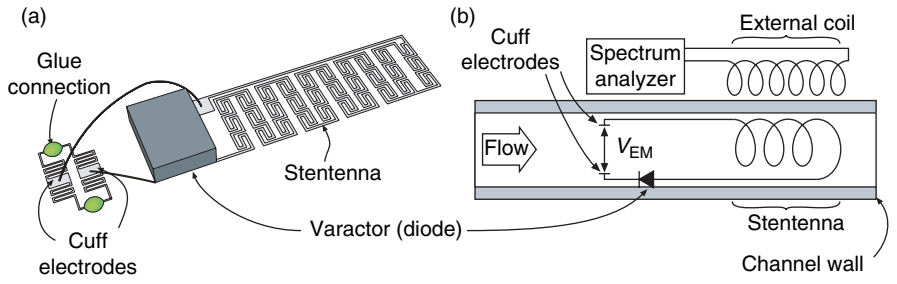


Figure 51 Electromagnetic flowmeter for implantation in blood vessels. (a) Pre-expansion and (b) equivalent circuit. (Source: Takahata K, Gianchandani Y 2005 Micromachined intraluminal devices for active and passive electromagnetic measurements of flow. *18th IEEE Int. Conf. Micro Electro Mechanical Systems 2005 (MEMS 2005)*, pp. 863–6. © 2005 IEEE.)

applied, the flow-dependent DC voltage generated in the cuff electrodes controls the varactor's capacitance. This affects the resonance frequency and the phase response of the system. Flow measurements were simulated but had not yet been performed.

The advantages of electromagnetic flow sensors are that they have a simple structure and generate no heat or pressure loss. For applications where the total flow through the channel is of interest, electromagnetic flow sensors provide a correct value, independent of the flow profile. The limitation is that the liquid needs to be conductive; the level of conductivity, however, does not affect the measurement.

A variant where the electrodes are not in ohmic contact with the fluid has been fabricated by Merkel *et al.* (2000) using microchannels in printed circuit boards (Merkel *et al.* 1999) (**Figure 52**). By varying the magnetic field, the voltage can be detected capacitively (see **Figure 52**). Flow rates ranging from 2 to 14 $\mu l\,min^{-1}$ were measured using a magnetic field intensity with an amplitude of 1200 $A\,m^{-1}$ varying at 65 kHz.

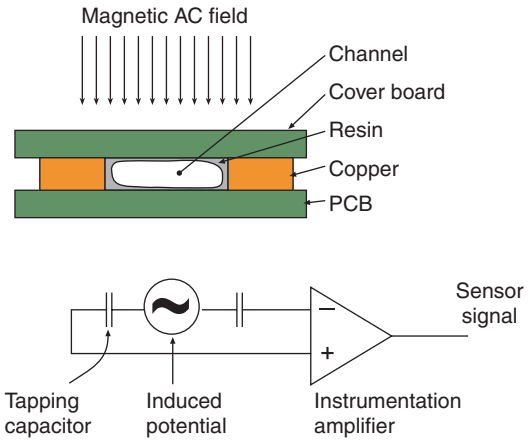


Figure 52 Cross-section and equivalent electrical circuit of the printed-circuit-board electromagnetic flow sensor. (Source: Merkel T, Pagel L, Glock H-W 2000 Electric fields in fluidic channels and sensor applications with capacitance. *Sens. Actuators A (Phys.)* **A80(1)**, 1–7. © 2000 Elsevier.)

2.07.9 CNT-based Flow Sensors

First discovered by Sumio Iijima in 1991 (Iijima 1991), CNTs are cylindrical carbon molecules with interesting thermal, mechanical, and electrical properties. CNTs can be seen as wrapped graphite sheets that, depending on the orientation, behave like semiconductors or like metals. A distinction is made between single-walled carbon nanotubes (SWCNTs) and MWCNTs, the latter consisting of several concentric cylinders. With the introduction of fabrication techniques that allow CNTs to be integrated with thin-film processes, more properties of CNTs have been discovered and different application areas have arisen, among which is flow sensing.

A truly novel phenomenon was observed by Ghosh *et al.* (2003). The voltage across bundles of randomly oriented SWCNTs exposed to liquid flows (**Figure 53**) was found to be nearly logarithmically dependent on the flow rate over six orders of magnitude (for water: from 10^{-7} to $10^{-1}\,m\,s^{-1}$ free flow velocity in a 3-cm-diameter glass tube). The sensitivity was observed to be dependent on the type of liquid, and it is believed to be dependent on the adsorption properties of the nanotube surfaces with respect to different types of ions (Sood *et al.* 2005). The underlying physics of the phenomenon is not yet fully understood, and several explanations have been suggested. In a review paper, Sood *et al.* (2005) argue for the explanation

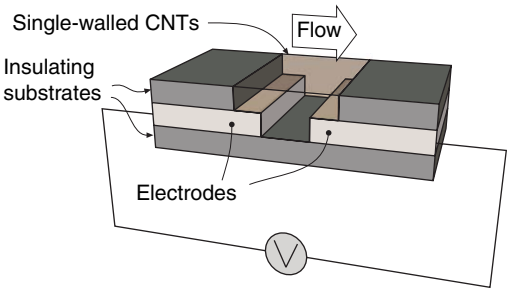


Figure 53 Schematic sketch of the nanotube flow sensor placed along the flow direction. (Source: Sood *et al*. 2005.)

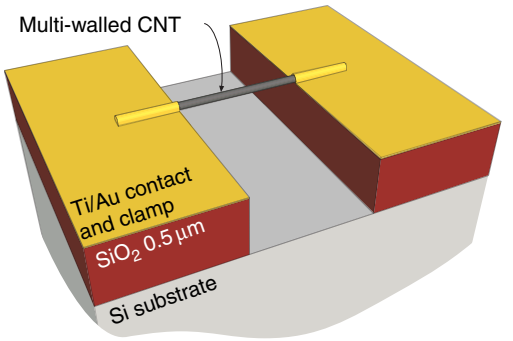


Figure 54 A suspended carbon nanotube (CNT) device. (Source: Chung *et al*. 2003.)

given in Ghosh *et al.* (2003, 2004), rejecting theories proposed by Král and Shapiro (2001) and Persson *et al.* (2004).

The response time of the sensor was reported to be better than a few milliseconds, the measurements being limited by the liquid flow velocity and the rate of data acquisition (Sood *et al.* 2005)

Chung *et al.* (2003) reported a resistance change in a suspended MWCNT subjected to a nitrogen flow (**Figure 54**). The effect is assumed to be a result of drag-induced deformation resulting in a resistance change (Tombler *et al.* 2000). Sood and Ghosh (2004) also presented a CNT gas flow sensor that uses the Seebeck effect in CNTs. This principle is described in Section 2.07.4.

2.07.10 On-chip Electronics

The use of semiconductor substrates for micromachined flow sensors provides the possibility of integrating electronics on the same chip as the sensor. This can reduce fabrication costs as fewer chips need to be combined in a package. The established fabrication facilities for complementary metal oxide semiconductor (CMOS) electronics can also provide low cost and high volume production. From a technical point of view, the output signal can be amplified or digitized before it leaves the chip thereby reducing the effects of noise. However, there are challenges both in fabrication and in operation. The overall majority of flow sensors with on-chip electronics are thermal and therefore require above-ambient, often-varying temperatures or temperature gradients that can affect the behavior of the analog electronics around them. Bruschi *et al.* (2004a), Makinwa and Huijsing (2002b), and Tong and Huang (1987) solved this problem by maintaining the entire chip at a constant, moderately elevated temperature (up to 55°C). Fricke (1994) used GaAs as a substrate material since the MESFET circuit integrated upon it can stand temperatures up to 240°C without its behavior being significantly affected. A more common solution, however, is to thermally isolate the heater from the rest of the chip. To achieve this, usually membranes need to be formed or grooves etched, and the etching methods used for this are often not compatible with, for example, CMOS electronics. One approach concentrates the thermal isolation micromachining steps in the pre- and post-processing stages before and after the CMOS fabrication (e.g., Malcovati *et al.* 1995). Another, more comprehensive, method integrates the MEMS-specific process steps in the CMOS fabrication (e.g., Yoon and Wise 1992).

Thermal flow sensors with on-chip circuitry can be categorized into two groups based on the function of the circuitry. The sensors of the first group are those in which the electronics amplify or digitalize the output signal. Bosman *et al.* (1992) used an on-chip voltage-to-frequency converter to convert the low Seebeck voltage signals, on the order of 0.1 mV, to a more noise-resistant variable-frequency signal. Bruschi *et al.* (2004a) based the function of their circuit on a current source that is proportional to absolute temperature (PTAT). Developed by Meijer (1986), this current source provides a signal linearly dependent on the absolute temperature of the device. This signal is used by Bruschi and coworkers as an input to an off-chip CT controller. Kersjes *et al.* (1993) placed a heater/sensor diode on a membrane to thermally isolate it from the bulk and used a differential amplifier at ambient temperature on the bulk of the chip to measure the temperature difference between the diode on the

membrane and the one on the bulk. A second chip then converted the analog signal into a pulse-width-modulated signal. One step beyond amplification is to digitalize the signal on the chip. Verhoeven and Huijsing (1994) used a PTAT current source and combined it with an on-chip sigma–delta converter to provide a TTL-compatible pulse-rate-modulated signal with a 10-bit resolution. Malcovati *et al.* (1995) based their design on a CMOS sigma–delta converter resulting in a 15-bit resolution.

In the second group are those sensors where the on-chip circuitry provides a feedback signal to operate the sensor in closed loop. The electrothermal oscillator by Stemme (1988), described above, was implemented at the bulk end of the sensor using metal gate CMOS technology with only the heating circuitry at the thermally isolated end. Chen *et al.* (2005) fabricated out-of-plane hot-wire structures starting from a passivated CMOS chip containing a constant temperature anemometry (CTA) circuit with only the contact pads free. (The process is described in Section 2.07.2.3.4.) A fully standard CMOS-processed chip was developed by Tong and Huang (1987). The chip is kept at a CT, and the voltage to the heating resistor is the output signal. This simple process skips any micromachining steps, but the trade-off is a time constant of around 40 s. Makinwa and Huijsing (2002b) also used a standard CMOS chip to make a hot-chip sensor for wind-meter applications. The feedback circuitry is based on three sigma–delta modulators, one to keep the chip at a constant average temperature and two to balance out the flow-induced temperature differences over the chip. This device has a time constant on the order of 10 seconds. The wide-range flow sensor developed by Yoon and Wise (1992) consists of two flow-rate sensors and a flow direction sensor. **Figure 55** shows a block diagram and a chip layout of the multisensor chip that, besides the flow-rate and direction sensors, has a gas-type sensor and a pressure sensor. The flow-rate sensors are driven in CT mode. To build the device, a dedicated process was developed that interleaved the CMOS and the micromachining steps. An analog multiplexer is used to select the desired output.

The demands on the fabrication of on-chip electronics in the differential-pressure blood-flow sensor presented by DeHennis and Wise (2006) are similar to the thermal flow sensors, but for different reasons. No heating is required, but the formation of the pressure membranes by wet etching and, specific for this sensor, the thinning of the substrate require pre- and postprocessing of the BiCMOS electronics used.

In the field of mechanical flow sensors, Schmidt *et al.* (1987) used a matched pair of depletion-mode metal gate positive-channel MOS (PMOS) transistors as the first stage to read out the capacitance variation of the polyimide shear-stress sensor described above (see **Figure 30(b)**). The signal is amplified further off-chip. On-chip feedback has also been demonstrated for mechanical skin-friction sensors. Pan *et al.* (1995a, b) (see **Figure 30(f)**) used

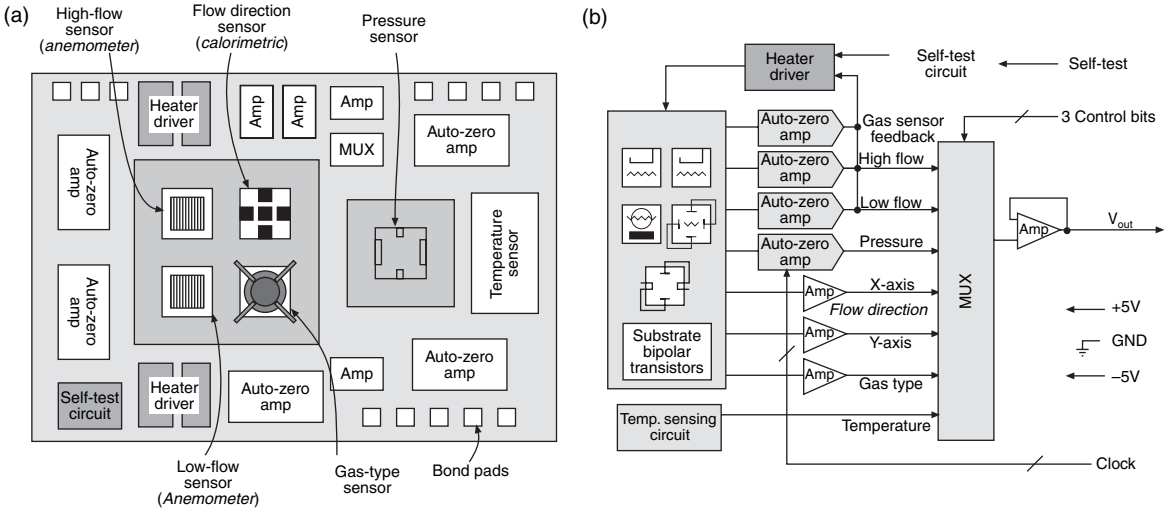


Figure 55 (a) Block diagram and (b) chip layout of the multisensor chip presented by Yoon and Wise (1992). (Source: Yoon E, Wise K 1992 An integrated mass flow sensor with on-chip CMOS interface circuitry. *IEEE Trans. Electron. Devices* **39(6)**, 1376–86. © 1992 IEEE.)

Table 6 Some micromachined flow sensors with on-chip electronics

Reference	*Type*	*Technology*	*On-chip feedback*	*Output*
Tong and Huang (1987)	Thermal	CMOS only	Yes	Amplified voltage
Schmidt *et al.* (1987)	Mechanical	PMOS + postprocessing	No	Amplified voltage
Stemme (1988)	Thermal	CMOS + postprocessing	Yes	PWM
Yoon and Wise (1992)	Thermal	Adapted CMOS	Yes	Multiplexed, amplified voltage
Bosman *et al.* (1992)	Thermal	BJT	No	Voltage-dependent frequency
Kersjes *et al.* (1993)	Thermal	Preprocessing + CMOS + postprocessing	No	Amplified voltage
Verhoeven and Huijsing (1994)	Thermal	Adapted CMOS (DIMES-01; Nanver *et al.* 1993)	No	$\Sigma\Delta$ bitstream
Fricke (1994)	Thermal	GaAs MESFET	No	Amplified voltage
Pan *et al.* (1995a)	Mechanical	BiCMOS + postprocessing	Yes	Amplified and modulated voltage
Malcovati *et al.* (1995)	Thermal	CMOS + postprocessing	No	$\Sigma\Delta$ bitstream
Makinwa and Huijsing (2002b)	Thermal	CMOS only	Yes	$\Sigma\Delta$ bitstream
Bruschi *et al.* (2004a)	Thermal	CMOS only	No	*Amplified voltage*
Chen *et al.* (2005)	Thermal	CMOS + postprocessing	Yes	Amplified voltage
DeHennis and Wise (2006)	Differential pressure	Preprocessing + BiCMOS + postprocessing	N/A	Amplitude-modulated RF signal

the control circuitry of an analog devices accelerometer (Kuehnel and Sherman 1994) to maintain the floating element at a constant position. Analog to CTA, the feedback signal provides a measure for the skin friction. **Table 6** provides an overview of flow sensors with integrated circuitry.

2.07.11 Applications

2.07.11.1 Skin-friction Measurements

Skin friction, also referred to as wall shear stress, is an important factor in the study of near-wall fluid flows. From a scientific point of view, measuring the skin friction provides information for understanding and modeling the complex behavior of the turbulent boundary layer. From an engineering point of view, this turbulent boundary layer determines the performance of aerodynamic surfaces such as wings or propellers. In flow-control applications, where one needs to control the transition from laminar to turbulent flow or vice versa, feedback from high-resolution skin-friction measurements is essential. The small size and rapid response times of MEMS sensors make them particularly interesting for such applications.

Besides a high bandwidth of up to 10 kHz for high Reynolds numbers and a spatial resolution of 100 μm (Padmanabhan *et al.* 1997, Sheplak *et al.* 2004), skin-friction sensors should interfere as little as possible with the flow. Since an array formation is desirable to obtain a picture of the shear-stress distribution over an area, restraints are placed on the mounting of the sensor and placing of the electrical leads.

Shear-stress-sensing methods can be divided into two categories: direct and indirect. The direct way to measure shear stress is by mechanical means – by means of the skin-friction subset noted above – where the displacement of a spring-suspended element is measured. The spring suspension provides high directional specificity. Care needs to be taken, however, that the mechanical structure does project into the flow or create a recess in the wall that would affect the flow. Of the direct skin-friction sensors presented in the literature, only the devices described in Hyman *et al.* (1999) and Pan *et al.* (1999) have been scaled up to array sensing.

Indirect mechanical shear-stress sensing is also possible. For flows of not too high Reynolds numbers, or if a slight influence of the flow by the sensor can be tolerated, hair-type cantilever flow sensor arrays can be used for shear-stress imaging. A prerequisite for

accuracy is that the hairs are short enough to remain completely within the viscous sublayer. Arrays of hair-type sensors have been presented by Chen *et al.* (2003), van Baar *et al.* (2005) and Brücker *et al.* (2005) using piezoresistive, electrostatic, and optical hair-deflection readout, respectively. Brücker *et al.* (2005) is the only one of this group who has demonstrated full 2D sensitivity but the application is limited by the optical readout system.

In addition to studies on mechanical shear-stress sensing, much research has been performed on thermal shear-stress sensing. Unlike mechanical skin-friction sensors, thermal sensors require no moving parts and can, in most cases, be easily integrated into a flat surface. The output from single hot-element flow sensors can be correlated empirically with the skin friction at the heating element. The spatial resolution of the skin-friction sensor is limited mainly by the size of the heated area, which includes part of the insulating structure (Yoshino *et al.* 2001). Furthermore, as the conductive heat transfer into the supporting structure is frequency-dependent, the relationship between the transferred heat and the skin friction varies with varying flow regimes, which results in a nonunique relationship (Sheplak *et al.* 2004). Therefore, thermal shear-flow sensors need to be calibrated for the situation in which they are used, which is not always possible as with complex flow phenomena.

Figure 56(a) illustrates the principle of the set of thermal shear-stress sensors developed by Tai and Ho and coworkers. A polysilicon heater is placed on a thermally isolating membrane over a 2-μm-deep vacuum cavity. A device based on the same principle was developed by Cain *et al.* (2000) who used platinum as a heater material and a 10-μm-deep vacuum cavity. Yoshino *et al.* (2003) used a more easily fabricated membrane over an air cavity, perforating the membrane to allow wet etchant to form the cavity. Thermally insulating the heated element by means of polyimide around a heated chip area was performed by Kälvesten *et al.* (1995) (see **Figure 56(b)**). **Figure 56(c)** illustrates the principle of Kim and Lee (2006) and Rouhanizadeh *et al.* (2006) where back-side contacting reduces the flow obstruction due to the contacting leads while at the same time isolating them from the fluid.

Löfdahl *et al.* (2003) used a different approach, controlling the heat flow to the substrate instead of trying to minimize it. In incompressible, no-slip turbulent flow, there is a viscous sublayer close by where the velocity profile is linear due to the

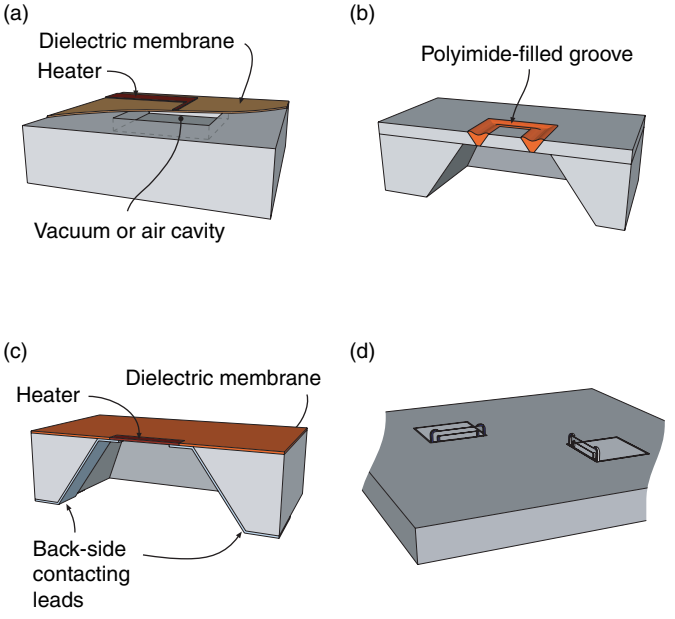


Figure 56 Thermal shear-flow sensors: (a) heater on vacuum cavity by Tai and Ho and coworkers (Huang *et al*. 1995, 1996, 1999, Jiang *et al*. 1995, 1996, 1997, Li *et al*. 2000, Lin *et al*. 2000, 2004, Liu *et al*. 1994, 1999, Tung *et al*. 2004, Xu *et al*. 2005a, b), alternatively on an air cavity. (Source: Yoshino *et al*. 2003.) (b) Groove-isolated membrane. (Source: Kälvesten *et al*. 1995.) (c) Thermally insulating membrane with back-side contacting. (Sources: Kim and Lee 2006, Rouhanizadeh *et al*. 2006.) (d) Prong-suspended hot wires. (Source: Löfdahl *et al*. 2003.)

dominance of the viscous forces (see **Figure 1**). The thickness of the viscous sublayer, λ_0, can be approximated as a function of the wall shear stress, the density of the fluid (ρ), and its kinematic viscosity (ν) (Löfdahl and Gad-el Hak 1999a):

$$\lambda_0 = 5\nu\sqrt{\frac{\rho}{\tau_0}}$$

By measuring the flow at a fixed distance from the wall, the wall shear stress can be derived from the following equation:

$$\tau_0 = \mu \left.\frac{\partial U}{\partial y}\right|_{y=0} = \mu\frac{U_d}{d}$$

where U_d is the velocity at a distance d from the wall (within the viscous sublayer). Micromachined aluminum hot wires at distances ranging from 50 to 250 μm were used (see **Figure 56(d)**). At greater distances from the wall, the shear-stress sensitivity increased and fluid temperature sensitivity decreased. However, the measurable range is decreased as the viscous sublayer becomes thinner with increasing wall shear stress. The heat flow to the substrate is controlled because the supports and the wall are thermally conductive and so remain at a CT (Haasl 2005).

van Oudheusden and Huijsing (1988) presented a direction-sensitive thermal shear-stress sensor in hot-chip mode using a thermopile to measure the temperature difference between the upstream and downstream ends of the chip. The set-up, in which the chip and the ceramic base are heated, had very large time constants (2–20 min), thus rendering it useless for dynamic applications.

Uncompensated thermal shear-stress sensors are extremely sensitive to fluid temperature changes; variations of 0.1°C have been reported to significantly affect the output signal (Löfdahl *et al.* 2003). Attempts to compensate for this by means of separate temperature sensors on-chip have been moderately successful within ranges of 0.7°C (Huang *et al.* 1999 and Jiang *et al.* 1996), which is acceptable for wind tunnel set-ups with tight temperature control. In general, for all thermal shear-stress sensors, increasing the overheat ratio $((T_{\text{wire}} - T_{\text{ambient}})/T_{\text{ambient}})$ increases the shear-stress sensitivity and decreases the fluid temperature sensitivity. At high overheat ratios, however, buoyancy effects caused by the heating of the flow by the wire start to affect the measurements (Appukuttan *et al.* 2003).

Despite these issues in accurately obtaining quantitative values for wall shear stress, flow control has been achieved with flexible arrays of shear-stress sensors held together by a polyimide film (Ho and Tai 1998). The shear-stress sensor arrays were attached to the leading edge of an unmanned aerial vehicle and provided feedback to micromachined balloon actuators to control its flight.

A micromachined optical shear-stress-sensing method has been presented by Fourguette *et al.* (2001). The principle, first developed with conventional optical components by Naqwi and Reynolds (1987), is derived from LDA. In LDA, two phase-shifted laser beams are crossed at a slight angle to generate an ellipsoid region with parallel interference fringes. As particles pass through this region, they scatter light with the intensity varying at a frequency $f = U/d_f$ where U is the velocity of the particles orthogonal to the fringes and d_f is the distance between the fringes. Naqwi and Reynolds (1987) instead used a diverging set of fringes (**Figure 57(a)**) and measured in a region within the viscous sublayer where the flow velocity is linearly dependent on the distance to the wall, $U_d = \tau_0 d/\mu$. As the distance between the fringes increases linearly with the distance from the wall (y), $d_f = k_f y$ (k_f is a constant), the frequency of refracted light off the particles becomes dependent only on the wall shear stress, $f = \tau_0/\mu k_f$. The micromachined optical interface between the laser and the photodetector is shown in **Figure 57(b)**. A 500-μm-thick fused silica or quartz substrate is used. A chromium pattern on the front side defines a pair of slits to create diffraction and a collecting window for the photodiode. On the back side, there are two polymethylmethacrylate (PMMA) diffractive lenses, which were fabricated by analog direct-write electron beam lithography followed by acetone development (Mouroulis *et al.* 1998, Wilson *et al.* 1996). The measuring volume was 15 μm × 20 μm × 90 μm, centered at 66 μm from the surface. As the viscous sublayer becomes thinner with increasing Reynolds numbers, the measurements were limited to Reynolds numbers up to 3×10^6 at 1 m downstream from a leading edge, at which the accuracy of a single-point measurement becomes 90%. Fourguette *et al.* (2003) presented a second generation of this sensor that uses three diffractive optical elements (DOEs): one transmitting DOE to split the incoming laser light into two beams and two receiver DOEs to measure the velocity at two different heights by means of photodetectors (**Figure 58**). The measurement volumes consist of a pair of 75-μm-long elongated light spots at 13.75 μm from each other, positioned at 65 and 110 μm from the surface. Using this set-up, and fitting the results to Spalding's formula

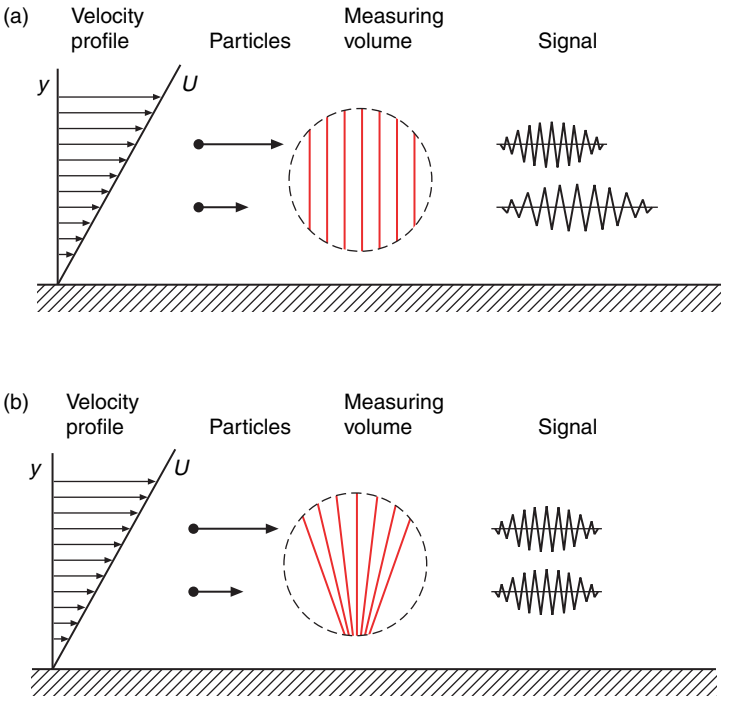


Figure 57 (a) Standard laser Doppler anemometry (LDA): parallel fringes result in a flow velocity-dependent signal as particles pass through the measuring volume. (b) LDA adapted for shear-stress sensing. The linearly diverging fringes result in a shear-stress-dependent signal. (Source: Naqwi A, Petrik S 1993 Fiber-optic dual-cylindrical wave sensor for measurement of wall velocity gradient in a fluid flow. *Appl. Opt.* **32(30)**, 6128–31.)

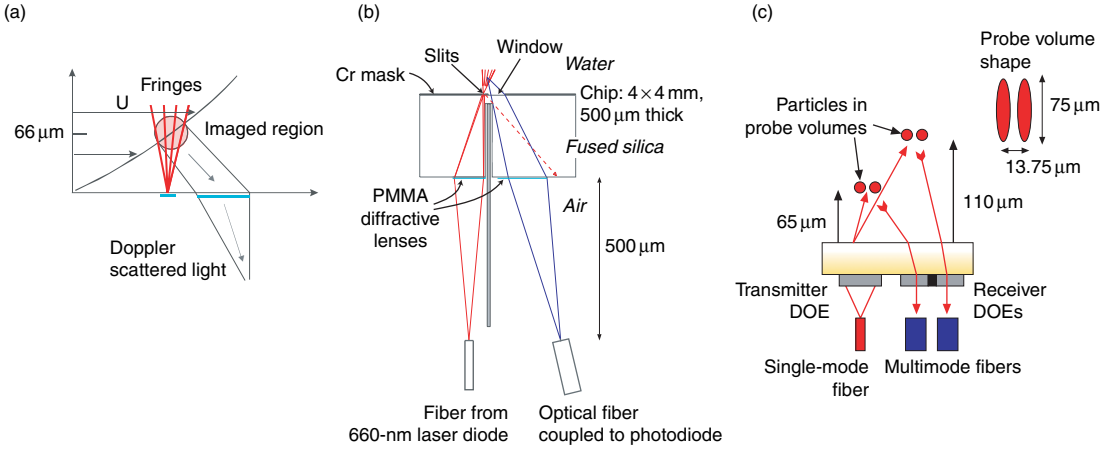


Figure 58 (a) Schematic of the optical shear-stress measurement. (b) Cross-section of the micro-optic shear-stress sensor. (Source: Fourguette D, Modarress D, Taugwalder F, Wilson D, Koochesfahani M, Gharib M 2001 Miniature and MOEMS flow sensors. *31st AIAA Fluid Dynamics Conference & Exhibition*, Anaheim, CA, USA.) (c) Cross-section of the second-generation optical shear-stress sensor. (Source: Fourguette D, Modarress D, Wilson D, Koochesfahani M, Gharib M 2003 An optical MEMS-based shear stress sensor for high Reynolds number applications. *41st Aerospace Sciences Meeting and Exhibition*, Reno, NV, USA.)

(Spalding 1961) that describes the mean velocity profile in a turbulent boundary layer up to a much larger distance than the viscous sublayer, 90% reliable results could be obtained up to a Reynolds number of 1.5×10^7.

The advantages of these optical methods are their nonintrusiveness (except for the seed particles), their linear response, and their high-frequency response. Their main drawback is the performance degradation

at large and unsteady velocity gradients because of the lack of near-wall seed particles (Naqwi and Reynolds 1991).

2.07.11.2 Medical Applications of Micromachined Flow Sensors

Micromachined flow sensors have been developed for use in three main areas of medicine: respiratory control, blood-flow measurements, and microdosing. Instruments that come in contact with the patient are often used only once to avoid contamination since disposing of them can be cheaper than disinfecting them. Microsystem technology with its low cost at high volumes, therefore, offers appealing prospects. The increasing use of implanted devices and minimally invasive surgery (MIS) has also created a need for miniature sensors that can be introduced with minimal tissue damage.

For respiratory control applications, the size requirement is not very stringent, the most important factor being low flow resistance, especially in situations where the patient can breathe partially or completely autonomously. Complicating aspects are the humidity and temperature variations in the breathing air and particles that may be carried with it and deposited on the sensor. Thermal flow sensors themselves do not generate pressure differences and generally have high sensitivity at low flow rates. However, because they usually measure flow only at one point, care must be taken of the varying ways in which the sensors are connected to the rest of the system as they might affect the flow profile. Respiratory flow sensors of the direction-insensitive thermal anemometer-type sensors have been presented (Bu *et al.* 1996, Svedin *et al.* 1998a), but an ability to distinguish inspiration from expiration is required if more than just the respiration frequency is to be reliably measured. Sensors based on both the double Wheatstone bridge thermal anemometry principle used by van Putten *et al.* (1997) and the calorimetric principle used by Bracio *et al.* (2000), Kaltsas and Nassiopoulou (2004), and Steurer and Kohl (1998) were developed with ventilator applications in mind, but many other thermal flow sensors would meet these specifications. Svedin *et al.* (2003a–c) developed two types of mechanical flow sensors based on lift-force measurement to obtain low pressure-drop measurements. As described in Section 2.07.3, Svedin *et al.* (2003a) combines a calorimetric flow sensor with the mechanical lift-force sensor to achieve high sensitivity over a large range. The main advantage of mechanical flow sensors is that they are less sensitive to particles and humidity variations.

The demands on blood-flow sensors are stringent. Those used in MIS have to be small enough to fit in the catheter that is used for introducing instruments. The packaged sensor should have a smooth surface to avoid blood clotting, and for thermal sensors, the allowed temperature variations are limited to about 5 K above ambient to avoid platelet and protein deposition as well as damage to the blood constituents (Kersjes *et al.* 1996, Tabata *et al.* 1987). The biocompatibility of the materials must be assured and the sensor has to be able to withstand standard sterilization procedures (Kotzar *et al.* 2002). As the entire sensor is placed in the liquid, care needs to be taken with packaging to prevent short circuits. Limited reverse flow can occur in blood vessels, which makes directional sensitivity desirable. Because of the size limitation, nearly all invasive blood-flow sensors are either thermal- or differential pressure-based. Kersjes *et al.* (1993, 1996) and Perichon *et al.* (2000) developed anemometer-type flow sensors with temperature compensation. Tanase *et al.* (2002) integrated a calorimetric flow sensor together with sensors for pressure and oxygen saturation into a single chip (**Figure 59(a)**). By integrating the sensor into the catheter, flow measurement can be performed online, while other surgical instruments or sensors are inserted with the guide wire. A significant disadvantage of anemometric and calorimetric flow sensing in blood vessels is that the dimensions of the blood vessels differ, which results in differing flow profiles within the vessel. The difficulty of controlling the lateral position of the sensor in the vessel adds further uncertainty to the volumetric blood-flow measurement. Differential-pressure measurement provides a partial solution, as the pressure does not vary much in the lateral direction of a blood vessel. Radi Medical Systems (Uppsala, Sweden) has developed a pressure sensor for measuring the pressure difference in the coronary arteries and hence to derive the volumetric flow and identify possible restrictions in the flow (stenoses). **Figure 60** illustrates how the sensor, mounted near the end of a guidewire, is inserted in the coronary vessels.

De Bruyne *et al.* (2001) and Pijls *et al.* (2002) used the resistive temperature compensation sensor on Radi's chip as the sensing part of a time-of-flight flow sensor. A small amount (3 ml) of saline at room temperature is quickly injected to cause a

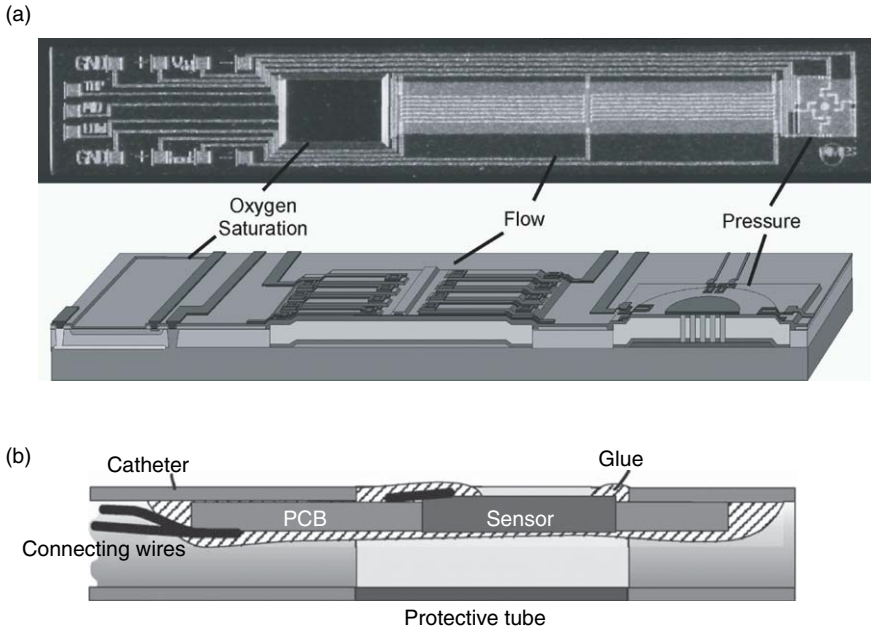


Figure 59 (a) Integrated sensor for measuring flow, pressure and oxygen saturation in the blood. (b) Sensor mount in the catheter. (Source: Tanase D, Goosen J F, Trimp P J, French P J 2002 Multi-parameter sensor system with intravascular navigation for catheter/guide wire application. *Sens. Actuators A Phys.* **97–98**, 116–24. © 2002 Elsevier.)

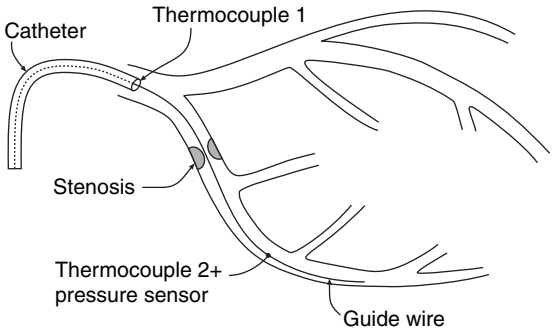


Figure 60 Thermal time-of-flight measurement to localize and evaluate stenoses. (Source: Pijls *et al*. 2002.)

temperature drop first at the tip of the catheter and then at the tip of the guide wire, 3 cm downstream. The injection time is measured by using the catheter tip as a thermistor. The small injection volume sufficed for accurate measurement and did not significantly affect the flow rate.

Micromachined flow sensors for blood perfusion measurements of the blood supply to tissue have been devised for both invasive and noninvasive applications. Kuttner *et al.* (1991) incorporated Ge thermistor arrays into the tip of a needle in order to probe the cerebral cortex of rabbits. Higurashi *et al.* (2003) and Ito *et al.* (1999) integrated a laser Doppler velocimetry system onto a substrate with an area less than 1 mm^2. A miniaturization of existing, bulky solutions can measure skin perfusion by observing the Doppler shift of red blood cells flowing just under the skin surface. Takahata and Gianchandani (2005) and Takahata *et al.* (2004) used a stent-like structure as an antenna to power differential-pressure and electromagnetic luminal flow sensors. DeHennis and Wise (2006) used a gold coil to provide power to a monolithic differential-pressure luminal flow sensor with integrated circuitry.

Blood shear stress has been found to be useful for monitoring blood clotting in cardiopulmonary bypass machines and implanted heart pumps. Luminal shear stress, i.e., the shear stress in blood vessels, has been shown to predict plaque formation in specific situations. For measuring shear stress, Rouhanizadeh *et al.* (2006) developed a sensor based on a back-side-contacted anemometer (see **Figure 56(c)**) that has been demonstrated in walls of artificial arteries (**Figure 61**).

Li *et al.* (2005) and Papageorgiou *et al.* (2001) designed flow sensors integrated in the base of neural microprobes, illustrated in **Figure 62**. The probes have electrodes integrated on the needle shaft to measure the electrical response of the neurons to electrical stimuli and a hollow shaft to allow drug injection. The flow sensor is used to control the

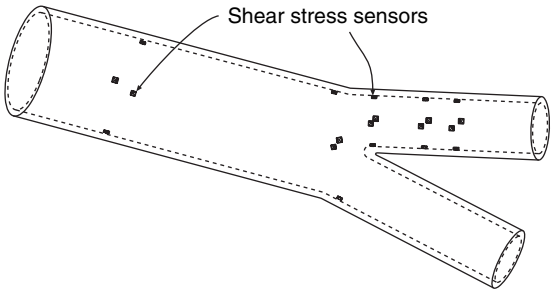


Figure 61 Shear-stress sensors placed on the wall of an artificial artery. (Source: Rouhanizadeh M, Soundararajan G, Lo R, Arcas D, Browand F, Hsiai T 2006 MEMS sensors to resolve spatial variations in shear stress in a 3D blood vessel bifurcation model. *IEEE Sens. J.* **6(1)**, 78–88. © 2006 IEEE.)

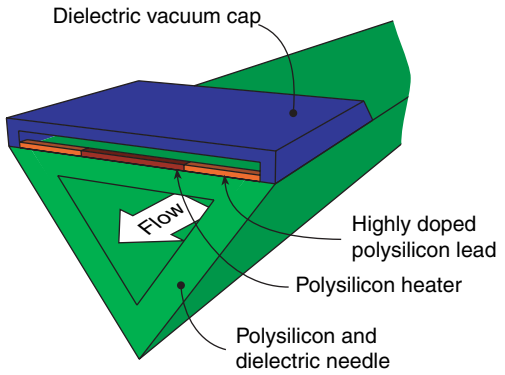


Figure 62 Anemometer integrated on the tip of a hollow needle. (Source: Li Y, Baek K, Gulari M, Lin D, Wise K 2005 A vacuum-isolated thermal microflowmeter for *in-vivo* drug delivery. *Sensors, 2005 IEEE*, pp. 668–71. © 2005 IEEE.)

dosage and to ensure that no blockage has occurred. The heating budget for probes in the central nervous system is even more stringent than for blood sensors as the temperature rise must be kept below 2°C to avoid neural damage. Since the entire sensor is in a liquid environment, the outside is thermally insulated by an air cavity to reduce unnecessary heat loss.

References

Adrian R J 2005 Twenty years of particle image velocimetry. *Exp. Fluids* **39(2)**, 159–69

Ahmad N 1985 Surface acoustic wave flow sensor. *Ultrasonics Symposium, IEEE 1985*, San Francisco, CA, USA, pp. 483–5

Al Khalfioui M, Michez A, Giani A, Boyer A, Foucaran A 2003 Anemometer based on Seebeck effect. *Sens. Actuators A Phys.* **107(1)**, 36–41

Appukuttan A, Shyy W, Sheplak M, Cattafesta L 2003 Mixed convection induced by MEMS-based thermal shear stress sensors. *Numer. Heat Transf. A (Appl.)* **43(3)**, 283–305

Arkilic E, Schmidt M, Breuer K 1997 Gaseous slip flow in long microchannels. *J. Microelectromech. Syst.* **6(2)**, 167–78

Ashauer M, Glosch H, Hedrich F, Hey N, Sandmaier H, Lang W 1998 Thermal flow sensor for liquids and gases. *Proceedings of the IEEE Micro Electro Mechanical Systems (MEMS)*, Heidelberg, Germany, pp. 351–5

Ashauer M, Glosch H, Hedrich F, Hey N, Sandmaier H, Lang W 1999 Thermal flow sensor for liquids and gases based on combinations of two principles. *Sens. Actuators A (Phys.)* **A73(1–2)**, 7–13

Ashauer M, Scholz H, Briegel R, Sandmaier H, Lang W 2001 Thermal flow sensors for very small flow rate. TRANSDUCERS '01. EUROSENSORS XV. 11th Int. Conf. Solid-State Sensors and Actuators. Digest of Technical Papers, Vol. 2, Munich, Germany, pp. 1464–7

Ayliffe H, Rabbitt R 2003 An electric impedance based microelectromechanical system flow sensor for ionic solutions. *Meas. Sci. Technol.* **14(8)**, 1321–7

Ayliffe H, Frazier A, Rabbitt R 1999 Electric impedance spectroscopy using microchannels with integrated metal electrodes. *J. Microelectromech. Syst.* **8(1)**, 50–7

van Baar J J, Wiegerink R J, Lammerink T S J, Krijnen G J M, Elwenspoek M 2001 Micromachined structures for thermal measurements of fluid and flow parameters. *J. Micromech. Microeng.* **11(4)**, 311–8

van Baar J, Dijkstra M, Wiegerink R, Lammerink T, Krijnen G 2003 Fabrication of arrays of artificial hairs for complex flow pattern recognition. *Proc. IEEE Sensors 2003*, Toronto, ON, Canada, Vol. 1, pp. 332–6

van Baar J, Dijkstra M, Wiegerink R, Lammerink T, De Boer R, Krijnen G 2005 Arrays of cricket-inspired sensory hairs with capacitive motion detection. *Proc. IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS)*, Miami, FL, USA, pp. 646–9

Barlian A, Narain R, Li J, Quance C, Ho A, Mukundan V, Pruitt B 2006 Piezo resistive MEMS underwater shear stress sensors. *Micro Electro Mechanical Systems 2006, 19th IEEE Int. Conf. MEMS 2006*, Istanbul, Turkey, pp. 626–9

Becker E, Ehrfeld W, Hagmann P, Maner A, Munchmeyer D 1986 Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming, and plastic moulding (LIGA process). *Microelectron. Eng.* **4(1)**, 35–56

Bedö G, Fannasch H, Müller R 2000 Silicon flow sensor for gases and liquids using AC measurements. *Sens. Actuators A Phys.* **85(1)**, 124–32

Bentley J P 1983a *Principles of Measurement Systems*. Longman, London, Chap. 12.3.1, Differential pressure flowmeters, pp. 283–91

Bentley J P 1983b *Principles of Measurement Systems*. Longman, London, Chap. 16, Ultrasonic measurement systems, pp. 408–12

Berberig O, Nottmeyer K, Mizuno J, Kanai Y, Kobayashi T 1997 Prandtl micro flow sensor (PMFS): A novel silicon diaphragm capacitive sensor for flow velocity measurement. *Int. Conf. Solid-State Sensors and Actuators, Transducers '97 Chicago, IL, USA, Proceedings*, Vol. 1, pp. 155–8

Betzner T M, Doty J R, Hamad A M A, Henderson H T, Berger F G 1996 Structural design and characteristics of a thermally isolated, sensitivity-enhanced, bulk-micromachined, silicon flow sensor. *J. Micromech. Microeng.* **6(2)**, 217–27

Billat S, Glosch H, Kunze M, Hedrich F, Frech J, Auber J, Lang W, Sandmaier H, Wimmer W 2001 Convection-based micromachined inclinometer using SOI technology. *Tech. Dig. MEMS 2001. 14th IEEE Int. Conf. Micro Electro Mechanical Systems*, Interlaken, Switzerland, pp. 159–61

Billat S, Glosch H, Kunze M, Hedrich F, Frech J, Auber J, Sandmaier H, Wimmer W, Lang W 2002 Micromachined

inclinometer with high sensitivity and very good stability. *Sens. Actuators A Phys.* **97–98**, 125–30

Boillat M, van der Wiel A, Hoogerwerf A, de Rooij N 1995 A differential pressure liquid flow sensor for flow regulation and dosing systems. *Proceedings of the Micro Electro Mechanical Systems, 1995, MEMS '95. IEEE*, Amsterdam, The Netherlands, pp. 350–2

Bosman J, De Bruijn J, Riedijk F, Van Oudheusden B, Huijsing J 1992 Integrated smart two-dimensional thermal flow sensor with Seebeck-voltage-to-frequency conversion. *Sens. Actuators A Phys.* **31(1–3 pt 3)**, 9–16

Bouwstra S, Kemna P, Legtenberg R 1989 Thermally excited resonating membrane mass flow sensor. *Sens. Actuators* **20(3)**, 213–23

Bouwstra S, Legtenberg R, Tilmans H, Elwenspoek M 1990 Resonating microbridge mass flow sensor. *Sens. Actuators A (Phys.)* **A21(1–3)**, 332–5

Brace J, Sanfelippo T, Joshi S 1989 Mass flow sensing using surface acoustic waves. *IEEE 1989 Ultrasonics Symposium Proceedings*, pp. 573–8

Bracio B, Fasching R, Kohl F, Krocza J 2000 A smart thin-film flow sensor for biomedical applications. *Proc. 22nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Vol. 4, pp. 2800–1

Breuer K S, Bayt R L, Nayaar A 1999 Measurement of shear stress and temperature using MEMS fabricated sensors. *ASME Microelectromechanical Systems.* Subdivision 1999, Vol. 1, American Society of Mechanical Engineers, Fairfield, NJ, pp. 229–33

Briers J D 2001 Laser Doppler, speckle and related techniques for blood perfusion mapping and imaging. *Physiol. Meas.* **22(4)**, R35–66

de Bree H-E 2003 An overview of microflown technologies. *Acta Acustica United Acustica* **89(1)**, 163–72

de Bree H-E, Leussink P, Korthorst T, Jansen H, Lammerink T, Elwenspoek M 1995 The mu-flown, a novel device measuring acoustical flows. *8th Int. Conf. Solid-State Sensors and Actuators and Eurosensors IX. Dig. Tech. Pap*, Stockholm, Sweden, pp. 536–9

de Bree H-E, Korthorst T, Leussink P, Janssen H, Elwenspoek M 1996 A method to measure apparent acoustic pressure, flow gradient and acoustic intensity using two micromachined flow microphones. *Eurosensors X*, Leuven, Belgium

de Bree H-E, Jansen H, Lammerink T, Krijnen G, Elwenspoek M 1999 Bi-directional fast flow sensor with a large dynamic range. *J. Micromech. Microeng.* **9(2)**, 186–9

Brücker C, Spatz J, Schrüder W 2005 Feasability study of wall shear stress imaging using microstructured surfaces with flexible micropillars. *Exp. Fluids* **39(2)**, 464–74

Bruschi P, Navarrini D, Piotto M 2004a A flow sensor for liquids based on a single temperature sensor operated in pulsed mode. *Sens. Actuators A (Phys.)* **A110(1–3)**, 269–75

Bruschi P, Navarrini D, Piotto M, Raffa G 2004b Sensitivity improvement of integrated thermal anemometers obtained by jet flow impingement. *Sens. Actuators A Phys.* **113(3)**, 301–6

Bruschi P, Diligenti A, Navarrini D, Piotto M 2005 A double heater integrated gas flow sensor with thermal feedback. *Sens. Actuators A Phys.* **123–124**, 210–15

Bruun H H 1995 *Hotwire Anemometry: Principles and Signal Analysis*. Oxford University Press Inc, New York

Bu J-U, Kim T-Y, Kim I-S, Jun Y-S, Shim Y-C, Kim S-T 1996 Silicon-based thermal comfort sensing device. *Sens. Actuators A Phys.* **54(1–3)**, 468–71

Buchner R, Maiwald M, Sosna C, Schary T, Benecke W, Lang W 2006 Miniaturised thermal flow sensors for rough environments. *19th IEEE Int. Conf. Micro Electro Mechanical Systems 2006 (MEMS 2006)*, Istanbul, Turkey, pp. 582–5

Cain A, Chandrasekaran V, Nishida T, Sheplak M 2000 Development of a wafer-bonded, siliconnitride membrane thermal shear-stress sensor with platinum sensing element. *Technical Digest Solid-State Sensor and Actuator Workshop (TRF)*, Hilton Head Island, SC, USA, pp. 300–3

Castaner L, Jimenez V, Dominguez M, Masana F, Rodriguez A 1997 Flow measurement hits home. *Circuits Devices Mag. IEEE* **13(1)**, 14–18

Chaehoi A, Dumas N, Mailly F, Latorre L, Nouet P 2005 Absolute pitch, roll and yaw measurement on CMOS. *Sensors, 2005 IEEE*, pp. 133–6

Chavan A, Prawdzik D, Manlove G, Long S, Dubois P, Dewes B 2003 Non-linear temperature compensation for a micromachined bi-directional thermal flow sensor. *IEEE 16th Annu. Int. Conf. Micro Electro Mechanical Systems 2003 (MEMS '03)*, Kyoto, Japan, pp. 510–13

Chen J, Liu C 2003 Development and characterization of surface micromachined, out-of-plane hotwire anemometer. *J. Microelectromech. Syst.* **12(6)**, 979–88

Chen J, Zou J, Liu C 2002 A surface micromachined, out-of-plane anemometer. *Tech. Dig.Proc. MEMS 2002 IEEE Int. Conf. 15th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2002)*, IEEE, Las Vegas, NV, USA, pp. 332–5

Chen J, Fan Z, Zou J, Engel J, Liu C 2003 Two-dimensional micromachined flow sensor array for fluid mechanics studies. *J. Aerospace Eng.* **16(2)**, 85–97

Chen J, Engel J, Chen N, Liu C 2005 A monolithic integrated array of out-of-plane hot-wire flow sensors and demonstration of boundary-layer flow imaging. *18th IEEE Int. Conf. Micro Electro Mechanical Systems, 2005 (MEMS 2005)*, Miami, FL, USA, pp. 299–302

Chen X, Lal A 2001 Integrated pressure and flow sensor in silicon-based ultrasonic surgical actuator. *Proc. IEEE Ultrason. Symp.* **2**, 1373–6

Cho S T, Najafi K, Lowman C E, Wise K D 1992 An ultrasensitive silicon pressure-based microflow sensor. *IEEE Trans. Electron. Devices* **39(4)**, 825–35

Chun D, Yaqiang W, Wang-Yuelin, Zhonghe J 1998 Discussion on the optical-coupling in silicon optical-type micromechanical sensors. *Proc. 5th Int. Conf. Solid-State and Integrated Circuit Technology*, pp. 953–6

Chung J, Lee K-H, Lee J 2003 Multi-walled carbon nanotube sensors. *TRANSDUCERS '03. 12th Int. Conf. Solid-State Sensors, Actuators and Microsystems. Dig. Tech. Pap.* Vol. 1, pp. 718–21

Cittadine A 2000 MEMS reshapes ultrasonic sensing. *Sensors* **17(2)**, 17–18

Collins J, Lee A 2004 Microfluidic flow transducer based on the measurement of electrical admittance. *Lab Chip* **4(1)**, 7–10

Dao D V, Toriyama T, Wells J, Sugiyama S 2002 Six-degree of freedom micro force-moment sensor for application in geophysics. *Tech. Dig. Proc. MEMS 2002 IEEE Int. Conf. 15th IEEE Int. Conf. Micro Electro Mechanical Systems*, Boston, MA, USA, pp. 312–15

Dao D, Toriyama T, Sugiyama S, Nguyen A, Wells J 2003 A MEMS-based microsensor to measure all six components of force and moment on a near-wall particle in turbulent flow. *TRANSDUCERS '03. 12th Int. Conf. Solid-State Sensors, Actuators and Microsystems. Dig. Tech. Pap.*, Vol. 1, pp. 504–7

De Bruyne B, Pijls N H J, Smith L, Wievegg M, Heyndrickx G R 2001 Coronary thermodilution to assess flow reserve: Experimental validation. *Circulation* **104**, 2003–6

DeHennis A, Wise K 2006 A fully integrated multisite pressure sensor for wireless arterial flow characterization. *J. Microelectromech. Syst.* **15(3)**, 678–85

Desai A V, Haque M A 2004 Design and fabrication of a direction sensitive MEMS shear stress sensor with high spatial and temporal resolution. *J. Micromech. Microeng.* **14(12)**, 1718–25

Dijkstra M, van Baar J, Wiegerink R, Lammerink T, de Boer J, Krijnen G 2005 Artificial sensory hairs based on the flow

sensitive receptor hairs of crickets. *J. Micromech. Microeng.* **15(7)**, 132–8

Dillner U, Kessler E, Poser S, Baier V, Muller J 1997 Low power consumption thermal gas-flow sensor based on thermopiles of highly effective thermoelectric materials. *Sens. Actuators A Phys.* **60(1–3)**, 1–4

Dittmann D, Ahrens R, Rummler Z, Schlote-Holubek K, Schomburg W 2001 Low-cost flow transducer fabricated with the AMANDA-process. *TRANSDUCERS '01. EUROSENSORS XV. 11th Int. Conf. Solid-State Sensors and Actuators. Dig. Tech. Pap.*, Vol. 2, pp. 1472–5

Dominguez D, Bonvalot B, Chau M T, Suski J 1993 Fabrication and characterization of a thermal flow sensor based on porous silicon technology. *J. Micromech. Microeng.* **3(4)**, 247–9

Dominguez M, Ricart J, Moreno A, Contesti X, Garriga S 2005 Low cost PCB thermal sigma–delta air flowmeter with improved thermal isolation. *Sens. Actuators A Phys.* **121(2)**, 388–94

Ebefors T 2000 Polyimide v-groove joints for three-dimensional silicon transducers. Ph.D. thesis, Royal Institute of Technology (KTH)

Ebefors T, Kälvesten E, Stemme G 1998 Three dimensional silicon triple-hot-wire anemometer based on polyimide joints. *Proc. MEMS 98. IEEE. 11th Annu. Int. Workshop Micro Electro Mechanical Systems*. IEEE, Heidelberg, Germany, pp. 93–8

Ebefors T, Ulfstedt-Mattsson J, Kaelvesten E, Stemme G 1999 3D micromachined devices based on polyimide joint technology. *Proc. SPIE – Int. Soc. Opt. Eng.* **3892**, 118–32

Elwenspoek M 1999 Thermal flow micro sensors.*1999 International Semiconductor Conference 1999. CAS '99 Proceedings*, Sinaia, Romania, Vol. 2, pp. 423–35

Elwenspoek M, Wiegerink R 2001 *Mechanical Microsensors*. Springer, Berlin, Chap. Flow Sensors, pp. 153–208

Engel J, Chen J, Chen N, Pandya S, Liu C 2005 Development and characterization of an artificial hair cell based on polyurethane elastomer and force sensitive resistors. *Sensors, 2005 IEEE*, Irvine, CA, USA, pp. 1014–17

Enoksson P, Stemme G, Stemme E 1997 A silicon resonant sensor structure for Coriolis mass-flow measurements. *J. Microelectromech. Syst.* **6(2)**, 119–25

Ernst H, Jachimowicz A, Urban G A 2002 High resolution flow characterization in Bio-MEMS. *Sens. Actuators A Phys.* **100(1)**, 54–62

Esashi M 1991 Micro flow sensor and integrated magnetic oxygen sensor using it. *TRANSDUCERS '91. 1991 Int. Conf. Solid-State Sensors and Actuators. Dig. Tech. Pap.*, San Francisco, CA, USA, pp. 34–7

Fan C-W, Nien C-C, Gwo T-T, Chu K-H 2002a Coriolis force type flow meter using optical interferometer. Tech. Rep. *US Pat.* 6 722 209, Industrial Technology Research Institute (Hsinchu Hsien, TW)

Fan Z, Chen J, Zou J, Bullen D, Liu C, Delcomyn F 2002b Design and fabrication of artificial lateral line flow sensors. *J. Micromech. Microeng.* **12(5)**, 655–61

Fourguette D, Modarress D, Taugwalder F, Wilson D, Koochesfahani M, Gharib M 2001 Miniature and MOEMS flow sensors. *31st AIAA Fluid Dynamics Conference & Exhibition*, Anaheim, CA, USA

Fourguette D, Modarress D, Wilson D, Koochesfahani M, Gharib M 2003 An optical MEMS-based shear stress sensor for high Reynolds number applications. *41st Aerospace Sciences Meeting and Exhibition*, Reno, NV, USA

Frechette L, Jacobson S, Breuer K, Ehrich F, Ghodssi R, Khanna R, Wong C W, Zhang X, Schmidt M, Epstein A 2000 Demonstration of a microfabricated high-speed turbine supported on gas bearings. *Technical Digest. Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, pp. 43–7

Frederick J, Zemel J, Goldfine N 1985 Pyroelectric anemometers: Experimental geometric considerations. *J. Appl. Phys.* **57(11)**, 4936–43

Fricke K 1994 A micromachined mass-flow sensor with integrated electronics on GaAs. *Sens. Actuators A Phys.* **45(2)**, 91–4

Friedberger A, Kreisl P, Muller G, Kassing R 2001 A versatile and modularizable micromachining process for the fabrication of thermal microsensors and microactuators. *J. Micromech. Microeng.* **11(6)**, 623–9

Fürjes P, Légrádi G, Dücso C, Aszódi A, Bársony I 2004 Thermal characterisation of a direction dependent flow sensor. *Sens. Actuators A Phys.* **115(2–3)**, 417–23

Furuberg L, Wang D, Vogl A, Merveille C, Hansen S I, Rogne H, Solli L 2003 New pressure sensor, optical mirrors and a flow rate sensor manufactured by bulk micromachining. *Proc. Sensor 2003*, Nuremberg, Germany, p. A6.3

Gajda M A, Ahmed H 1995 Applications of thermal silicon sensors on membranes. *Sens. Actuators A Phys.* **49(1–2)**, 1–9

Gamage S K, Okulan N, Henderson H T 2000 Behavior of bulk micromachined silicon flow sensor in the negative differential resistance regime. *J. Micromech. Microeng.* **10(3)**, 421–9

Gass V, van der Schoot B, de Rooij N 1993 Nanofluid handling by micro-flow-sensor based on drag force measurements. *Proc. IEEE Micro Electro Mechanical Systems. An Investigation of Micro Structures, Sensors, Actuators, Machines and Systems*, Fort Lauderdale, FL, USA, pp. 167–72

Ghosh S, Sood A, Kumar N 2003 Carbon nanotube flow sensors. *Science* **299(5609)**, 1042–4

Ghosh S, Sood A, Ramaswamy S, Kumar N 2004 Flow-induced voltage and current generation in carbon nanotubes. *Phys. Rev. B (Condens. Matter Mater. Phys.)* **70(20)**, 205423–1

Glaninger A, Jachimowicz A, Kohl F, Chabicovsky R, Urban G 2000 Wide range semiconductor flow sensors. *Sens. Actuators A (Phys.)* **A85(1–3)**, 139–46

Go J S, Kim S J, Lim G, Yun H, Lee J, Song I, Pak Y E 2001 Heat transfer enhancement using flow-induced vibration of a microfin array. *Sens. Actuators A Phys.* **90(3)**, 232–9

Gongora-Rubio M, Sola-Laguna L M, Moffett P J, Santiago-Aviles J J 1999 The utilization of low temperature co-fired ceramics (LTCC-ML) technology for meso-scale EMS, a simple thermistor based flow sensor. *Sens. Actuators A Phys.* **73(3)**, 215–21

Goodfellow Corporation Website. Online resource, access date 2007: URL http://www.goodfellow.com

Haasl S 2005 Assembly of microsystems for optical and fluidic applications. Ph.D. thesis, Royal Institute of Technology

Haasl S, Mucha D, Chernoray V, Ebefors T, Enoksson P, Löfdahl L, Stemme G 2005 Hybrid-mounted micromachined aluminum hotwires for wall shear-stress measurements. *J. Microelectromech. Syst.* **14(2)**, 254–60

Häberli A, Mayer F, Jaeggi D, Baltes H 1997 IC microsensors – Between system and technology. *Proc. 2nd IEEE-CAS Region 8 Workshop Analog and Mixed IC Design, 1997*, Baveno, Italy, pp. 36–40

Han G, Bird J C, Johan K, Westin A, Cao Z, Breuer K S 2004 Infrared diagnostics for measuring fluid and solid motion inside silicon microdevices. *Microscale Thermophys. Eng.* **8(2)**, 169–82

Haronian D 1999 Direct integration (DI) of solid state stress sensors with single crystal micro-electromechanical systems for integrated displacement sensing. *12th IEEE Int. Conf. Micro Electro Mechanical Systems 1999 (MEMS '99)*, Orlando, FL, USA, pp. 88–93

Hauptmann P, Hoppe N, Puttmer A 2002 Application of ultrasonic sensors in the process industry. *Meas. Sci. Technol.* **13(8)**, 73–83

Higurashi E, Sawada R, Ito T 2003 An integrated laser blood flowmeter. *J. Lightwave Technol.* **21(3)**, 591–5

Himmelhaus M, Bley P, Mohr J, Wallrabe U 1992 Integrated measuring system for the detection of the revolutions of LIGA microturbines in view of a volumetric flow sensor. *J. Micromech. Microeng.* **2(3)**, 196–8

Ho C-M, Tai Y-C 1998 Micro-electro-mechanical systems (MEMS) and fluid flows. *Annu. Rev. Fluid Mech.* **30**, 579–612

Holman J P 1997 *Heat Transfer*, 8th edn. McGraw-Hill, New York, ISBN 0-07-114320-3

Holmes A, Hong G, Pullen K, Buffard K 2004 Axial-flow microturbine with electromagnetic generator: Design, CFD simulation, and prototype demonstration. *17th IEEE Int. Conf. Micro Electro Mechanical Systems, 2004 (MEMS '04)*, Maastricht, The Netherlands, pp. 568–71

van Honschoten J, van Baar J, de Bree H E, Lammerink T, Krijnen G, Elwenspoek M 2000 Application of a microflown as a low-cost level sensor. *J. Micromech. Microeng.* **10(2)**, 250–3

van Honschoten J, Svetovoy V, Krijnen G, Elwenspoek M 2005 Optimization of a thermal flow sensor for acoustic particle velocity measurements. *J. Microelectromech. Syst.* **14(3)**, 436–3

Horowitz S, Chen T, Chandrasekaran V, Tedjojuwono K, Nishida T, cattafesta L, Sheplak M 2004 A micromachined geometric Moire interferometric floating-element shear stress sensor. *AIAA Pap.*, pp. 6309–18

Hsieh H Y, Zemel J N 1995 Pyroelectric anemometry: Frequency, geometry and gas dependence. *Sens. Actuators A Phys.* **49(3)**, 133–40

Hsieh H, Spetz A, Zemel J 1991 Wide range pyroelectric anemometers for gas flow measurements. *Solid-State Sensors and Actuators 1991. Digest of Technical Papers, TRANSDUCERS '91*, pp. 38–40

Hsieh H, Spetz A, Zemel J 1995a Pyroelectric anemometry: Vector and swirl measurements. *Sens. Actuators A (Phys.)* **A49(3)**, 141–7

Hsieh H Y, Zemel J N, Bau H H 1995b Pyroelectric anemometry: Theory of operation. *Sens. Actuators A Phys.* **49(3)**, 125–32

Huang J-B, Tong Q-Y 1989 Integrated multi-function sensor for flow velocity, temperature and vacuum measurements. *Sens. Actuators* **19(1)**, 3–11

Huang J-B, Liu C, Jiang F, Tung S, Tai Y-C, Ho C-M 1995 Fluidic shear-stress measurement using surface-micromachined sensors. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, Hong Kong, pp. 16–19

Huang J-B, Tung S, Ho C-M, Liu C, Tai Y-C 1996 Improved micro thermal shear-stress sensor. *IEEE Trans. Instrum. Meas.* **45(2)**, 570–4

Huang J, Jiang F, Tai Y, Ho C 1999 A micro-electro-mechanical-system-based thermal shear-stress sensor with self-frequency compensation. *Meas. Sci. Technol.* **10(8)**, 687–96

Hung S-T, Wong S-C, Fang W 2000 The development and application of microthermal sensors with a mesh-membrane supporting structure. *Sens. Actuators A Phys.* **84(1–2)**, 70–5

Hyman D, Pan T, Reshotko E, Mehregany M 1999 Microfabricated shear stress sensors, part 2: Testing and calibration. *AIAA J.* **37(1)**, 73–8

Iijima S 1991 Helical microtubules of graphitic carbon. *Nature* **354(6348)**, 56–8

Ito T, Sawada R, Higurashi E 1999 Integrated microlaser Doppler velocimeter. *J. Lightwave Technol.* **17(1)**, 30–4

Jang J, Wereley S 2004 A capacitive micro gas flow sensor based on slip flow. *17th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2004). Tech. Dig.*, Maastricht, The Netherlands, pp. 540–3

Jiang F, Tai Y-C, Ho C-M, Karan R, Garstenauer M 1994a Theoretical and experimental studies of micromachined hot-wire anemometers. *Technical Digest. International Electron Devices Meeting 1994*. IEEE, San Francisco, CA, USA, pp. 139–42

Jiang F, Tai Y-C, Ho C-M, Li W 1994b A micromachined polysilicon hot-wire anemometer. *Technical Digest. Solid-State Sensor and Actuator Workshop. Transducer Res. Found*, Hilton Head Island, SC, USA, pp. 264–7

Jiang F, Tai Y-C, Huang J-B, Ho C-M 1995 Polysilicon structures for shear stress sensors. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, Hong Kong, pp. 12–15

Jiang F, Tai Y-C, Gupta B, Goodman R, Tung S, Huang J-B, Ho C-M 1996 A surface-micromachined shear stress imager. *Proc. IEEE, 9th Annu. Int. Workshop Micro Electro Mechanical Systems. An Investigation of Micro Structures, Sensors, Actuators, Machines and Systems*, San Diego, CA, USA, pp. 110–15

Jiang F, Tai Y-C, Walsh K, Tsao T, Lee G-B, Ho C-M 1997 A flexible MEMS technology and its first application to shear stress sensor skin. *Proc. IEEE. 10th Annu. Int. Workshop Micro Electro Mechanical Systems. An Investigation of Micro Structures, Sensors, Actuators, Machines and Robots (Cat. No.97CH36021)*, Nagoya, Japan, pp. 465–70

Jin X, Ladabaum I, Khuri-Yakub B 1998 The microfabrication of capacitive ultrasonic transducers. *J. Microelectromech. Syst.* **7(3)**, 295–302

Johnson R G, Higashi R E 1987 A highly sensitive silicon chip microtransducer for air flow and differential pressure sensing applications. *Sens. Actuators* **11**, 63–72

Jørgensen F E 1971 Directional sensitivity of wire and fibre-film probes. *DISA Info.* **11**, 31–7

Joshi S 1989 Use of a surface-acoustic-wave (SAW) device to measure gas flow. *IEEE Trans. Instrum. Meas.* **38(3)**, 824–6

Joshi S 1994 Flow sensors based on surface acoustic waves. *Sens. Actuators A (Phys.)* **A44(3)**, 191–7

Joshi S, Jin Y 1990 Application of a surface-acoustic-wave device for measurement of liquid flow rate. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **37(5)**, 475–7

Kaltsas G, Nassiopoulou A 1999 Novel CMOS compatible monolithic silicon gas flow sensor with porous silicon thermal isolation. *Sens. Actuators A (Phys.)* **A76(1–3)**, 133–8

Kaltsas G, Nassiopoulou A G 2004 Gas flow meter for application in medical equipment for respiratory control: Study of the housing. *Sens. Actuators A Phys.* **110(1–3)**, 413–22

Kaltsas G, Nassiopoulos A, Nassiopoulou A 2002 Characterization of a silicon thermal gas-flow sensor with porous silicon thermal isolation. *IEEE Sens. J.* **2(5)**, 463–75

Kälvesten E, Vieider C, Löfdahl L, Stemme G 1995 Integrated pressure-flow sensor for correlation measurements in turbulent gas flows. *Int. Conf. Solid-State Sensors Actuators, Proc.*, IEEE, Piscataway, NJ, USA, Vol. 2, pp. 428–31

Kälvesten E, Vieider C, Löfdahl L, Stemme G 1996 An integrated pressure-flow sensor for correlation measurements in turbulent gas flows. *Sens. Actuators A Phys.* **52(1–3)**, 51–8

Kang J, Park S 2000 Integrated comfort sensing system on indoor climate. *Sens. Actuators A Phys.* **82(1)**, 302–7

Kersjes R, Mokwa W 1995 Fast liquid flow sensor with thermal isolation by oxide-filled trenches. *Sens. Actuators A Phys.* **47(1–3 pt 4)**, 373–9

Kersjes R, Eichholz J, Langerbein A, Manoli Y, Mokwa W 1993 An integrated sensor for invasive blood-velocity measurement. *Sens. Actuators A Phys.* **37–38**, 674–8

Kersjes R, Liebscher F, Spiegel E, Manoli Y, Mokwa W 1996 An invasive catheter flow sensor with on-chip CMOS readout electronics for the on-line determination of blood flow. *Sens. Actuators A (Phys.)* **A54(1–3)**, 563–7

Kim I C, Lee S J 2006 Characterization of a miniature thermal shear-stress sensor with backside connections. *Sens. Actuators A (Phys.)* **128(2)**, 305–11

Kim S, Nam T, Park S 2004 Measurement of flow direction and velocity using a micromachined flow sensor. *Sens. Actuators A Phys.* **114(2–3)**, 312–18

Kim Y-M, Seo C-T, Eun D-S, Park S-G, Jo C-S, Lee J-H 2003 Characteristics of cantilever beam fabricated by silicon micromachining for flow sensor application.*Sensors 2003 Proc. IEEE*, Toronto, Canada, Vol. 1, pp, 642–6

King L V 1914 On the convection of heat from small cylinders in a stream of fluid: Determination of the convection constants of small platinum wires with applications to hot-wire anemometry. *Philos. Trans. R. Soc.* **214**, 373–432

Koch M, Schabmueller C, Evans A, Brunnschweiler A 1999 Micromachined chemical reaction system. *Sens. Actuators A (Phys.)* **A74(1–3)**, 207–10

Kohl F, Jachimowicz A, Steurer J, Glatz R, Kuttner J, Biacovsky D, Olcaytug F, Urban G 1994 A micromachined flow sensor for liquid and gaseous fluids. *Sens. Actuators A (Phys.)* **A41(1–3)**, 293–9

Kotzar G, Freas M, Abel P, Fleischman A, Roy S, Zorman C, Moran J M, Melzak J 2002 Evaluation of MEMS materials of construction for implantable medical devices. *Biomaterials* **23(13)**, 2737–50

Král P, Shapiro M 2001 Nanotube electron drag in flowing liquids. *Phys. Rev. Lett.* **86(1)**, 131–4

Kraus T, Gunther A, de Mas N, Schmidt M, Jensen K 2004 An integrated multiphase flow sensor for microchannels. *Exp. Fluids* **36(6)**, 819–32

Kuehnel W, Sherman S 1994 A surface micromachined silicon accelerometer with on-chip detection circuitry. *Sens. Actuators A (Phys.)* **A45(1)**, 7–16

van Kuijk J, Lammerink T S, de Bree H-E, Elwenspoek M, Fluitman J H J 1995 Multiparameter detection in fluid flows. *Sens. Actuators A (Phys.)* **A47(1–3)**, 369–72

Kuttner H, Urban G, Jachimowicz A, Kohl F, Olcaytug F, Goiser P 1991 Microminiaturized thermistor arrays for temperature gradient, flow and perfusion measurements. *Sens. Actuators A Phys.* **27(1–3)**, 653–6

Kwok C Y, Lin K M, Huang R S 1996 A silicon thermocapacitive flow sensor with frequency modulated output. *Sens. Actuators A (Phys.)* **A57(1)**, 35–9

Ladabaum I, Jin X, Soh H, Atalar A, Khuri-Yakub B 1998 Surface micromachined capacitive ultrasonic transducers. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **45(3)**, 678–90

Lai P T, Liu B, Zheng X, Li B, Zhang S, Wu Z 1997 Monolithic integrated spreading-resistance silicon flow sensor. *Sens. Actuators A Phys.* **58(1)**, 85–8

Lambert D K 1986 Thermal diffusion flow sensor. Tech. Rep. *US Pat*. 4 576 050, General Motors Corporation, Detroit, MI

Lambert D K, Harrington C R 1986 An air flow sensor based on interface thermal wave propagation. *J. Appl. Phys.* **59(1)**, 59–65

Lammerink T, Tas N, Elwenspoek M, Fluitman J 1993 Micro-liquid flow sensor. *Sens. Actuators A (Phys.)* **A37–A38**, 45–50

Lammerink T S, Dijkstra F, Houkes Z, van Kuijk J 1995 Intelligent gas-mixture flow sensor. *Sens. Actuators A Phys.* **47(1–3 pt 4)**, 380–4

Lammerink T, Tas N, Krijnen G, Elwenspoek M 2000 A new class of thermal flow sensors using Delta $T = 0$ as a control signal. *Proc. IEEE 13th Annu. Int. Conf. Micro Electro Mechanical Systems*, Miyavaki, Japan, pp. 525–30

Lammerink T, Tas N, van Honschoten J, Krijnen G, van Baar J, Elwenspoek M 2001 AC-driven temperature-balance flow sensor.*TRANSDUCERS '01. EUROSENSORS XV. 11th Int. Conf. Solid-State Sensors and Actuators. Dig. Tech. Pap.*, Munich, Germany, Vol. 2, pp. 1448–51

Lee G-B, Kuo T-Y, Wu W-Y 2002 A novel micromachined flow sensor using periodic flapping motion of a planar jet impinging on a V-shaped plate. *Exp. Therm. Fluid Sci.* **26(5)**, 435–44

Legtenberg R, Bouwstra S, Fluitman J H J 1991 Resonating microbridge mass flow sensor with low-temperature glass-bonded cap wafer. *Sens. Actuators A Phys.* **27(1–3)**, 737–40

Leung A, Jones J, Czyzewska E, Chen J, Pascal M 1997 Micromachined accelerometer with no proof mass. *International Electron Devices Meeting, 1997. Technical Digest*, Washington, DC, USA, pp. 899–902

Leung A, Jones J, Czyzewska E, Chen J, Woods B 1998 Micromachined accelerometer based on convection heat transfer. *Proc. 11th Annu. Int. Workshop Micro Electro Mechanical Systems 1998 (MEMS 98)*, Heidelberg, Germany, pp. 627–30

Li Y, Baek K, Gulari M, Lin D, Wise K 2005 A vacuum-isolated thermal microflowmeter for *in-vivo* drug delivery. *Sensors, 2005 IEEE*, Irvine, CA, USA, pp. 668–71

Lide D R (ed.) 2006 *CRC Handbook of Chemistry and Physics*, 87th edn. ,CRC Press, Boca Raton, FL

Lin L, Jones J 2005 A liquid-filled buoyancy-driven convective micromachined accelerometer. *J. Microelectromech. Syst.* **14(5)**, 1061–9

Lin K M, Kwok C, Huang R S 1996 Integrated thermo-capacitive type MOS flow sensor. *IEEE Electron. Device Lett.* **17(5)**, 247–9

Lin Q, Jiang F, Wang X-Q, Han Z, Tai Y-C, Lew J, Ho C-M 2000 MEMS thermal shear stress sensors: Experiments, theory and modeling. *Technical Digest. Solid-State Sensor and Actuator Workshop (TRF)*, Hilton Head, SC, USA, pp.304–7

Lin Q, Jiang F, Wang X, Xu Y, Han Z, Tai Y, Lew J, Ho C 2004 Experiments and simulations of MEMS thermal sensors for wall shear-stress measurements in aerodynamic control applications. *Journal of Micromechanics and Microengineering* **14(12)**, 1640–9

Liu B, Lai P, Xueren Z, Zhaohui W, Bin L 1997 A two-dimensional flow sensor using integrated silicon spreading-resistance temperature detectors. *Rev. Sci. Instrum.* **68(10)**, 3785–9

Liu C, Tai Y, Huang J, Ho C 1994 Surface micromachined thermal shear stress sensor. *Proceedings of the 1994 International Mechanical Engineering Congress and Exposition, American Society of Mechanical Engineers, Fluids Engineering Division (Publication) FED.* New York, NY, USA, Vol. 197, pp. 9–15

Liu C, Huang J, Zhu Z, Jiang F, Tung S, Tai Y, Ho C 1999 Micromachined flow shear-stress sensor based on thermal transfer principles. *Journal of Microelectromechanical Systems* **8(1)**, 90–9

Löfdahl L, Gad-el Hak M 1999a MEMS applications in turbulence and flow control. *Progr. Aerospace Sci.* **35(2)**, 101–203

Löfdahl L, Gad-el Hak M 1999b MEMS-based pressure and shear stress sensors for turbulent flows. *Meas. Sci. Technol.* **10(8)**, 665–86

Löfdahl L, Stemme G, Johansson B 1989a A sensor based on silicon technology for turbulence measurements. *J. Phys. E Sci. Instrum.* **22(6)**, 391–3

Löfdahl L, Stemme G, Johansson B 1989b Turbulence measurements using sensors based on silicon technology. ICIASF '89 Record. Int. Congr. Instrum. Aerospace Simul. Facilities, Göttingen, West Germany, pp.95–103

Löfdahl L, Stemme G, Johansson B 1992 Silicon based flow sensors used for mean velocity and turbulence measurements. *Exp. Fluids* **12(4–5)**, 270–6

Löfdahl L, Chernoray V, Haasl S, Stemme G, Sen M 2003 Characteristics of a hot-wire microsensor for time-dependent wall shear stress measurements. *Exp. Fluids* **35(3)**, 240–51

Luo X B, Yang Y J, Zheng F, Li Z X, Guo Z Y 2001 An optimized micromachined convective accelerometer with no proof mass. *J. Micromech. Microeng.* **11(5)**, 504–8

Luo X, Li Z, Guo Z, Yang Y 2002 Thermal optimization on micromachined convective accelerometer. *Heat Mass Transf./Waerme- und Stoffuebertragung* **38(7–8)**, 705–12

Lyons C, Friedberger A, Welser W, Muller G, Krotz G, Kassing R 1998 A high-speed mass flow sensor with heated silicon carbide bridges. *Proc. MEMS 98. IEEE. 11th Annu. Int. Workshop Micro Electro Mechanical Systems. An Investigation of Micro Structures, Sensors, Actuators, Machines and Systems*, Heidelberg, Germany, pp. 356–60

Ma Y, Ma S, Fang W, Wang T 1995 Air-flow sensor and humidity sensor application to neonatal infant respiration monitoring. *Sens. Actuators A Phys.* **49(1–2)**, 47–50

MacHattie L 1979 The transistor as an anemometer. *J. Phys. E (Sci. Instrum.)* **12(8)**, 754–60

Mailly F, Giani A, Bonnot R, Temple-Boyer P, Pascal-Delannoy F, Foucaran A, Boyer A 2001 Anemometer with hot platinum thin film. *Sens. Actuators A Phys.* **94(1–2)**, 32–8

Makinwa K, Huijsing J 2002a Constant power operation of a two-dimensional flow sensor. *IEEE Trans. Instrum. Meas.* **51(4)**, 840–4

Makinwa K A A, Huijsing J H 2002b A smart wind sensor using thermal sigma–delta modulation techniques. *Sens. Actuators A Phys.* **97–98**, 15–20

Makinwa K, Huijsing J, Hagedoorn A 2001 Industrial design of a solid-state wind sensor. *Sensor for Industry, 2001, Proceedings of the First ISA/IEEE Conference*, Rosemont, IL, USA, pp. 68–71

Malcovati P, Häberli A, Mayer F, Paul O, Maloberti F, Baltes H 1995 Combined air humidity and flow CMOS microsensor with on-chip 15 bit sigma–delta A/D interface. *IEEE Symp. VLSI Circuits, Dig. Tech. Pap.*, Kyoto, Japan pp.45–6

Marshall D C 1958 Measurement of sap flow in conifers by heat transport. *Plant Physiol.* **33(6)**, 385–96

Mastrangelo C, Muller R 1988 A constant-temperature gas flowmeter with a silicon micromachined package. *1988 Solid State Sensor and Actuator Workshop. Technical Digest*, Hilton Head, SC, USA, pp. 43–6

Mayer F, Paul O, Baltes H 1995 Influence of design geometry and packaging on the response of thermal CMOS flow sensors. *8th Int. Conf. Solid-State Sensors and Actuators and Eurosensors IX. Dig. Tech. Pap.*, Stockholm, Sweden, Vol. 1, pp. 528–31

Mayer F, Hintermann M, Jacobs H, Paul O, Baltes H 1996 Thermoelectric CMOS anemometers. *Proc. SPIE – Int. Soc. Opt. Eng.* **2882**, 236–46

Meijer G 1986 Thermal sensors based on transistors. *Sens. Actuators* **10(1–2)**, 103–25

Merkel T, Graeber M, Pagel L 1999 A new technology for fluidic microsystems based on PCB technology. *Sens. Actuators A (Phys.)* **A77(2)**, 98–105

Merkel T, Pagel L, Glock H-W 2000 Electric fields in fluidic channels and sensor applications with capacitance. *Sens. Actuators A (Phys.)* **A80(1)**, 1–7

Mierij Meteo B V 2006 Solid-state wind sensor MMW 005, product data sheet. Online resource, access date June 2006: http://www.mierijmeteo.nl/

Milanovic V, Bowen E, Tea N, Suehle J, Payne B, Zaghloul M, Gaitan M 1998 Convection-based accelerometer and tilt sensor implemented in standard CMOS. *Micro-Electro-Mechanical Systems (MEMS) – 1998. ASME International Mechanical Engineering Congress and Exposition*, Anaheim, CA, USA, pp. 487–90

Mireles J J 2005 Micromachined sensor design for optical-fiber flow measurement. *Sens. Rev.* **25(1)**, 33–9

Mizuno Y, Liger M, Tai Y-C 2004 Nanofluidic flowmeter using carbon sensing element. *17th IEEE Int. Conf. Micro Electro Mechanical Systems 2004 (MEMS)*, Maastricht, The Netherlands, pp. 322–5

Moser D, Baltes H 1993 A high sensitivity CMOS gas flow sensor on a thin dielectric membrane. *Sens. Actuators A Phys.* **37–38**, 33–7

Moser D, Lenggenhager R, Baltes H 1991 Silicon gas flow sensors using industrial CMOS and bipolar IC technology. *Sens. Actuators A Phys.* **27(1–3)**, 591–5

Moser D, Lenggenhager R, Wachutka G, Baltes H 1992 Fabrication and modelling of CMOS microbridge gas-flow sensors. *Sens. Actuators B (Chem.)* **B6(1–3)**, 165–9

Mouroulis P, Wilson D, Maker P, Muller R 1998 Convex grating types for concentric imaging spectrometers. *Appl. Opt.* **37(31)**, 7200–8

Nagata M, Stevens M, Swart N, Dravia T, Nathan A 2001 Optimization of two-element flow microsensors using quasi 3-D numerical electrothermal analysis. *Sens. Actuators A Phys.* **90(1–2)**, 102–10

Nakayama Y 1999 *Introduction to Fluid Mechanics*. Butterworth-Heinemann, Oxford

Nanver L, Goudena E, van Zeijl H 1993 Dimes-01, a baseline BIFET process for smart sensor experimentation. *Sens. Actuators A (Phys.)* **A36(2)**, 139–47

Naqwi A, Petrik S 1993 Fiber-optic dual-cylindrical wave sensor for measurement of wall velocity gradient in a fluid flow. *Appl. Opt.* **32(30)**, 6128–31

Naqwi A A, Reynolds W C 1987 Dual cylindrical wave laser Doppler method for measurement of skin friction in fluid flow. *Tech. Rep.*, Stanford University

Naqwi A, Reynolds W 1991 Measurement of turbulent wall velocity gradients using cylindrical waves of laser light. *Exp. Fluids* **10(5)**, 257–66

Naughton J W, Sheplak M 2002 Modern developments in shear-stress measurement. *Prog. Aerospace Sci.* **38(6–7)**, 515–70

Neda T, Nakamura K, Takumi T 1996 A polysilicon flow sensor for gas flow meters. *Sens. Actuators A Phys.* **54(1–3)**, 626–31

Ng K-Y, Shajii J, Schmidt M A 1991 A liquid shear-stress sensor fabricated using wafer bonding technology. Solid-State Sensors and Actuators 1991 Digest of Technical Papers, TRANSDUCERS '91, San Francisco, CA, USA, pp. 931–4

Nguyen N, Kiehnscherf R 1995 Low-cost silicon sensors for mass flow measurement of liquids and gases. *Sens. Actuators A (Phys.)* **A49(1–2)**, 17–20

Nguyen N, Meng A, Black J, White R 2000 Integrated flow sensor for *in situ* measurement and control of acoustic streaming in flexural plate wave micropumps. *Sens. Actuators A (Phys.)* **A79(2)**, 115–21

Nishimoto T, Shoji S, Esashi M 1994 Buried piezoresistive sensors by means of MeV ion implantation. *Sens. Actuators A (Phys.)* **A43(1–3)**, 249–53

Norlin P, Ohman O, Ekstrom B, Forssen L 1998 A chemical micro analysis system for the measurement of pressure, flow rate, temperature, conductivity, UV-absorption and fluorescence. *Sens. Actuators B (Chem.)* **B49(1–2)**, 34–9

Oda S, Anzai M, Uematsu S, Watanabe K 2003 A silicon micromachined flow sensor using thermopiles for heat transfer measurements. *IEEE Trans. Instrum. Meas.* **52(4)**, 1155–9

Ohnstein T, Johnson R, Higashi R, Burns D, Holmen J, Satren E, Johnson G, Bicking R, Johnson S 1990 Environmentally rugged, wide dynamic range microstructure airflow sensor. *Technical Digest. IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, pp. 158–60

Oosterbroek R, Lammerink T, Berenschot J, van den Berg A, Elwenspoek M 1997 Designing, realization and characterization of a novel capacitive pressure/flow sensor.*Tranducers 97. 1997 Int. Conf. Solid-State Sensors and Actuators. Dig. Tech. Pap.*, Chicago, IL, USA, Vol. 1, pp. 151–4

Oosterbroek R, Lammerink T, Berenschot J, Krijnen G, Elwenspoek M, van den Berg A 1999 A micromachined

pressure/flow-sensor. *Sens. Actuators A (Phys.)* **A77(3)**, 167–77
van Oudheusden B, Huijsing J 1988 Integrated flow friction sensor. *Sens. Actuators* **15(2)**, 135–44
van Oudheusden B, Huijsing J 1990 An electronic wind meter based on a silicon flow sensor. *Sens. Actuators A (Phys.)* **A22(1–3)**, 420–4
van Oudheusden B, de Bruijn J, Hoogeboom P, Beaufort D, Huijsing J 1989 Integrated sensor for non-invasive monitoring of flow in pipes. *Sens. Actuators* **18(3–4)**, 259–67
Ozaki Y, Ohyama T, Yasuda T, Shimoyama I 2000 An air flow sensor modeled on wind receptor hairs of insects. *Proc. IEEE 13th Annu. Int. Conf. Micro Electro Mechanical Systems*, Miyavaki, Japan, pp. 531–6
Padmanabhan A, Goldberg H, Breuer K, Schmidt M 1995 A silicon micromachined floating element shear-stress sensor with optical position sensing by photodiodes.*8th Int. Conf. Solid-State Sensors and Actuators and Eurosensors IX. Dig. Tech. Pap.*, Stockholm, Sweden, Vol. 2, pp. 436–9
Padmanabhan A, Sheplak M, Breuer K, Schmidt M 1997 Micromachined sensors for static and dynamic shear-stress measurements in aerodynamic flows.*Tranducers 97. 1997 Int. Conf. Solid-State Sensors and Actuators. Dig. Tech. Pap.*, Chicago, IL, USA, Vol. 1, pp. 137–40
Pagonis D N, Kaltsas G, Nassiopoulou A G 2004 Fabrication and testing of an integrated thermal flow sensor employing thermal isolation by a porous silicon membrane over an air cavity. *J. Micromech. Microeng.* **14(6)**, 793–7
Pan T, Hyman D, Mehregany M, Reshotko E, Willis B 1995a Calibration of microfabricated shear stress sensors.*Int. Conf. Solid-State Sensors and Actuators, Proc.*, Stockholm, Sweden, Vol. 2, pp. 443–6
Pan T, Hyman D, Mehregany M, Reshotko E, Willis B 1995b Characterization of microfabricated shear stress sensors. *ICIASF '95 Record. Int. Congr. Instrumentation in Aerospace Simulation Facilities*, Wright-Patterson AFB, OH, USA, pp. 6–1
Pan T, Hyman D, Mehregany M, Reshotko E, Garverick S 1999 Microfabricated shear stress sensors, part 1: Design and fabrication. *AIAA J.* **37(1)**, 66–72
Papageorgiou D, Bledsoe S, Gulari M, Hetke J, Anderson D, Wise K 2001 A shuttered probe with in-line flowmeters for chronic *in-vivo* drug delivery. 14th IEEE Int. Conf. Micro Electro Mechanical Systems 2001 (MEMS 2001), Interlaken, Switzerland, pp. 212–15
von Papen T, Steffes H, Ngo H, Obermeier E 2002 A micro surface fence probe for the application in flow reversal areas. *Sens. Actuators A Phys.* **97–98**, 264–70
von Papen T, Buder U, Ngo H, Obermeier E 2004 A second generation MEMS surface fence sensor for high resolution wall shear stress measurement. *Sens. Actuators A (Phys.)* **A113(2)**, 151–5
Park H, Pak J J, Son S Y, Lim G, Song I 2003a Fabrication of a microchannel integrated with inner sensors and the analysis of its laminar flow characteristics. *Sens. Actuators A Phys.* **103(3)**, 317–29
Park S, Kim S, Kim S, Kim Y 2003b A flow direction sensor fabricated using MEMS technology and its simple interface circuit. *Sens. Actuators B (Chem.)* **B91(1–3)**, 347–52
Perichon S, Roussel P, Lysenko V, Remaki B, Barbier D, Tritto J, Delhomme G, Dittmar A 2000 Micro-blood flow measurement using thermal conductivity micro-needles: A new CMOS compatible manufacturing process onto porous silicon. *Proc. 1st Annu. Int. IEEE-EMBS Spec. Top. Conf. Microtechnologies in Medicine and Biology*, Lyon, France, pp. 184–7
Persson B, Tartaglino U, Tosatti E, Ueba H 2004 Electronic friction and liquid-flow-induced voltage in nanotubes. *Phys. Rev. B (Condens. Matter Mater. Phys.)* **69(23)**, 235410–1
Pfann W, Thurston R 1961 Semiconducting stress transducers utilizing the transverse and shear piezoresistance effects. *J. Appl. Phys.* **32(10)**, 2008–19
Pijls N H, De Bruyne B, Sith L, Aarnoudse W, Barbato E, Bartunek J, Bech G J W, Frans V D V 2002 Coronary thermodilution to assess flow reserve: Validation in humans. *Circulation* **105**, 2482–6
Platzer G E J 1976 Solid state fluid flow sensor. Tech. Rep. *US Pat.* 3 992 940, Southfield, MI
Plöchinger H 1994 Neigungs- und beschleunigungs- sensor. Tech. Rep. *German Pat.* 40 34 962 C1
Poghossian A, Berndsen L, Schoening M 2003a Chemical sensor as physical sensor: ISFET-based flow-velocity, flow-direction and diffusion-coefficient sensor. *Sens. Actuators B (Chem.)* **B95(1–3)**, 384–90
Poghossian A, Schultze J, Schoening M 2003b Multi-parameter detection of (bio-)chemical and physical quantities using an identical transducer principle. *Sens. Actuators B Chem.* **91(1–3)**, 83–91
Polla D, Muller R, White R 1983 Monolithic integrated zinc-oxide on silicon pyroelectric anemometer. *International Electron Devices Meeting 1983. Technical Digest*, Washington, DC, USA, pp. 639–42
van Putten A 1983 An integrated silicon double bridge anemometer. *Sens. Actuators* **4(3)**, 387–96
van Putten A F P 1988 A constant voltage constant current Wheatstone bridge configuration. *Sens. Actuators* **13(2)**, 103–15
van Putten A, Middelhoek S 1974 Integrated silicon anemometer. *Electron. Lett.* **10(21)**, 425–6
van Putten M, van Putten M, van Putten A, Pompe J, Bruining H 1997 A silicon bidirectional flow sensor for measuring respiratory flow. *IEEE Trans. Biomed. Eng.* **44(2)**, 205–8
Qiu L, Hein S, Obermeier E, Schubert A 1996 Micro gas-flow sensor with integrated heat sink and flow guide. *Sens. Actuators A Phys.* **54(1–3)**, 547–51
Quandt E 1998 MEMS actuators based on smart film materials. *Proc. SPIE – Int. Soc. Opt. Eng.* **3514**, 136–46
Radhakrishnan S, Lal A 2005 Scalable microbeam flowsensors with electronic readout. *J. Microelectromech. Syst.* **14(5)**, 1013–22
Rahnamai H, Zemel J 1981 Pyroelectric anemometers: Preparation and flow velocity measurements. *Sens. Actuators* **2(1)**, 3–16
Rapoport S D, Reed M L 1991 Fabrication and testing of a microdynamic rotor for blood flow measurements. *J. Micromech. Microeng.* **1(1)**, 60–5
Rasmussen A, Gaitan M, Locascio L, Zaghloul M 2001 Fabrication techniques to realize CMOS-compatible microfluidic microchannels. *J. Microelectromech. Syst.* **10(2)**, 286–97
Rebiere D, Dejous C, Pistre J, Aucouturier J-L, Tiret C, Planade R 1994 Acoustic wave devices to measure gas flow: Comparison between surface acoustic wave (SAW) and shear horizontal acoustic plate mode (SH-APM) oscillators. *Sens. Actuators A Phys.* **42(1–3)**, 384–8
Richter A, Plettner A, Hofmann K, Sandmaier H 1991 Electrohydrodynamic pumping and flow measurement. *Proceedings. IEEE Micro Electro Mechanical Systems. An Investigation of Micro Structures, Sensors, Actuators, Machines and Robots*, Nara, Japan, pp. 271–6
Richter M, Wackerle M, Woias P, Hillerich B 1999 A novel flow sensor with high time resolution based on differential pressure principle. *Tech. Dig. IEEE Int. MEMS 99 Conf. 12th IEEE Int. Conf. Micro Electro Mechanical Systems*, Orlando, FL, USA, pp. 118–23
Robadey J, Paul O, Baltes H 1995 Two-dimensional integrated gas flow sensors by CMOS IC technology. *J. Micromech. Microeng.* **5(3)**, 243–50

Rodrigues R, Furlan R 2003 Design of microsensor for gases and liquids flow measurements. *Microelectron. J.* **34(5–8)**, 709–11

Rouhanizadeh M, Soundararajan G, Lo R, Arcas D, Browand F, Hsiai T 2006 MEMS sensors to resolve spatial variations in shear stress in a 3-d blood vessel bifurcation model. *IEEE Sens. J.* **6(1)**, 78–88

Sabate N, Santander J, Fonseca L, Gracia I, Cane C 2004 Multi-range silicon micromachined flow sensor. *Sens. Actuators A Phys.* **110(1–3)**, 282–8

Schmid U 2002 A robust flow sensor for high pressure automotive applications. *Sens. Actuators A Phys.* **97–98**, 253–63

Schmidt M, Howe R, Senturia S, Haritonidis J 1987 Fabrication and testing of a micromachined shear sensor. *1987 International Electron Devices Meeting, IEDM. Technical Digest*, Washington, DC, USA, pp. 282–5

Schmitz G J, Brücker C, Jacobs P 2005 Manufacture of high-aspect-ratio micro-hair sensor arrays. *J. Micromech. Microeng.* **15(10)**, 1904–10

Schoening M J, Poghossian A, Luth H, Schultze J 2001 Field-effect based multifunctional hybrid sensor module for the determination of both (bio-)chemical and physical parameters. *Proc. SPIE – Int. Soc. Opt. Eng.* **4576**, 149–59

Sheplak M, Chandrasekaran V, Cain A, Nishida T, Cattafesta L III 2002 Characterization of a silicon-micromachined thermal shear-stress sensor. *AIAA J. 39th Aerospace Sci. Meet.* **40(6)**, pp. 1099–104

Sheplak M, Cattafesta L, Nishida T 2004 MEMS shear stress sensors: Promise and progress. *24th AIAA Aerodynamic Measurement Technology and Ground Testing Conference*, Portland, OR, USA, June 28 – July 1, 2004

Shida K, Yuji J 1996 Discrimination of material property by pressure-conductive rubber sheet sensor with multi-sensing function.*Proc. IEEE Int. Symp. Industrial Electronics 1996 ISIE '96*, Warsaw, Poland, Vol. 1, pp. 54–9

Shin W C, Besser R S 2006 A micromachined thin-film gas flow sensor for microchemical reactors. *J. Micromech. Microeng.* **16(4)**, 731–41

Shoji S, Esashi M 1994 Micro flow devices. *Proc. Int. Symp. Micromechatronics and Human Science*, Nagoya, Japan, pp. 89–95

Sood A K, Ghosh S 2004 Direct generation of a voltage and current by gas flow over carbon nanotubes and semiconductors. *Phys. Rev. Lett.* **93(8)**, 086601.1

Sood A K, Ghosh S, Das A 2005 Flow-driven voltage generation in carbon nanotubes. *Pramana J. Phys.* **65(4)**, 571–9

Spalding D 1961 A single formula for the law of the wall. *Trans ASME Ser. A J. Appl. Mech.* **28(3)**, 444–58

Stemme G 1986 A monolithic gas flow sensor with polyimide as thermal insulator. *IEEE Trans. Electron. Devices* **ED-33(10)**, 1470–4

Stemme G 1988 A CMOS integrated silicon gas-flow sensor with pulse-modulated output. *Sens. Actuators* **14(3)**, 293–303

Stenberg M, Stemme G, Kittilsland G, Pedersen K 1988 A silicon sensor for measurement of liquid flow and thickness of fouling biofilms. *Sens. Actuators* **13(3)**, 203–21

Stephan C, Zanini M 1991 A micromachined, silicon mass-air-flow sensor for automotive applications. *1991 Int. Conf. Solid-State Sensors and Actuators 1991. Dig. Tech. Pap., TRANSDUCERS '91*, San Francisco, CA, USA, pp. 30–3

Steurer J, Kohl F 1998 Adaptive controlled thermal sensor for measuring gas flow. *Sens. Actuators A (Phys.)* **A65(2–3)**, 116–22

Su Y, Evans A, Brunnschweiler A 1996 Micromachined silicon cantilever paddles with piezoresistive readout for flow sensing. *J. Micromech. Microeng.* **6(1)**, 69–72

Su Y, Evans A, Brunnschweiler A, Ensell G 2002 Characterization of a highly sensitive ultrathin piezoresistive silicon cantilever probe and its application in gas flow velocity sensing. *J. Micromech. Microeng.* **12(6)**, 780–5

Sultan M F, Harrington C R, Hile J W 1993 Differential ac anemometer. *Tech. rep*, General Motors Corporation, Detroit, MI

Sundeen J, Buchanan R 2001 Thermal sensor properties of cermet resistor films on silicon substrates. *Sens. Actuators A (Phys.)* **A90(1–2)**, 118–24

Svedin N, Kalvesten E, Stemme E, Stemme G 1998a Lift-force flow sensor designed for acceleration insensitivity. *Sens. Actuators A Phys.* **68(1–3 pt 2)**, 263–8

Svedin N, Kalvesten E, Stemme E, Stemme G 1998b A new silicon gas-flow sensor based on lift force. *J. Microelectromech. Syst.* **7(3)**, 303–8

Svedin N, Kalvesten E, Stemme G 2003a A lift force sensor with integrated hot-chips for wide range flow measurements. *Sens. Actuators A (Phys.)* **A109(1–2)**, 120–30

Svedin N, Kälvesten E, Stemme G 2003b A new edge-detected lift force flow sensor. *J. Microelectromech. Syst.* **12(3)**, 344–54

Svedin N, Stemme E, Stemme G 2003c A static turbine flow meter with a micromachined silicon torque sensor. *J. Microelectromech. Syst.* **12(6)**, 937–46

Tabata O 1986 Fast-response silicon flow sensor with an on-chip fluid temperature sensing element. *IEEE Trans. Electron. Devices* **ED-33(3)**, 361–5

Tabata O, Inagaki H, Igarashi I 1987 Monolithic pressure-flow sensor. *IEEE Trans. Electron. Devices* **34(12)**, 2456–62

Tadigadapa S, Tsai C, Zhang Y, Najafi N 2002 Micromachined fluidic apparatus. Tech. Rep. *US Pat. 6 477 901*, Integrated Sensing Systems Inc. (US)

Tai Y-C, Muller R 1988 Lightly-doped polysilicon bridge as a flow meter. *Sens. Actuators* **15(1)**, 63–75

Takahata K, Gianchandani Y 2005 Micromachined intraluminal devices for active and passive electromagnetic measurements of flow. *18th IEEE Int. Conf. Micro Electro Mechanical Systems 2005 (MEMS 2005)*, pp. 863–6

Takahata K, DeHennis A, Wise K, Gianchandani Y 2004 A wireless microsensor for monitoring flow and pressure in a blood vessel utilizing a dual-inductor antenna stent and two pressure sensors. *17th IEEE Int. Conf. Micro Electro Mechanical Systems*, Maastricht, The Netherlands, pp. 216–19

Tan Z, Shikida M, Hirota M, Sato K 2006 Characteristics of on-wall in-tube thermal flexible mass-flow sensors. *19th IEEE Int. Conf. Micro Electro Mechanical Systems 2006 (MEMS 2006)*, Istanbul, Turkey, pp. 622–5

Tanase D, Goosen J F, Trimp P J, French P J 2002 Multi-parameter sensor system with intravascular navigation for catheter/guide wire application. *Sens. Actuators A Phys.* **97–98**, 116–24

Thürlemann B 1941 Methode zur elektrischen Geschwindigkeitsmessung in Flüssigkeiten. *Helv. Phys. Acta* **14(5–6)**, 383–419

Thürlemann B 1955 On the electromagnetic speed measurement of fluid. *Helv. Phys. Acta* **28**, 483

Toda K, Sanemasa I, Ishikawa K 1996 Simple temperature compensation of thermal air-flow sensor. *Sens. Actuators A Phys.* **57(3)**, 197–201

Tombler T W, Zhou C, Alexseyev L, Kong J, Dai H, Liu L, Jayanthi C S, Tang M, Wu S-Y 2000 Reversible electromechanical characteristics of carbon nanotubes under local-probe manipulation. *Nature* **405(6788)**, 769–72

Tong Q-Y, Huang J-B 1987 A novel CMOS flow sensor with constant chip temperature (CCT) operation. *Sens. Actuators* **12(1)**, 9–21

Trautweiler S, Paul O, Stahl J, Baltes H 1996 Anodically bonded silicon membranes for sealed and flush mounted microsensors. *IEEE, 9th Annu. Int. Workshop Micro Electro*

Mechanical Systems, 1996 (MEMS '96), Proceedings. An Investigation of Micro Structures, Sensors, Actuators, Machines and Systems, pp. 61–6

Tseng F-G, Lin C-J 2003 Polymer MEMS-based Fabry–Perot shear stress sensor. *IEEE Sens. J.* **3(6)**, 812–17

Tung S, Maines B, Jiang F, Tsao T 2004 Development of a MEMS-based control system for compressible flow separation. *J. Microelectromech. Syst.* **13(1)**, 91–9

Van Der Schoot B, Boillat M, De Rooij N 2001 Micro-instruments for life science research. *IEEE Trans. Instrum. Meas.* **50(6)**, 1538–42

Vass G 1997 Ultrasonic flowmeter basics. *Sensors* **14(10)**, 73–8

Verhoeven H J, Huijsing J H 1994 An integrated gas flow sensor with high sensitivity, low response time and a pulse-rate output. *Sens. Actuators A Phys.* **41(1–3)**, 217–20

Weiping Y, Chong L, Jianhua L, Lingzhi M, Defang N 2005 Thermal distribution microfluidic sensor based on silicon. *Sens. Actuators B Chem.* **108(1–2 Spec Iss)**, 943–6

Wenzel S, White R 1988 A multisensor employing an ultrasonic Lamb-wave oscillator. *IEEE Trans. Electron. Devices* **35(6)**, 735–43

van der Wiel A J, Linder C, de Rooij N F, Bezinge A 1993 A liquid velocity sensor based on the hot-wire principle. *Sens. Actuators A Phys.* **37–38**, 693–7

van der Wiel A, Boillat M, de Rooij N 1995 A bi-directional silicon orifice flow sensor characterised for fluid temperature and pressure. *8th Int. Conf. Solid-State Sensors and Actuators and Eurosensors IX. Dig. Tech. Pap.*, Stockholm, Sweden, Vol. 2, pp. 420–3

Wilson D, Maker P, Muller R 1996 Binary optic reflection grating for an imaging spectrometer. *Proc. SPIE – Int. Soc. Opt. Eng.* **2689**, 255–66

Wu J, Sansen W 2002 Electrochemical time of flight flow sensor. *Sens. Actuators A (Phys.)* **A97–98**, 68–74

Wu S, Lin Q, Yuen Y, Tai Y-C 2001 MEMS flow sensors for nano-fluidic applications. *Sens. Actuators A (Phys.)* **A89(1–2)**, 152–8

Xie J, Shih J, Tai Y-C 2003 Integrated surface-micromachined mass flow controller. *IEEE 16th Annu. Int. Conf. Micro Electro Mechanical Systems 2003 (MEMS-03)*, Kyoto, Japan, pp. 20–3

Xu Y, Lin Q, Lin G, Katragadda R, Jiang F, Tung S, Tai Y-C 2005a Micromachined thermal shear-stress sensor for underwater applications. *J. Microelectromech. Syst.* **14(5)**, 1023–30

Xu Y, Chiu C, Jiang F, Lin Q, Tai Y 2005b A MEMS multi-sensor chip for gas flow sensing. *Sensors and Actuators A: Physical* **121(1)**, 253–61

Yang C, Soeberg H 1992 Monolithic flow sensor for measuring millilitre per minute liquid flow. *Sens. Actuators A Phys.* **33(3)**, 143–53

Yoon E, Wise K 1988 Dielectrically-supported multi-element mass flow sensor. *Technical Digest – International Electron Devices Meeting*, San Francisco, CA, USA, pp. 670–3

Yoon E, Wise K 1992 An integrated mass flow sensor with on-chip CMOS interface circuitry. *IEEE Trans. Electron. Devices* **39(6)**, 1376–86

Yoon H J, Kim S Y, Lee S W, Yang S S 2000 Fabrication of a micro electromagnetic flow sensor for micro flow rate measurement. *Proc. SPIE – Int. Soc. Opt. Eng.* **3990**, 264–71

Yoshino T, Suzuki Y, Kasagi N, Kamiunten S 2001 Assessment of the wall shear stress measurement with arrayed micro hot-film sensors in a turbulent channel flow. *Proc. 2nd Int. Symp. Turbulence and Shear Flow Phenomena*, Stockholm, Sweden, Vol. 2, pp. 153–8

Yoshino T, Suzuki Y, Kasagi N, Kamiunten S 2003 Optimum design of micro thermal flow sensor and its evaluation in wall shear stress measurement. *Proc. IEEE Micro Electro Mechanical Systems (MEMS)*, Kyoto, Japan, pp. 193–6

Yu D, Hsieh H, Zemel J 1993 Microchannel pyroelectric anemometer. *Sens. Actuators A (Phys.)* **A39(1)**, 29–35

Zemel J N, Rahnamai H 1980 Pyroelectric anemometer. Tech. Rep. *US Pat*. 4 332 157,University of Pennsylvania (USA)

Zhang L, Ye X, Zhou Z, Yao J 1997 A micromachined single-crystal silicon flow sensor with a cantilever paddle. *1997 Int. Symp. Micromechanics and Human Science*, Nagoya, Japan, pp. 225–9

Zhang Y, Tadigadapa S, Najafi N 2001 A micromachined Coriolis-force-based mass flowmeter for direct mass flow and fluid density measurement.*TRANSDUCERS '01. EUROSENSORS XV. 11th Int. Conf. Solid-State Sensors and Actuators. Dig. Tech. Pap.*, San Francisco, CA, USA, Vol. 2, pp. 1460–3

Zhe J, Modi V, Farmer K R J 2005 A microfabricated wall shear-stress sensor with capacitative sensing. *J. Microelectromech. Syst.* **14(1)**, 167–75

Zou J, Chen J, Liu C, Schutt-Aine J 2001 Plastic deformation magnetic assembly (PDMA) of out-of-plane microstructures: Technology and application. *J. Microelectromech. Syst.* **10(2)**, 302–9

Biographies

Göran Stemme was born in 1958 in Stockholm, Sweden. He received the M.Sc. degree in Electrical Engineering in 1981 and the Ph.D. degree in Solid State Electronics in 1987, both from the Chalmers University of Technology, Gothenburg, Sweden. In 1981, he joined the Department of Solid State Electronics, Chalmers University of Technology, Gothenburg, Sweden. There, in 1990, he became an associate professor heading the silicon sensor research group. In 1991, Dr. Stemme was appointed Professor at The Royal Institute of Technology, Stockholm, Sweden, where he heads the Microsystem Technology group at the School of Electrical Engineering. His research is devoted to microsystem technology based on micromachining of silicon. His work spans over a broad range of technological and application fields such as medical technology, biochemistry, biotechnology, microfluidics, optical applications, wafer-level packaging, and device integration. Some of the results have successfully been commercialized. Dr. Stemme is a member of the Editorial Board of the IEEE/ASME *Journal of Microelectromechanical Systems* since 1997 and was a member of the Editorial Board of the Royal Society of Chemistry journal *Lab On A Chip* between 2000 and 2005. He has published more than 150 research journals and conference papers and has more than 12 patent proposals or granted patents. Dr. Stemme is a member of the Royal Swedish Academy of Sciences (KVA) and he is an IEEE fellow.

Sjoerd Haasl was born in 1976 in Leuven, Belgium. He received the M.Sc. degree with high honors in Electrotechnical Engineering, specialization Automation and Computer Systems at the University of Leuven, Belgium, in 1999. He received the Ph.D. degree in Microsystem Technology at the Royal Institute of Technology, Stockholm, Sweden, in 2005. At present, he is employed as Research Scientist and Project Leader at IMEGO AB, Gothenburg, Sweden. His current research interests lie in the fields of inertial sensors, flow sensors, and chemical sensors. Dr. Haasl is a member of the IEEE.

2.08 Gas Micropumps

Hanseup Kim, Khalil Najafi, and Luis P Bernal, University of Michigan, Ann Arbor, MI, USA

Glossary

CR Compression Ratio
ESP Electrostatic Ion Pump
LIGA An acronym from German words for lithography, electroplating, and molding
MEMS Microelectromechanical Systems
PDMS Polydimethylsiloxane
PVDF Polyvinylidene Flouride
PZT Lead Zirconate Titanate
SIP Sputter Ion Pump
VAMP Valve and Micropump

2.08.1 Introduction

A gas micropump is a small-volume device that can pump compressible gaseous fluids at different flow rates and pressures. The development of gas micropumps has been facilitated through the use of microelectromechanical systems (MEMS) technology that allows high-precision manufacturing of microdomain structures. Gas micropumps have several significant advantages, such as small size, low power consumption, and portability, that are highly desirable in many emerging applications, including health and environmental monitoring (Zellers *et al.* 2004, Kim *et al.* 2006c, Watson 1997), microelectronics cooling (Chou *et al.* 2002, Gromoll 1994, Jiang *et al.* 2002, Kim *et al.* 2006b, Rodgers *et al.* 2005, Tuckerman and Pease 1981, Zhang *et al.* 2002), micropropulsion (Chou *et al.* 2001, Kim *et al.* 2006b, Kwon *et al.* 2005, Wang *et al.* 2006, Yu and Kim 2003), and power generation (Arana *et al.* 2003, Seo and Cho 2004). They can fit in a small volume (<0.1 cc) and consume little power (<100 mW) (Kim *et al.* 2006a); in contrast,

conventional mesoscale pumping systems, such a gas tanks or mechanical pumps, take up much space, consume high power, and have limited precision to be used for controlling small doses of fluids in the microscale, as illustrated in **Figure 1**.

The development of gas micropumps has lagged behind liquid micropumps (Laser and Santiago 2004), primarily due to gas compressibility (Stehr *et al.* 1996), sealing challenges, and small actuation forces and amplitudes that are achievable in the microdomain. As of 2006, there have been only about more than 20 conference and journal papers reported in the literature on gas micropumps fabricated using MEMS technology (**Figure 2**). This is a small number compared with more than 200 archival papers published on liquid micropumps.

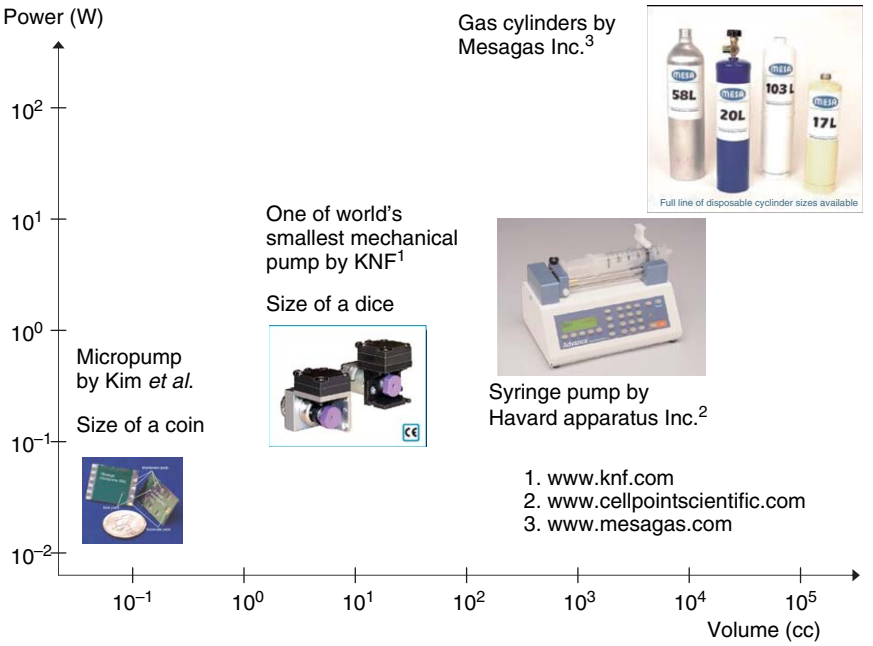


Figure 1 Gas pumping devices in various sizes and power consumption. (Sources: Kim H 2006 An integrated electrostatic peristaltic gas micropump with active microvalves. Ph.D. thesis, University of Michigan; KNF Inc. 2006 http://www.knf.com; Mesagas Inc. http://www.mesagas.com.) Also shown for comparison is a syringe pump. (Source: Cellpoint Scientific Inc. http://cellpointscientific.com).

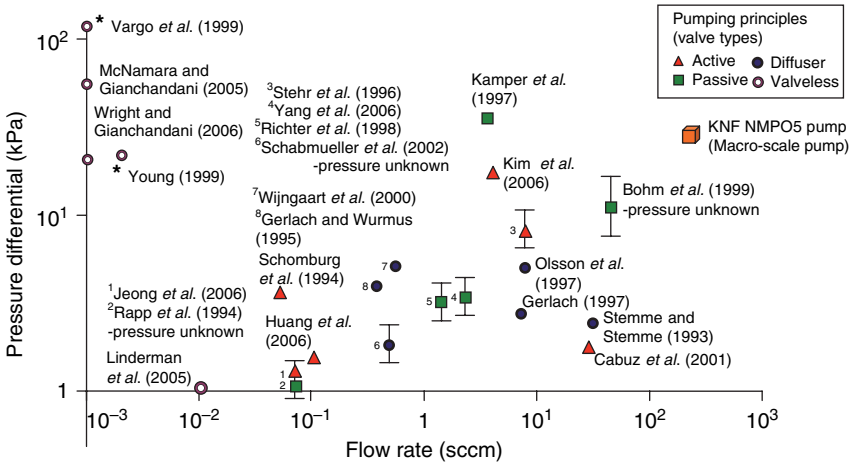


Figure 2 Flow rate and pressure differential of past gas micropumps with different pumping principles (asterisk indicates theoretical values). (Source: Kim H 2006 An integrated electrostatic peristaltic gas micropump with active microvalves. Ph.D. thesis, University of Michigan.)

The earliest MEMS-fabricated liquid micropumps were developed by van Lintel (van Lintel *et al.* 1988), then by Esashi (Esashi *et al.* 1989), and Smits (Smits 1989). The first MEMS gas micropumps were reported in the mid-1990s. The first reported micromachined pump that was targeted for gas pumping was the surface micromachined pump developed by Judy *et al.* (1991), but no flow data were available or reported in that paper. The first experimental measurements in a gas micropump were reported by Stemme in 1993 (Stemme and Stemme 1993). Since then, significant progress has been made in the development of gas micropumps due to better understanding of the relationship between flow resonances and pumping in the microdomain, and because of improved fabrication technologies and device structures. Stehr utilized the gas resonance effect to develop a bidirectional pump that produces reverse flow at high frequency (Stehr *et al.* 1996). Astle developed a theory to analyze the effect of gas compressibility and its use in a resonance-based micropump (Astle *et al.* 2007). Industry has also played an important role in the development of gas micropumps. Westinghouse developed a piezoelectric peristaltic gas micropump (Young *et al.* 1995), and Honeywell introduced a dual-membrane structure to minimize gas leakage and reported a high flow rate and efficient gas pumping (Cabuz *et al.* 2001). **Figure 2** shows the performance of past gas micropumps in terms of differential pressure generation and flow rate capabilities. It is noted that these parameters alone do not provide complete information on the performance of a particular micropump, because other performance metrics, such as size or power consumption that are not shown in the graph, are also important in many applications (**Table 1**).

This chapter will provide a systematic review of gas micropumps. Section 2.08.2 discusses the basic operation and structure of gas pumps and identifies important components, such as valves, that characterize gas micropumps. Section 2.08.3 provides a review of past gas micropumps based on four different pumping structures (valve types): passive, active, diffuser, and valveless pumps. Section 2.08.4 reviews gas micropumps based on their actuation techniques and fabrication technologies. Section 2.08.5 concludes the chapter with a design road map that summarizes all past efforts in developing gas micropumps based on some specific metrics and briefly discusses the remaining challenges and future directions in the development of gas micropumps.

2.08.2 Fundamental Characteristics

Gas micropumps have used different structures based on various working principles and have produced a wide range of performance specifications. This section discusses the general structure and important performance metrics, and design of gas micropumps.

2.08.2.1 General Pump Structure

A gas micropump can be simply described as a closed cavity with two fluidic paths, as illustrated in **Figure 3**. It consists of an inlet, an inlet control (usually an inlet valve), a main pumping cavity, an outlet control (outlet valve), and an outlet. The inlet is a passage through which gas is introduced into the main pumping cavity. Gas flow through the inlet into the main pumping cavity is often controlled by a valve. The inlet valve also minimizes flow loss in the unwanted reverse direction and plays an important role in defining pumping efficiency. The main pumping cavity is where gas goes through the pumping action to generate gas flow. In most cases, the pumping action involves a physical volume change of the cavity that is often performed by a compliant wall or membrane. In other cases, the pumping action involves a change in some properties of gas inside the cavity. For example, a change in temperature leads to thermal expansion in the volume, which in turn results in a higher pressure inside the cavity. By such pumping actions, gas is pumped out of the cavity and is directed to the outlet through the outlet valve. In the microdomain, the pumping action is generally limited due to the limited volume displacement or force.

Gas pumps operate based on the pressure change in the pumping cavity caused by gas compression or expansion. Because gases are compressible fluids, the inlet and outlet valves play an important role in regulating the pressure changes in gas pumps. The valves control the leak rate and limit the reverse flow through the pump, and adjust the valve timing that determines how much the gas is compressed before it is allowed to flow out of the main pumping chamber. Furthermore, the valves are an integral part of the pumping process because they can be used to modulate the phase between gas compression and flow cycles. In liquid pumps, where the liquid is incompressible, the primary role of the inlet and outlet valves is simply to minimize reverse flows in the

Table 1 Summary of past gas micropumps

	Pumping mechanism		*Pump structure*		*Fabrication*			*Pumping metrics*				
Reference(s)	*Pumping principle: valve type*	*Pumping driver (value)*	*Membrane material (size in mm^2)*	*Multistages (number)*	*Membrane*	*Chamber*	*Assembly*	*Flow rate (sccm)*	*Pressure (kPa)*	*Efficiency (%) (approximately)*	*Power (mW)*	*Membrane size[c] / (pack. volume) (mm^2, mm^3)*
Judy *et al*. (1991)	Active (peristaltic)	Electrostatic (75 V)	Silicon/nitride (0.16)	No	Surface micro-machining	Surface micro-machining	N/A	N/R	N/R	N/A	N/R	0.16
Stehr *et al*. (1996)	Active[a] (restricted)	Piezoelectric (100 V)	Silicon (28.09)	No	Surface micro-machining	N/A	Adhesive	8	N/R	1.44	N/R	28.09
Cabuz *et al*. (2001)	Active[b] (restricted)	Electrostatic (200 V)	Metalized Kapton (225)	No	Mechanical fixation	Injection molding	Mechanical fixation	30	2	1.85	8	225
Kim *et al*. (2006a)	Active	Electrostatic (±100 V)	Parylene (4)	Yes (18)	Membrane transfer	Bulk micro-machining	Wafer bonding (Parylene)	4	17.5	9.8	57	4
Jeong *et al*. (2006)	Active (peristaltic)	Pneumatic (62.5 kPa)	PDMS (0.785)	No	Membrane transfer	Molding	Wafer bonding (PDMS)	0.086	N/R	N/R	N/R	0.79
Huang *et al*. (2006)	Active (peristaltic)	Pneumatic (172 kPa)	PDMS (5.4[d])	No	Surface micro-machining	Molding	Wafer bonding (PDMS)	0.108	1.6	N/R	N/R	5.4
Rapp *et al*. (1994)	Passive	Pneumatic (50 kPa)	Titanium (10)	No	Membrane transfer	LIGA	Adhesive	0.086	N/R	N/R	N/R	10
Bustgens *et al*. (1994), Schomburg *et al*. (1994)	Passive	Thermopneumatic (15 V)	Polyimide (70)	No	Membrane transfer	Injection molding	Adhesive	0.044	3.8	N/R	450	70
Kamper *et al*. (1998)	Passive	Piezoelectric (N/R)	Polycarbornate (78.5)	No	Laser Welding	Injection molding	Laser welding	3.5	35	N/R	N/R	78.5
Richter *et al*. (1998)	Passive	Piezoelectric (N/R)	Silicon (N/R)	No	Bulk micro-machining	Bulk micro-machining	Wafer bonding (anodic)	1.4	N/R	N/R	N/R	N/R
Bohm *et al*. (1999)	Passive	Electromagnetic (100 mA)	Mylar (100)	No	Adhesive	Injection molding	Adhesive	40	N/R	N/R	N/R	100

Yang *et al*. (2006)	Passive	Piezoelectric (20 V)	Stainless steel (960)	No	N/R	Precision machining	Mechanical fixation	4.5	N/R	N/R	3.18	960
Stemme and Stemme (1993)	Diffuser	Piezoelectric (20 V)	Brass (283.4)	No	Mechanical fixation	N/R	N/R	35	2.3	N/R	N/R	288.39
Gerlach and Wurmus (1995)	Diffuser	Piezoelectric (N/R)	Pyrex glass (49)	No	Bulk micro-machining	Bulk micro-machining	Adhesive	0.4	3.9	N/R	N/R	49
Gerlach (1997)	Diffuser	Piezoelectric (150 V)	Pyrex glass (121)	No	Bulk micro-machining	Bulk micro-machining	Wafer bonding (anodic)	7.5	2.8	N/R	N/R	121
Olsson *et al*. (1997)	Diffuser	Piezoelectric (24 V)	N/R (113.04)	No	Mechanical fixation	Precision machining	Mechanical fixation	8	5	N/R	N/R	113.04
Wijngaart *et al*. (2000)	Diffuser	Piezoelectric (200 V)	Silicon (16)	No	Bulk micro-machining	Bulk micro-machining	Wafer bonding (anodic)	0.545	5	N/R	N/R	16
Schabmueller *et al*. (2002)	Diffuser	Piezoelectric (300 V)	Silicon (49)	No	Bulk micro-machining	Bulk micro-machining	Adhesive	0.69	N/R	N/R	N/R	49
Young (1999)	Valveless (open passage)	Knudsen (N/R)	N/A	Yes (26[a])	N/A	N/A	N/A	0.002[d]	19.9[d]	N/A	1800[d]	N/R
McNamara and Gianchandani (2005)	Valveless (open passage)	Knudsen	N/A	No	N/A	Bulk micro-machining	Wafer bonding (anodic)	1×10^{-6}	46.6	N/A	80	0.4
Wright and Gianchandani (2006)	No valve (closed cavity)	Sputter ion	N/A	No	N/A	Bulk micro-machining	Wafer bonding (solder)	0	22.4	N/A	N/R	1419 (6330)
Vargo *et al*. (1999)	Valveless (open passage)	Knudsen	N/A	No	N/A	N/A	N/A	N/A	100[d]	N/A	2400[d]	N/R
Linderman *et al*. (2005)	No valve (open passage)	Electrostatic fan pump	Polysilicon (0.0918)	No	Membrane transfer	N/A	N/A	0.01	N/A	N/A	N/R	0.09

[a]Pumping is controlled by a single membrane shared by a valve and pumping membrane.
[b]Pumping is controlled by two (dual) membranes containing holes at different locations respectively and by their certain order of movements.
[c]The areas and volumes of a pumping membrane are approximate values.
[d]Values are theoretical calculations.
PDMS, polydimethylsiloxane.

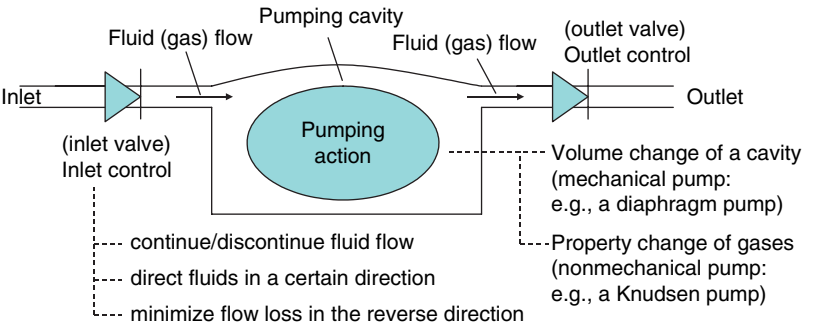


Figure 3 The general structure of a gas micropump. (Source: Kim H 2006 An integrated electrostatic peristaltic gas micropump with active microvalves. Ph.D. thesis, University of Michigan.)

pump. In these liquid pumps, the pumping process is mainly controlled by the pumping action in the cavity.

2.08.2.2 Performance Metrics

Before proceeding further, it is informative to define a set of performance metrics for gas micropumps. These metrics are useful in comparing different pumping mechanisms and fabrication technologies. The most important performance metrics are summarized below:

- Pressure differential
- Flow rate
- Pumping efficiency
- Size
- Power consumption
- Cost

Pressure differential is the pressure difference created between the inlet and the outlet and indicates how much pressure a micropump can work against. Flow rate is the volume (or mass) of gas that is transferred per unit time from the inlet to the outlet. Generally, pressure differential and flow rate are good indicators of the quality of the pump. Pumping efficiency is defined as the ratio between the ideal stroke volume per unit time in a diaphragm pump and the real flow rate. It indicates how much of the compressed gas is actually pumped out. In gas pumps, pumping efficiency is typically less than 20% because gases are compressible; in liquid pumps, most of the compressed liquid can be pumped out. Device size, power consumption, and cost are all obvious but important parameters and all need to be minimized. Among the five performance metrics, the three most critical metrics are (1) pressure differential, (2) flow rate, and (3) pumping efficiency. Figure 4 shows typical performance curves for a gas micropump illustrating these features.

2.08.2.2.1 Pressure differential

One distinctive feature of a gas micropump is the maximum pressure differential or the maximum back pressure it can produce. Figure 4 illustrates a typical performance curve of a gas pump, which shows that the pressure differential created by the pump decreases as the pump flow rate is increased. The maximum back pressure is the pressure at which the pump flow rate is zero. It indicates the maximum pressure that the micropump can work against while avoiding negative flow rates.

In a gas micropump, thermodynamic considerations show that the maximum pressure differential is determined by the compression ratio (CR) of the pump, as described in Figure 5. Here the compression ratio is defined as the change in volume divided by the original volume of the pumping chamber

$$\mathrm{CR} = \frac{\Delta p}{p_{\mathrm{in}}} = \frac{\Delta V}{V_1} \qquad [1]$$

where Δp is the maximum back pressure, p_{in} is the absolute inlet pressure, ΔV is the volume change of the pumping chamber, and V_1 is the pumping cavity volume. Here isothermal compression is assumed. For example, if the pump is compressed by 20% of its original volume, the pressure inside the chamber increases by 20%. Thus, a high pressure differential can be achieved by obtaining a high compression ratio, which could result from either by directly compressing the gas by moving a pumping membrane or by expanding the volume of the gas thermally in the pumping cavity. The pump by Rapp *et al.* (1994) employs a flexible polyimide membrane to create a large deflection to compress the

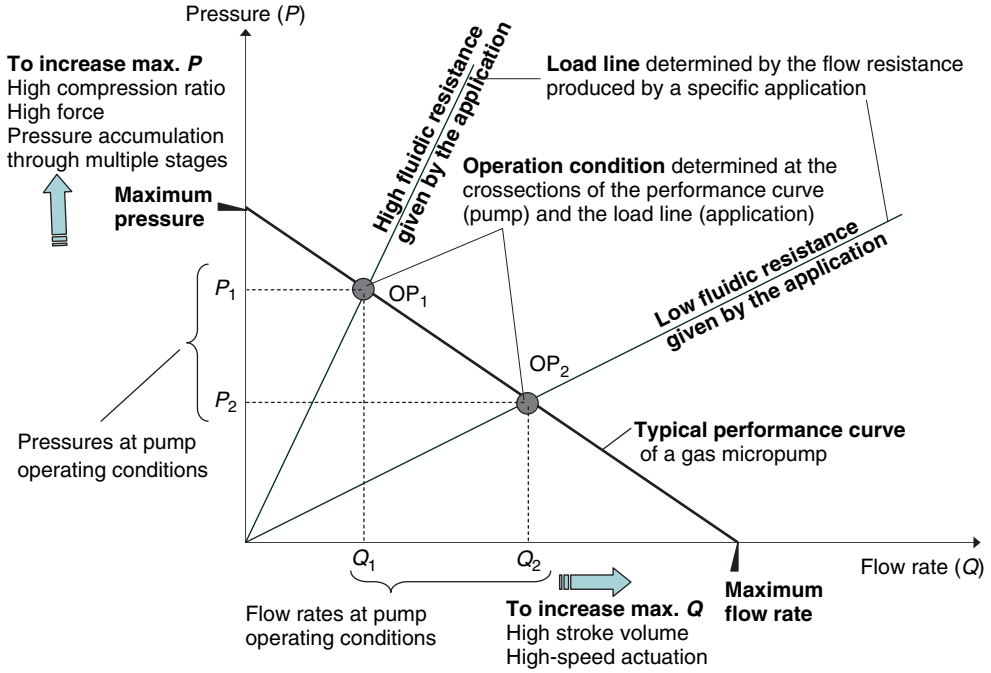


Figure 4 Typical performance curves including important performance metrics (pressure differential, flow rate, and operation modes) observed in a gas micropump.

pumping chamber with a high compression ratio, whereas the pump reported by Cabuz *et al.* (2001) produces only a small pressure differential of 2 kPa because it transfers gas without compression cycles. Other researchers have taken advantage of the high thermal expansion of gas to achieve higher effective compression ratio than typical diaphragm pumps (McNamara and Gianchandani 2005, Vargo *et al.* 1999, Young 1999). In practice, the compression ratio is lower than predicted by theory due to imperfect valves and leakage.

Pressure differential can be accumulated using multiple pumping chambers connected in series, each with a certain compression ratio. **Figure 5** shows the increase in pressure when multiple pumping chambers are used. If the compression ratio is large (CR = 0.5), the pressure differential in the last few stages is much larger than that in the first few stages. However, when the compression ratio is small (CR = 0.02), the differential pressures across all the stages are approximately equal (Astle *et al.* 2002, 2007). This feature of the multiple-chamber configuration is very attractive for MEMS micropumps because it allows to distribute the required pumping actuation force uniformly across many pumping chambers. A high overall pressure differential can be achieved while limiting the compression ratio and the pressure differential for each stage. For example, Kim *et al.* cascaded several micropumps and reported the first functional peristaltic configuration achieving a high pressure differential (>17 kPa) for the whole pump. The compression ratio for each stage in this pump was less than 5% (Kim *et al.* 2006a).

2.08.2.2.2 Flow rate

Another important feature of a gas micropump is the maximum gas flow rate, which represents the maximum volume/mass of gas transferred per unit time, often defined in standard cubic centimeter per minute (sccm) or milliliters per minute. As discussed earlier the maximum gas flow rate for a typical gas pump is achieved when the pressure differential is zero.

Thermodynamic analysis shows that the maximum gas flow rate, Q_{max}, is determined by the product of the volume displacement of a moving membrane, its frequency of oscillation, and the pumping efficiency, as shown in eqn [2], where ΔV indicates the volume displacement per stroke, f the actuation frequency, and α the pumping efficiency. In gas micropumps that do not utilize moving parts such as a moving membrane, flow rate is typically small because the effective volume displacement is small. In diaphragm-based gas micropumps, a high

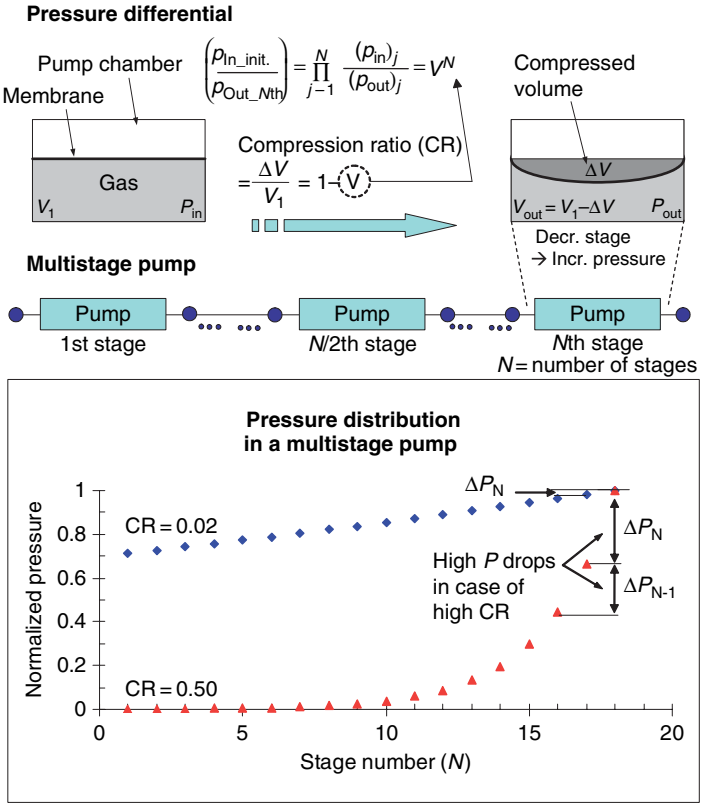


Figure 5 Pressure differential is determined by the compression ratio of each stage. In a multistage pump, the compression ratio is low for each stage, but a high total pressure differential is achieved. (Source: Astle A, Paige A, Bernal L P, Munfakh J, Kim H, Najafi K 2002 Analysis and design of multistage electrostatically-actuated micro vacuum pumps. *Proc. ASME Int. Mechanical Engineering Congress and Exposition (ASME '02).* New Orleans, LA, USA, IMECE 2002-39308.)

gas flow rate can be achieved by operating the micropump at a high frequency.

$$Q_{max} = \Delta V_{pump\_membrane} * f * \alpha \qquad [2]$$

Typically, the flow rate in gas pumps has a peak at a certain operating frequency, as shown in **Figure 6**,

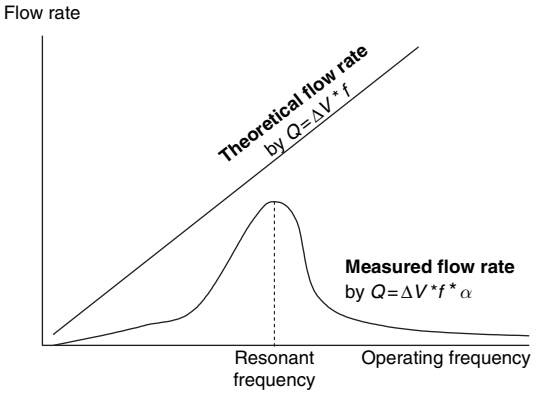


Figure 6 Output flow rate shows a resonance behavior caused by the compressibility of gas.

which is not often observed in liquid pumps. Flow rate increases with frequency and peaks at the resonant frequency, showing the maximum gas transfer, and then decreases as frequency increases despite the higher volume displacement per unit time. Such a resonant behavior in gas flow could be the result of gas dynamics resonance, structural resonance, or actuator resonance. Muller *et al.* provided an insightful analysis of gas flow resonance effects in the microdomain for a pumping membrane operated at a high frequency based on Helmholtz theory (Eckart 1948, Filippi *et al.* 1999, Muller *et al.* 2002, Rayleigh 1916). This analysis shows that an optimum frequency exists where gas transfer is maximized given by

$$f_o = \frac{a}{2\pi}\sqrt{\frac{A_t}{L_t V_c}} \qquad [3]$$

where a is the speed of sound, A_t the effective cross-sectional area, L_t the effective length of small flow passages in the pump, and V_c the pumping cavity

volume. At the optimum frequency, the flow rate is a maximum. At higher frequencies, the flow rate decreases rapidly because the phase shift between membrane compression and gas motion leads to inefficient pumping. It is worth noting that the fluidic resonance and the associated highest flow rate occur at a high frequency in micropumps due to their small dimensions, as shown in eqn [3].

The flow rate in a gas micropump can be improved by increasing the volume displacement (ΔV), operating frequency (f) of the pumping membrane, or pumping efficiency (α). Volume displacement is maximized by compressing gas by using high-force and large-deflection actuation mechanisms and by making the membrane using high-elongation flexible materials, such as polymers. The operating frequency, often the resonance frequency of the device, as shown in eqn [3], is optimized through proper scaling of the pump geometry. It follows that the high operation frequency compensates for the small volume displacement in the microdomain gas pumps. Pumping efficiency is enhanced by operating the gas pump at the fluidic resonance where gas volume transfer becomes highly efficient and by minimizing gas leakage by using appropriate microvalve structures, as explained in detail in the next section.

2.08.2.2.3 Pumping efficiency

Pumping efficiency (α) is defined as the ratio of the actual gas flow rate to the ideal gas flow rate. The ideal gas flow rate can be calculated from the product of stroke volume of the membrane and the operating frequency, as described earlier in eqn [2]. Pumping efficiency is typically lower than 20% even at fluidic resonance and is determined by gas leakage through imperfect microvalves and by the volume transfer efficiency of the compressible gas caused by the phase difference between the actual membrane deflection and microvalve timing.

Pumping efficiency is influenced by different microvalve designs used in micropumps: passive valves, diffuser valves, and active valves. Generally, pumping with passive valves provides the highest pumping efficiency, whereas diffuser-based micropumps provide the lowest pumping efficiency of less than a few percent. Active pumps provide a pumping efficiency between those of passive pumps and diffuser pumps. For example, a pumping efficiency of $\geq$10% can be estimated from the parameters reported in the passive valve pump reported by Rapp *et al.* (1994) where a flow rate of 0.086 sccm for a 5-mm diameter membrane was reported at an operating frequency of 5 Hz. The diffuser pump reported by Stehr *et al.* (1996) provided a measured flow rate of 8.0 sccm for a 5.3-mm membrane deflected over $\sim$20 μm at a frequency of 16.5 kHz, which results in an estimated efficiency of 1.44%. Active pumps have demonstrated pumping efficiencies of 2.2% and 9.8%, respectively, in the dual-membrane pump by Cabuz (Cabuz *et al.* 2001) and the 18-stage pump by Kim (Kim *et al.* 2006a). While Cabuz's dual-membrane pump provided a flow rate of 30 sccm using a 15-mm diameter membrane at an operation frequency of 100 Hz, Kim's pump demonstrated a flow rate of 4 sccm using a 2-mm square membrane at an operation frequency of 17.5 kHz.

2.08.3 Review of Pumping Techniques

Pumping techniques play an important role in producing the desired pressure differential, flow rate, and pumping efficiency. They include passive, diffuser, active, and valveless pumping. In this section, we briefly review several types of common pumping techniques used in micropumps, with an emphasis on their operation principle, basic structure, and advantages and disadvantages. **Figure 7** summarizes the types of micropumps and categorizes them by their pumping technique and actuation mechanism. This latter categorization will be discussed in more detail in Section 2.08.4.

2.08.3.1 Passive Pumping

Passive pumping is the most basic and traditional pumping technique that has been widely used in most macroscale pumps and some microscale pumps. It restricts gas flow passively by using the pressure changes in the pumping chamber. The pressure change is caused by the pumping action primarily through the membrane motion. Passive pumping typically uses a unidirectional normally-closed valve located over the inlet and the outlet paths. The valve remains closed until the pressure difference across it becomes sufficiently large to force it open and to allow gas to flow (Zengerle *et al.* 1995), as illustrated in **Figure 8**. Because the valve normally closes the flow path, passive pumping minimizes reverse flows or leakage, which results in a high pumping efficiency. It is notable that gas can be pumped only in one direction because the valve opens only in one direction. Passive pumping does not work at high frequencies because it is governed by the geometry and mechanical

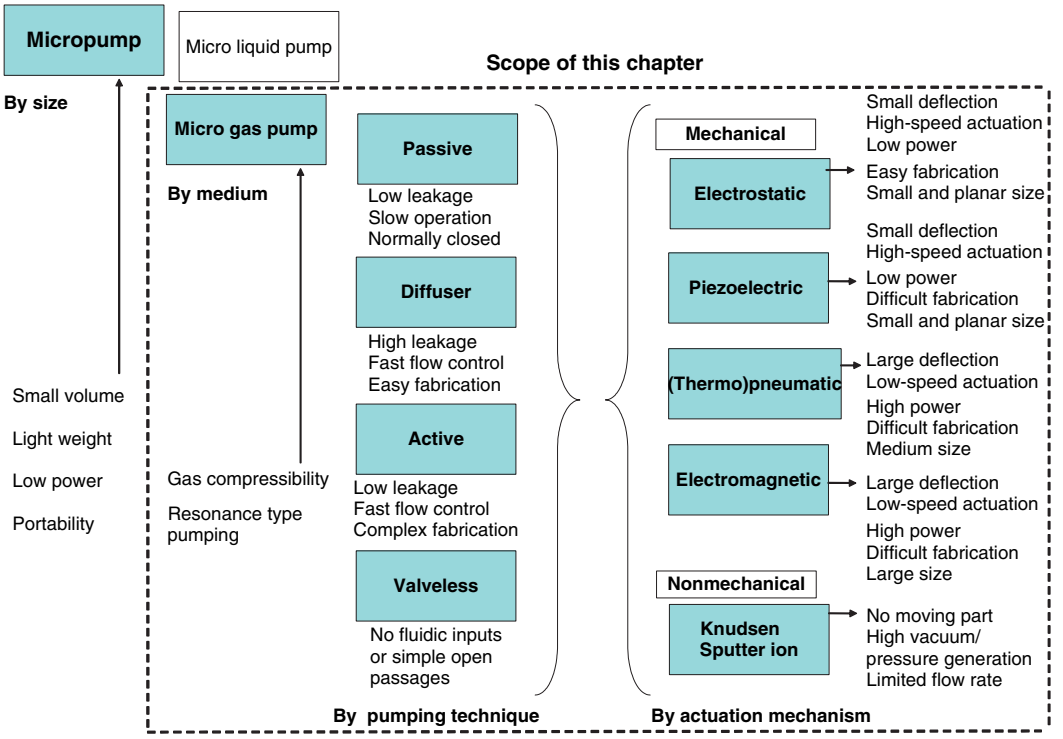


Figure 7 Gas micropumps categorized by pumping technique (valve types) and actuation mechanism. (Source: Kim H 2006 An integrated electrostatic peristaltic gas micropump with active microvalves. Ph.D. thesis, University of Michigan.)

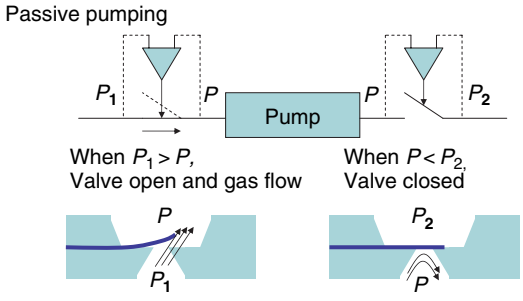


Figure 8 Simplified diagram for a passive pump (*top*) and operation principle of a passive valve (*bottom*).

characteristics of the valve, which limit the response time to the changes of pressure at high frequencies.

Passive pumps have used two different valve configurations: a single cantilever valve and a dual-diaphragm check valve, as shown in **Figure 9**. The cantilever valve, the most common valve used in passive pumping, is deflected to open or close the fluidic path depending on the pressure change across it. Using the cantilever valve Richter investigated the relationship between the required gas compression ratio to actuate the passive valve and the resultant flow rate from passive pumping (Richter *et al.* 1998). In his work, he formed a cantilever valve made of silicon under pumping cavities with various heights from the pumping membrane in piezoelectric and electrostatic gas micropumps. Passive pumping using the cantilever valve produced higher flow rates at higher compression ratio of above ~2% but some flow was observed when compression ratio was higher than 1.7%. Additionally, a gas flow rate of up to 1.4 sccm was measured when the compression ratio was 8.5% and the membrane was operating at 300 Hz. For efficient passive pumping, the cantilever valve has typically been used together with electromagnetic (Bohm *et al.* 1999) and thermopneumatic actuation (Bustgens *et al.* 1994, Schomburg *et al.* 1994). These actuation techniques produce a larger deflection and thus a higher compression than piezoelectric and electrostatic actuation. Bohm developed an electromagnetic pump that deflects the membrane using a high electromagnetic force of 0.11 N. To further facilitate the motion of the valve, the cantilever valve was made of a flexible polymer, Mylar, with a thickness of 7 μm. This passive pump produced a measured gas flow rate of 40 sccm, while consuming a power of 0.5 W (5 V and 100 mA) in a packaged volume of 10 mm × 10 mm × 10 mm.

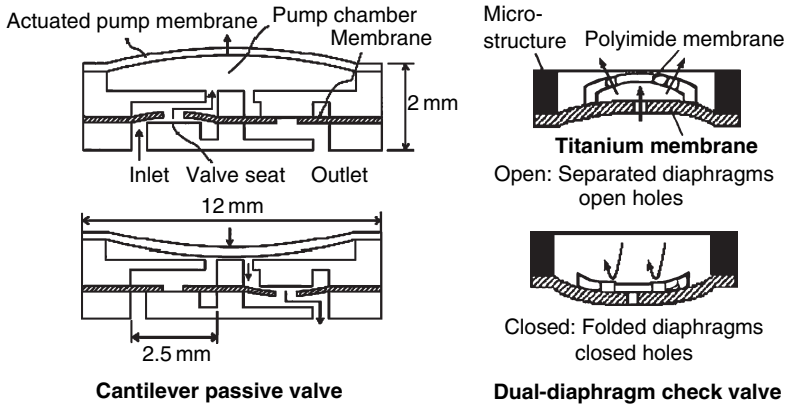


Figure 9 *Left*: Typical cantilever passive valve in an electromagnetic pump. (Source: Bohm S, Olthuis W, Bergveld P 1999 A plastic micropump constructed with conventional techniques and materials. *Sens. Actuators A* **77**, 223–8.) *Right*: Dual-diaphragm check valve. (Source: Rapp R, Schomburg W K, Maas D, Schulz J, Stark W 1994 LIGA micropump for gases and liquids. *Sens. Actuators A* **40**, 57–61.)

Bustgens used thermopneumatic actuation to achieve a high compression ratio of 10% by deflecting a 70-mm^2 membrane over a 100-μm distance. This thermopneumatic passive pump produced a flow rate of 0.044 sccm and a pressure of 3.8 kPa at a pumping efficiency of ~2.1%.

The dual-diaphragm check valve, the second valve design used in passive pumps, consisted of two separate membranes that were mounted on top of each other with valve holes staggered relative to each other, as shown in **Figure 9** (*right*). Because the top membrane was made of a flexible polyimide film, it could be moved above the bottom titanium membrane, thus opening the holes to allow gas flow when gas was pumped out of the pumping chamber in the forward direction. The top polyimide membrane collapsed down to the bottom titanium membrane closing the holes in the reverse direction, thus preventing any gas flow (Rapp *et al.* 1994). Rapp reported a high pumping efficiency of 14.6% using the passive check valve using a high-force (>50 kPa of external pressure) pneumatic action. This passive pumping produced a flow rate of 0.086 sccm and a pressure of 4.7 kPa at a frequency of 10 Hz. Rapp also reported that when the passive check valve is larger in dimension or thinner in thickness, it reacts in a more responsive way by allowing higher gas flows under some specific conditions.

2.08.3.2 Diffuser Pumping

Diffuser pumping was first used by Stemme (Stemme and Stemme 1993) in the microdomain to take advantage of the high-frequency stroke of a piezoelectric pumping membrane to provide a high stroke volume per unit time (high flow rate). It directs gas flow using different flow resistance in the forward and the reverse directions through tapered inlet and outlet passages, as shown in **Figure 10**. Its tapered passages work as a diffuser valve where the flow resistance is much smaller in the forward direction than in the reverse direction. The basic operation of the pump is quite straightforward, as illustrated in **Figure 11**. In the supply cycle, where gas is pumped into the chamber

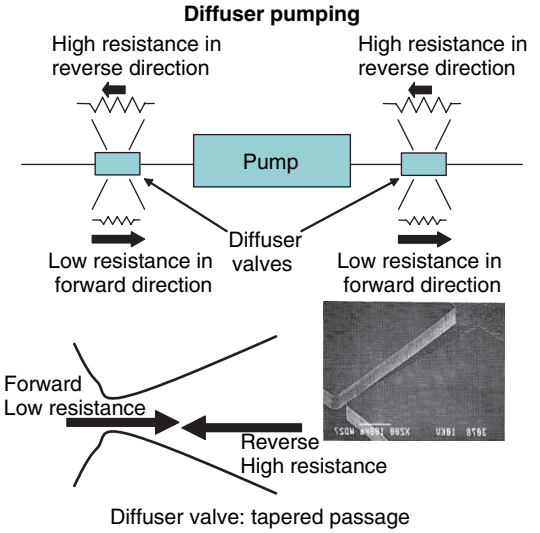


Figure 10 Basic structure of a diffuser pump (*top*) and operation principle of a diffuser valve and a SEM of a fabricated diffuser valve (*bottom*). (Source: Olsson A, Enoksson P, Stemme G, Stemme E 1996 An improved valve-less pump fabricated using deep reactive ion etching. *Proc. 9th Int. Conf. Micro Electro Mechanical Systems (MEMS '96)*, San Diego, CA, USA, pp. 479–84.)

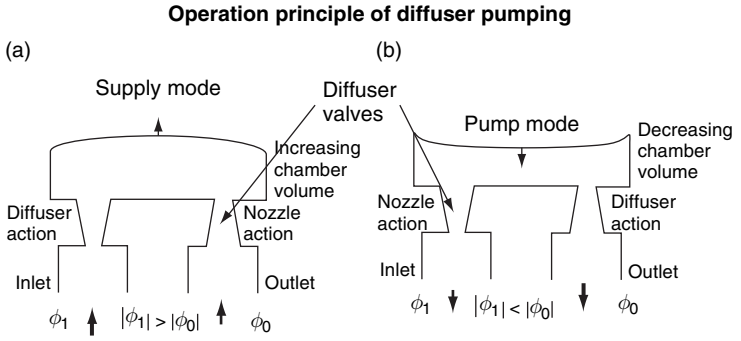


Figure 11 Operation principle of diffuser pump. (Source: Stemme E, Stemme G 1993 Valveless diffuser/nozzle-based fluid pump. *Sens. Actuators A* **A39(2)**, 159–67.)

from the inlet, the pumping membrane moves up, thus expanding the chamber volume. Because flow resistance is smaller from outside to inside for the inlet diffuser than for the outlet diffuser, more gas flows into the chamber from the inlet than through the outlet. In the pump cycle, the membrane is deflected downward, thus compressing the pumping chamber. Because flow resistance is smaller from inside to outside through the outlet diffuser than through the inlet diffuser, more gas is pumped out through the outlet than through the inlet. Therefore, during a complete pumping cycle, there is a net flow of gas from the inlet to the outlet. The theory for deriving pumping flow rate and pressure as well as efficiency has been developed by several investigators. While Stemme developed a theory for generic diffuser pumping for both gas and liquid, Olsson investigated a theory for gas pumping considering the compressibility of gases through an analogy to electrical circuits (Olsson *et al.* 1996). Olsson *et al.* derived a two-resonance frequency equation where the higher resonance frequency represents the operation frequency for gas pumping, whereas the lower resonance frequency represents that for liquid pumping.

Because diffuser pumping does not require moving valves, it responds to gas pressure changes at high frequency and allows high-speed pumping. In contrast, in passive pumping the frequency of operation is limited by the material property of the moving valve and its resonant frequency. Additionally, diffuser pumping does not suffer from wear or fatigue. However, diffuser pumping has a relatively low pumping efficiency of less than a few percent because it has flow losses in the unwanted direction through the always-open passages. It is notable that diffuser pumping has been realized only in piezoelectrically actuated pumps.

Stemme reported a flow rate of 35 sccm and a pressure of 2.3 kPa in his earliest diffuser pump, while Olsson generated a flow rate of 8 sccm and a pressure of 5 kPa in his diffuser pump designed only for gas pumping. Olsson's diffuser pump operated a 12-mm diameter membrane at a frequency of 11 kHz with a pumping efficiency of ~0.36%. Gerlach investigated a 3D diffuser (out-of-plane) and investigated the dynamic changes in fluid resistance at high flux velocity (Gerlach 1997, Gerlach and Wurmus 1995), as shown in **Figure 12**. The 3D diffuser pumping solved a

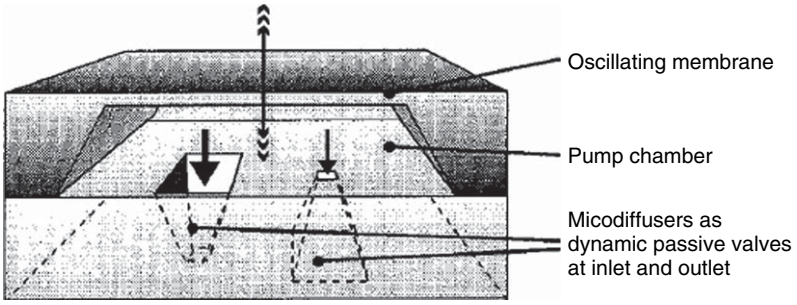


Figure 12 The structure of a 3D diffuser pump. (Source: Gerlach T 1997 Pumping gases by a silicon micro pump with dynamic passive valves. *Proc. Int. Conf. Solid-State Sensors and Actuators (Transducers '03)*, Chicago, IL, USA, pp. 357–60.)

limitation in the typical planar diffusers by providing a large area to obtain necessary flow rectification capability. Gerlach's theoretical model precisely predicted the passive pumping flow rate of 7.8 sccm from a microfabricated prototype within 4% error (measured flow rate 7.5 sccm), while it over-predicted the gas pressure differential of 4.8 kPa by 60% (measured pressure differential 2.8 kPa). The 3D diffuser pumps actuated a 11-mm diameter membrane at a frequency of 12.1 kHz and the resulting pumping efficiency was ~0.56%. Later, a high operation frequency of up to 33 kHz was reported by Wijngaart (Wijngaart *et al.* 2002). This diffuser pump actuated 4 mm × 4 mm membrane producing a flow rate of 0.545 sccm and a pressure of 5 kPa with a pumping efficiency of ~0.86%.

2.08.3.3 Active Pumping

Active pumping is intended to combine the advantages of passive and diffuser pumping by compressing gas in a confined chamber for high and efficient pressure rise and releasing it only during a certain period for minimized reverse flow. For such operation, active pumping controls individual valves independently using external signals, as shown in **Figure 13**. Active pumping controls the timing and the duration of opening and closing of microvalves to reduce flow leakage and operate the pump at high frequency, and create various output gas flows. Active pumping generally requires a more complex structure and fabrication process than other pumping techniques.

Active pumping has been pursued in various configurations, among which four commonly used configurations are discussed here: (1) single membrane pumping (valve and micropump (VAMP)) (Stehr *et al.* 1996), (2) dual-membrane pumping (Cabuz *et al.* 2001), (3) serial (peristaltic) pumping (Huang *et al.* 2006), and (4) independent valve pumping (Kim 2006).

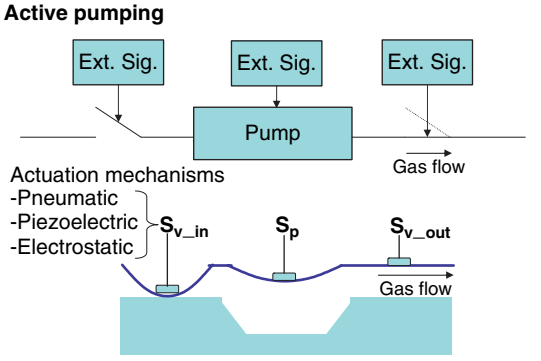


Figure 13 General structure of active pumping (*top*) and an active micropump (*bottom*).

Stehr introduced a single-membrane valve in a diffuser pump called VAMP to minimize unwanted flow losses while maintaining a high-frequency operation. VAMP uses one active membrane that works as a valve and as a pump membrane and another passive membrane that compresses the gas (**Figure 14**, *left*) (Stehr *et al.* 1996). In VAMP, gas is inhaled as the active membrane moves up and down, then it is compressed as the passive membrane moves with the delay from the motion of the active membrane, and finally it is pumped out through the outlet. The VAMP operated a 6-mm diameter membrane piezoelectrically at a frequency of 16.5 kHz producing a flow rate of 8 sccm. The pumping efficiency is estimated to be between 1.45% and 3.4%.

Another active pumping structure using dual membranes was reported by Cabuz *et al.* (2001). The pump consists of two polymer membranes (metallized Kapton) containing gas transfer holes to form a check valve. It transfers a certain volume of gas without compressing it by folding and moving the two membranes with closed gas transfer holes on each membrane (**Figure 14**, *right*). It completes a cycle by separating the dual membranes and returning them separately to their original position while letting gas leak through the open transfer holes. This dual-membrane pumping generated a gas flow rate of 30 sccm and a pressure of 2 kPa by actuating the two membranes electrostatically at a frequency of 100 Hz with an operation voltage of 200 V while consuming a power of 8 mW, which resulted in an efficiency of 1.85% (Cabuz *et al.* 2001). Note that, as explained in the earlier sections, pressure rise is produced by gas compression, and because the dual-membrane pump does not compress gas during its cycle, it produces a low pressure differential of only 2 kPa. Also note that the operation frequency of the dual-membrane pumping is relatively low (100 Hz).

Serial (peristaltic) pumps, which have been popular in liquid pumping, have also been used in active gas pumping. Serial pumps squeeze gas from one side to the other by moving membranes in a sequential order (**Figure 15**). Such sequential motion is especially useful in transferring incompressible fluids, such as liquid, while minimizing leakage. In compressible gas pumping, the minimization of dead volume between chambers or under the incompletely deflected membrane is critical to high-performance

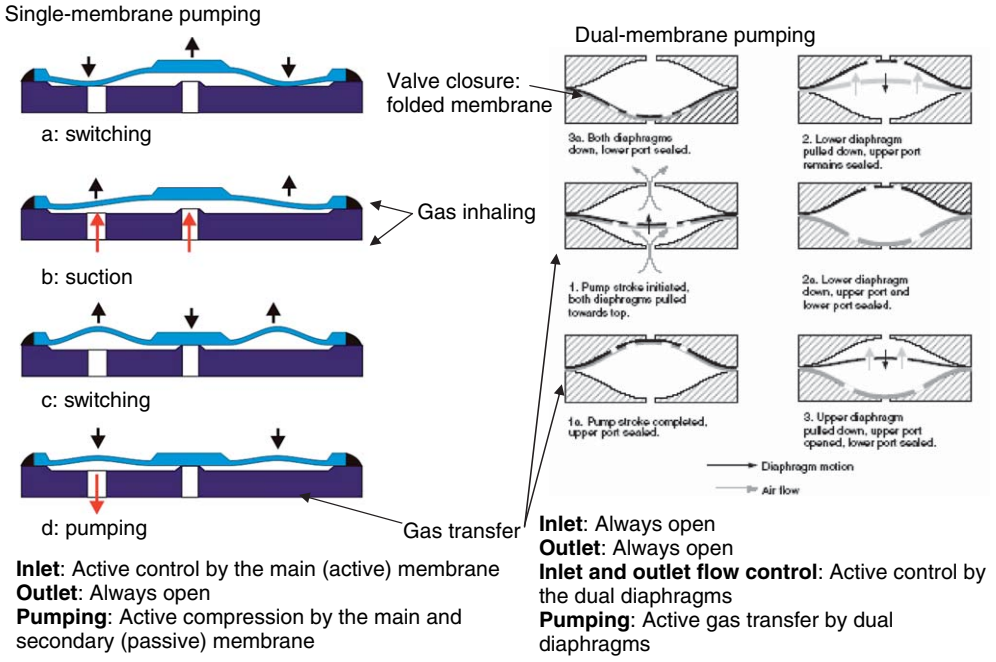


Figure 14 *Left*: Operation principle of active pumping using a single (active) membrane structure. (Source: Stehr M, Messner S, Sandmaier H, Zengerle R 1996 The VAMP – A new device for handling liquid or gases. *Sens. Actuators A* **57**, 153–7.) *Right*: Operation principle of active pumping using a dual-diaphragm structure. (Source: Cabuz C, Herb W R, Cabuz E I, Lu S T 2001 The dual diaphragm pump. *Proc. 14th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2001)*, Interlaken, Switzerland, pp. 519–22.)

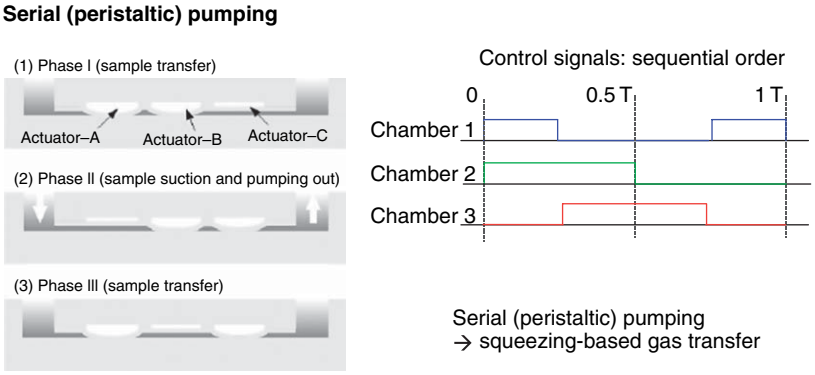


Figure 15 General structure (*left*) and operation principle (*right*) of active pumping in a serial (peristaltic) configuration. (Source: Jeong O, Morimoto T, Watanabe Y, Konish S 2006 Peristaltic PDMS pump with perfect dynamic valves for both gas and liquid. *Proc. 19th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '06)*, Istanbul, Turkey, pp. 782–5.)

operation. Thus, serial pumping is used frequently with high compression actuation, such as thermo-pneumatic and pneumatic actuation. Serial pumping was first used in the first gas micropump by Judy in 1991 (Judy *et al.* 1991) where the successful sequential electrostatic actuation of three membranes was reported, but no measured gas flow was observed. This was most likely because the employed surface micromachining technology limited the maximum size (0.4 mm × 0.4 mm) of the pumping membrane, which resulted in a small volume displacement and a gas flow that was too small to be monitored. Measured flow rates of 0.086 and 0.108 sccm were reported in later pneumatic micropumps reported by Jeong (Jeong *et al.* 2006) and Huang (Huang *et al.* 2006), respectively. They formed three serially connected pumping chambers with polydimethylsiloxane (PDMS) membranes in each chamber. The

PDMS membranes were actuated pneumatically using 60 and 172 kPa of external pressure, respectively, at an operating frequency of 10 Hz. Huang's pump generated a pressure of 1.6 kPa. It is notable that the serial pumping process does not involve active gas compression in its cycle, thus resulting in a low pressure rise.

The most advanced configuration of active pumping requires independent valve and membrane operation to optimize pumping performance, as shown in **Figure 16** (Kim 2006). Individual microvalves (inlet and outlet) operate independently from each other and from the pump membrane. An inlet valve signal governs the movement of the inlet valves and the gas flow from top chambers to bottom chambers. An outlet valve signal governs the movement of the outlet valves and the gas flow from the bottom chambers to the top chambers. The timing of the two electrical signals for inlet and outlet microvalves can be adjusted to control micropump performance variables such as flow rate and pressure generation. Different from the pump membrane signal, each valve signal can be delayed or shortened within a cycle, allowing the pressure to build up longer or the flow to transfer faster. The flow rate and pressure rise of the pump are determined by the adjusted duration of the valve opening (gas transfer), as shown in **Figure 16**. When the opening duration is long (close to half the cycle: 0.5 T), the pump flow rate is maximized and the pressure rise is very small because most gases transfer without being confined and compressed. When the opening duration is short (e.g., 0.25 T), the duration of the membrane compression is increased so that the pressure increases and the flow rate decreases. Kim applied independent valve control to an 18-stage peristaltic gas micropump. This pump consists of 18 pumping stages and 19 microvalves, as shown in **Figure 17** (Kim *et al.* 2006a). By independently controlling the individual valves, they (1) proved that gas pressure could be controlled and accumulated using multiple stages by confining the gas during compression, (2) generated gas flows using fluidic resonance at a high frequency by controlling individual pumping membrane, and (3) demonstrated multiple pumping modes by modulating the duration of gas compression. Independent valve pumping produced a high pressure of 17.5 kPa and a high flow rate of 4.0 sccm with a high pumping efficiency of 9.8% while consuming 57 mW. Each membrane was 2 mm × 2 mm and was operated at a frequency of ~15 kHz. It also demonstrated pumping mode control (high flow rate or high pressure generation) by adjusting the valve opening duration.

2.08.3.4 Valveless Pumping

Valveless pumping uses a nonmechanical approach (no moving parts) to generate high pressure differential in the pumping chamber. In the macroscale, several mechanisms have been used, including thermal transpiration (Knudsen pump), ion capturing using chemisorption (electrostatic ion pump (ESP)), physisorption (sputter ion pump (SIP)), and particle implantation on pumping surfaces (getter pump). In

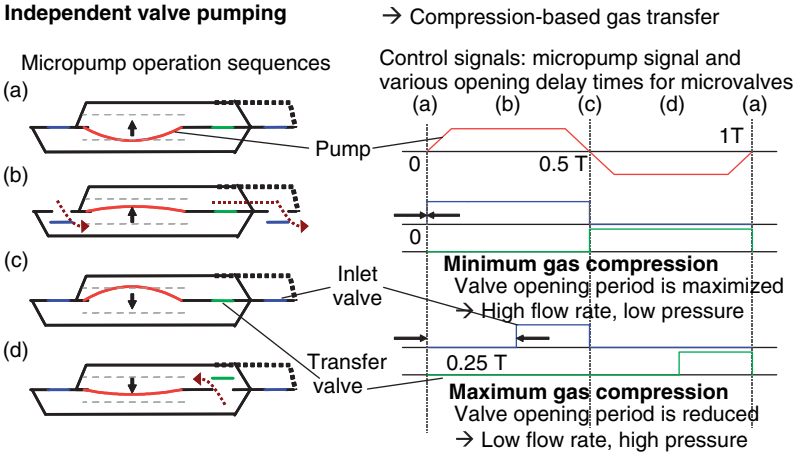
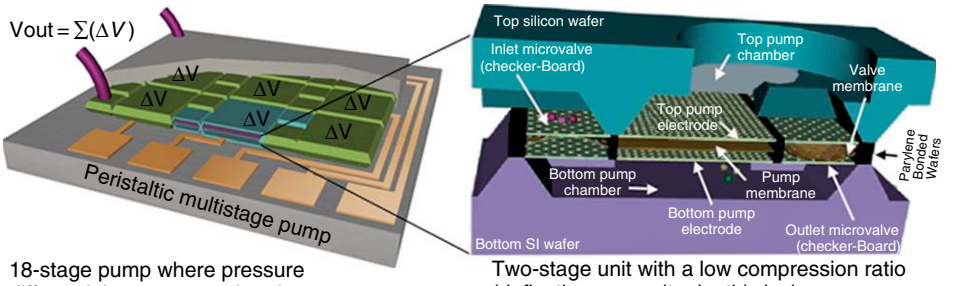


Figure 16 *Left*: Operation principle of active pumping using independent valve control. *Right*: The inlet and outlet valve timing to regulate pumping performance. (Source: Kim H 2006 An integrated electrostatic peristaltic gas micropump with active microvalves. Ph.D. thesis, University of Michigan.)



18-stage pump where pressure differentials are accumulated

Two-stage unit with a low compression ratio (deflection vs. cavity depth) design

Figure 17 *Left*: Structure of the 18-stage gas micropump. *Right*: Details of a two-stage pump unit. (Source: Kim H, Astle A, Najafi K, Bernal L, Washabaugh P 2006a Integrated peristaltic 18-stage electrostatic gas micro pump with active microvalves. *Proc. 12th Solid-State Sensor, Actuator and Microsystems Workshop (Hilton Head 2006)*, Hilton Head Island, SC, USA, pp. 292–5.)

the microdomain, only Knudsen pumping and SIPs have been realized.

Thermal transpiration pumping was originally discovered by Knudsen in 1910 and was theoretically demonstrated in a microdomain Knudsen pump by Vargo and Young in 1999 (Vargo *et al.* 1999, Young 1999). A Knudsen pump consists of several cascaded chambers, each of which is connected to very tiny diameter channels, as shown in **Figure 18**. One side of a given chamber is connected to the adjoining chamber using several microchannels, whereas the other side of the chamber is connected to another adjoining chamber using fewer channels. Because the diameter of these channels is very small, gas flow (through these channels) is minimized because gas flows in these channels in the free molecular regime instead of the continuum regime and leads to a pressure rise proportional to the temperature difference, $\frac{P_H}{P_C} = \sqrt{\frac{T_H}{T_C}}$. A theoretical estimation suggested that a cavity would be evacuated from 152 torr (~20.26 kPa) down to 3 torr (~0.4 kPa) using 16 stages of cold and hot chambers at 300 and 600 K, respectively (Young 1999). Later, a microfabricated Knudsen pump was reported to have evacuated a cavity down to 0.46 atm (~45 kPa) using a power of only 80 mW (McNamara and Gianchandani 2005). This microfabricated version used a single stage of cold, hot, and cold chambers where hot chambers were heated up to 600°C. It is notable that the Knudsen pumps had only minimal flow rate (less than 0.010 sccm).

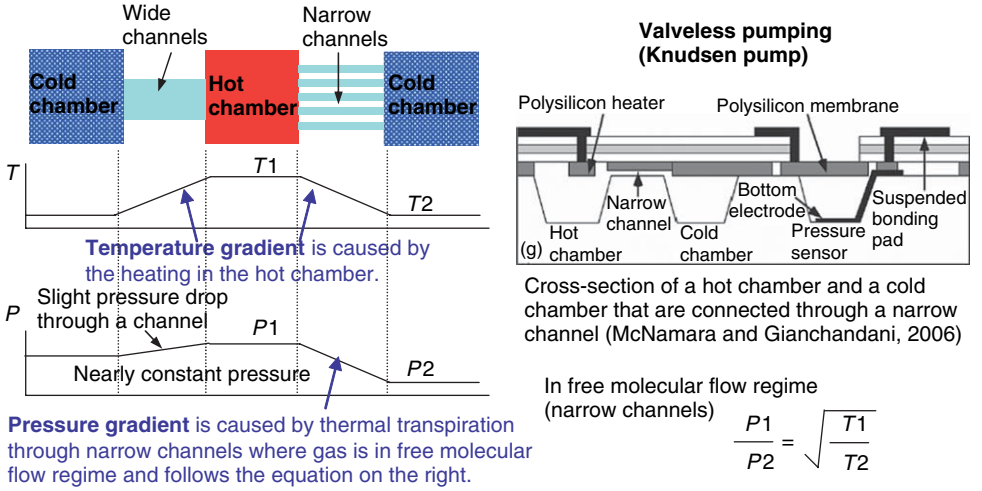


Figure 18 *Left*: Operation principle of a Knudsen pump where pressure differential is generated up by heating of gas. (Source: Kim H 2006 An integrated electrostatic peristaltic gas micropump with active microvalves. Ph.D. thesis, University of Michigan.) *Right*: Cross-sectional view of a Knudsen micropump. (Source: McNamara S, Gianchandani Y B 2005 On-chip vacuum generated by micromachined Knudsen pump. *J. Microelectromech. Syst.* **14(4)**, 741–6.)

An SIP was also recently realized in the micro-domain (Wright and Gianchandani 2006). This pump captures gas molecules and creates a vacuum in an isolated chamber. The developed micro SIP used a high voltage (500–1000 V) to create discharges in the closed chamber. The discharges, in turn, caused the sputtering of titanium, which was previously deposited in the chamber. The sputtered titanium would capture gas ions in the chamber, thus lowering the pressure inside. It was reported that a vacuum of 168 torr (~22.4 kPa) could be produced in a volume of 6.33 cc.

2.08.3.5 Summary and Comparison

Table 2 summarizes the performance and characteristics of these four pumping techniques. Passive pumps have low reverse flow and leakage but cannot operate at high frequencies. Diffuser pumps provide the simplest flow control method as well as easier fabrication. Although they can operate at high frequency, they suffer from reverse flow and leakage. Active pumps combine the advantages of passive and diffuser pumps, allowing gas pumping at high frequency with minimized reverse flow and leakage. Valveless pumps are easy to fabricate in the micro-scale, can generate high pressure differential, but produce very low flow rates.

2.08.4 Review of Pump Actuation Mechanisms

The selection of the actuation mechanism plays an important role in determining the performance of a gas micropump because it influences the speed, force, and deflection distance of the pumping membrane and the microvalve. Gas micropumps can be divided into two categories based on their actuation mechanism: mechanical and nonmechanical. Mechanical pumps need actuators to move the membrane and valves, whereas nonmechanical pumps rely on gas

Table 2 Comparison of pumping principles in gas micropumps

	Passive	*Diffuser*	*Active*	*Valveless*
References	Bohm *et al*. (1999), Kamper *et al*. (1998), Rapp *et al*. (1994), Richter *et al*. (1998), Yang *et al*. (2006)	Gerlach (1997), Gerlach and Wurmus (1995), Olsson *et al*. (1997), Schabmueller *et al*. (2002), Stemme and Stemme (1993), Wijngaart *et al*. (2002)	Cabuz *et al*. (2001), Huang *et al*. (2006), Jeong *et al*. (2006), Judy *et al*. (1991), Kim (2006), Kim *et al*. (2006a), Schomburg *et al*. (1994), Stehr *et al*. (1996, 1997)	Linderman *et al*. (2005), McNamara and Gianchandani (2005), Vargo *et al*. (1999), Wright and Gianchandani (2006), Young (1999)
Flow leakage (pumping efficiency in %)	Low (>10)	High (<3)	Low (2.2–9.8)	Minimal
Operation frequency	Low (<1 kHz)	High (3.4–33 kHz)	High (10–18 kHz)	–
Fabrication	Medium	Easy	Complex	Easy
Pressure control	Medium: inlet closed and outlet open	Difficult: inlet and outlet always open	Easy: inlet closed and outlet closed	Easy: minimal flow loss
Control	Easy	Easy	Complex (external signal)	Easy
Power[a] (mW)	Low (3.18)	Low (N/R)	High (8–450)	High (80–2400)
Flow rate (sccm)	0.044–40	0.4–35	0.1–30	0–0.002
Pressure (kPa)	N/R	2.3–5	2–35	20–46
Other comments	Minimal reverse flow and leakage, low-frequency operation	Reverse flow and leakage, high-frequency operation	Minimized reverse flow and leakage, high-frequency operation	No moving parts, high pressure differential, and low flow rate

[a]Includes the power of the membrane actuator.

heating or other interactions to pump gas or create a vacuum. Clearly, the actuator plays an important role in mechanical pumps, and so different actuation mechanisms and their use in gas micropumps will be reviewed in this section.

2.08.4.1 Electrostatic Actuation

Electrostatic actuation uses the attractive electrostatic force between two oppositely charged electrodes to move a membrane that is one of the two electrodes, as shown in **Figure 19**. When a voltage is applied across the electrodes, electrostatic actuation moves the membrane against the other fixed electrode, resulting in the compression of the space enclosed under the membrane. When the applied voltage is removed, the movable membrane returns to its original position because of its mechanical stiffness.

Electrostatic actuation is easy to apply in microdevices because of its compatible fabrication process, the ability to form small air gaps (10 μm), low power consumption, high speed operation, and easy electrical control of various components. An additional advantage is resistance to temperature and humidity changes. However, electrostatic actuation is susceptible to particulates, although particulates could be filtered using an inlet filter, and it requires high operation voltages (>80 V).

In the design of an electrostatic pump, the fundamental challenge lies in the fact that the maximum out-of-plane (vertical) deflection is physically limited by the small gap separation between the two electrodes. The small gap limits the maximum volume displacement and thus the flow rate. Increasing the gap separation would require much higher operating voltages. For most electrostatic micropumps with a membrane size less than 3 mm × 3 mm, a gap distance larger than 10 μm becomes impractical requiring high operating voltages (>500 V).

To overcome these issues (limited volume displacement and high operating voltage), several techniques have been developed, including the use of polymer membranes, dual drive electrodes, and curved electrodes (Cabuz *et al.* 2001, Kim *et al.* 2006a). A polymer membrane can deflect over a relatively large distance due to its high elongation rates; however, its use was limited in electrostatic gas micropumps due to its limited stiffness and restoring force. This can be overcome by using dual electrostatic electrodes to pull the compliant membrane in two directions as needed. The dual electrodes also increase (double) the volume displacement of the membrane. The curved electrode decreases the operating voltages compared with conventional flat electrodes (Legtenberg *et al.* 1997). It provides larger electrostatic forces, due to smaller air gap at the edges. Thus, the membrane is moved to a much larger vertical deflection with a lower voltage because a large force is created around the edges where the two electrodes are closest.

2.08.4.2 Piezoelectric Actuation

Piezoelectric actuation uses the deformation of piezoelectric materials, such as polyvinylidene fluoride (PVDF), zinc oxide (ZnO), and lead zirconate titanate (PZT), under an electric field, to vibrate a thin membrane that the materials are attached to, as shown in **Figure 20**. Piezoelectric actuation is the most commonly used actuation technique in both macro- and micropumps, thanks to its high-frequency operation,

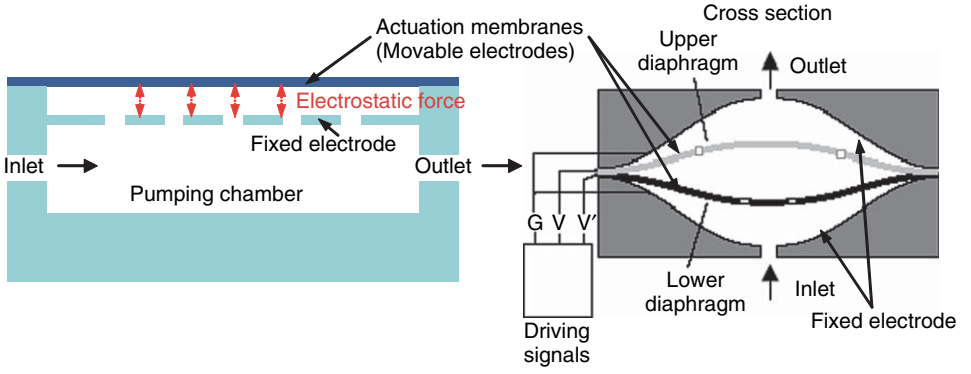


Figure 19 *Left*: Basic structure of an electrostatic pump. (Kim H 2006 An integrated electrostatic peristaltic gas micropump with active microvalves. Ph.D. thesis, University of Michigan.) *Right*: An electrostatic gas micropump. (Source: Cabuz C, Herb W R, Cabuz E I, Lu S T 2001 The dual diaphragm pump. *Proc. 14th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2001)*, Interlaken, Switzerland, pp. 519–22.)

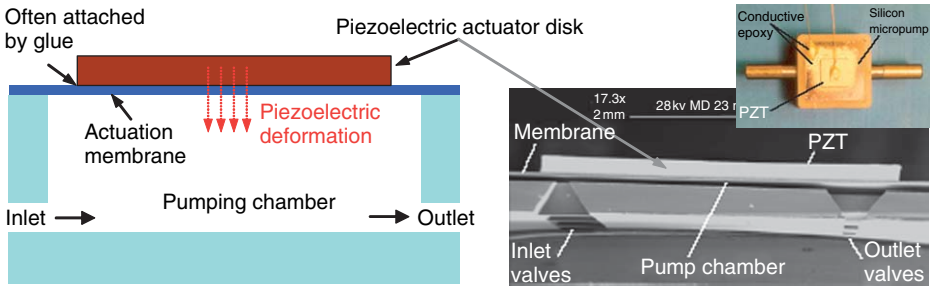


Figure 20 *Left*: Basic structure of a piezoelectric pump. (Kim H 2006 An integrated electrostatic peristaltic gas micropump with active microvalves. Ph.D. thesis, University of Michigan.) *Right*: A developed piezoelectric gas pump with diffuser microvalves. PZT, lead zirconate titanate. (Source: Schabmueller C, Koch M, Mokhtari M, Evans A G R, Brunnschweiler A, Sehr H 2002 Self-aligning gas/liquid micropump. *J. Micromech. Microeng.* **12**, 420–4.)

precise motion control, low power consumption, high stroke force, and operation at cryogenic temperatures (Yang and Wiberg 2003). However, piezoelectric actuators have limited deflection and require high operating voltages (>80 V).

The main challenge in realizing a piezoelectric micropump is the fabrication process that is complex, is hard to integrate and often requires manual assembly (gluing), and is limited in making small-size actuators. Some researchers have developed precision machining technology. For example, Kamper combined injection molding with precision assembly to build a piezoelectric gas micropump with a 10-mm diameter pumping membrane and predicted a cost of only ~$3 (Kamper *et al.* 1998). In the pump, the piezoelectric material (piezoceramic disk) was attached to and used to actuate a 200-μm-thick polycarbonate diaphragm at a frequency of 100 Hz, resulting in a gas flow rate of 3.5 sccm and a pressure differential of 35 kPa.

Another challenge is the limited deflection of piezoelectric membrane actuators. Although a stack of piezoelectric layers can increase the deflection amplitude, it generally cannot operate at a higher frequency (>10 kHz). Thus, researchers have used a single piezoelectric layer for high-frequency operation (~33 kHz) while enlarging the size of the pumping membrane to obtain a high mass flow rate of 8 sccm (Olsson *et al.* 1997).

2.08.4.3 Thermopneumatic and Pneumatic Actuation

Thermopneumatic or pneumatic actuation uses the volume expansion of a material or a pumping membrane when heated or inflated by external sources, as shown in **Figure 21**. This volume expansion is often generated by microheaters built on top of the membrane in a thermopneumatic pump (Bustgens *et al.* 1994) and by external pressure in a pneumatic pump (Huang *et al.* 2006, Jeong *et al.* 2006, Rapp *et al.* 1994). Thermopneumatic actuation provides a large deflection and a high force but responds slowly (less than

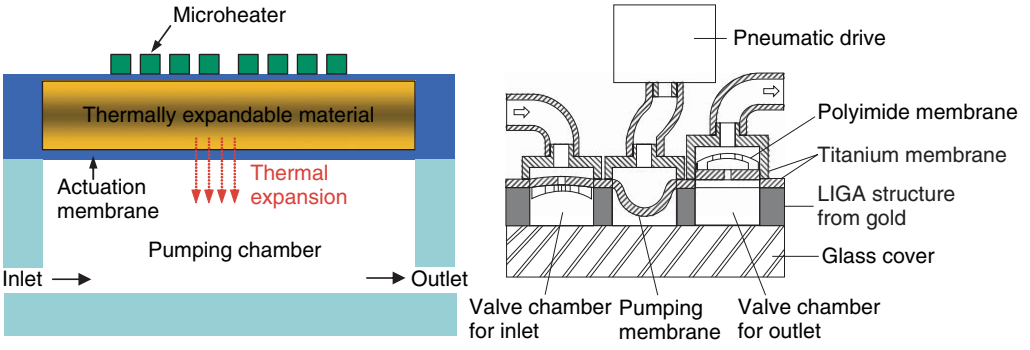


Figure 21 *Left*: Basic structure of a thermopneumatic micropump. LIGA, an acronym from German words for lithography, electroplating, and molding. (Source: Kim H 2006 An integrated electrostatic peristaltic gas micropump with active microvalves. Ph.D. thesis, University of Michigan; reproduced with permission.) *Right*: Cross-sectional view of a pneumatic gas micropump. (Source: Rapp R, Schomburg W K, Maas D, Schulz J, Stark W 1994 LIGA micropump for gases and liquids. *Sens. Actuators A* **40**, 57–61.)

10 Hz, often decided by the thermal time constant), requires high power consumption, and produces a low flow rate (<0.1 sccm). Pneumatic micropumps use external gas pressures to actuate the membrane. By applying the external pressure (as high as >100 kPa), the membrane is forced into large deflection. The requirement for an external pressure source generally prevents the use of pneumatic pumps in integrated microsystems.

2.08.4.4 Electromagnetic Actuation

Electromagnetic actuation is based on an electromagnetic field generated by either magnets or microcoils. The most common structure for an electromagnetic actuator in the microdomain uses a permanent magnet above a fixed coil, as shown in **Figure 22** (*top*). Depending on the direction of current flow through the coil, the direction of the electromagnetic field changes, causing the magnets to realign along the field. Such realignment of magnets leads to the up and the down motion of the membrane. Another common structure for an electromagnetic actuator has a slightly modified design: a permanent magnet is placed inside a fixed coil, as shown in **Figure 22** (*bottom*). Depending on the direction of current flow in the coil, the magnet moves up or down, thus actuating the attached pumping membrane.

Electromagnetic actuation has not been used widely for gas pumping in the microscale because of the difficulty in scaling and integrating magnets. Instead, it has been applied in a few mesoscale micropumps (Bohm *et al.* 1999). Electromagnetic actuation does have some advantages in the microdomain because it provides bidirectional (push-pull) actuation, thus larger volume displacement, and can operate at high frequencies. It is functional over a relatively large displacement and is especially desirable for use in dust-filled environments or in conductive fluids. Because it requires a large current (more than a few amperes) to obtain a reasonable amount of force, it often consumes a large power of more than a few watts.

Bohm developed a mesoscale electromagnetic pump using the magnet-coil configuration in a volume of 11 mm × 11 mm × 11 mm, as shown in **Figure 22** (*bottom*). The membrane was attached to a small permanent magnet that was actuated using a 1000-turn coil formed around the pump package. When the permanent magnetic material (NdFeB, 5 mm diameter) moves out-of-plane by the Lorentz force generated from the current-carrying microcoils, the pumping membrane is deflected and compresses the pumping chamber. Such movements generated a large

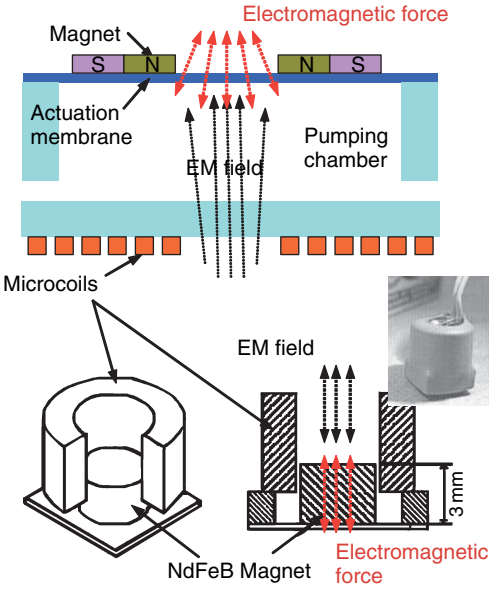


Figure 22 *Top*: Basic structure and operating principle of an electromagnetic pump with a membrane with embedded permanent magnets and a microfabricated actuation micro coil. (Kim H 2006 An integrated electrostatic peristaltic gas micropump with active microvalves. Ph.D. thesis, University of Michigan.) *Bottom*: A mesoscale electromagnetic gas pump. (Source: Bohm S, Olthuis W, Bergveld P 1999 A plastic micropump constructed with conventional techniques and materials. *Sens. Actuators A* **77**, 223–8.)

force of 0.1 N and a large deflection of several hundred microns, which resulted in large flow rate of 50 sccm.

2.08.4.5 Summary and Comparison

Gas micropumps have employed different actuation techniques to pump compressible gases, as summarized in Table 3. The selection of the actuation mechanism depends on the goals of a particular application. Electrostatic actuation would be the best choice for low power consumption, high-speed operation, and easy microfabrication without material restrictions, despite its high operating voltage and susceptibility to particulates. Piezoelectric actuation offers low power consumption, high-speed operation, and operation at a low temperature, but it requires a more complex fabrication process. Thermopneumatic actuation provides high force and large deflection. Electromagnetic actuation offers high force and immunity to dust.

2.08.5 Conclusions and Discussion

Figure 23 shows the design road map of past gas micropumps and the design choices and performance achievements in the evolution of past gas micropumps. All of the key components and characteristics required in gas pumps are listed in the figure, along with the citation to the paper that addresses that requirement or utilizes a particular technology to achieve a specific requirement. Despite recent successes and remarkable improvements in performance, several challenges and issues remain for gas micropumps. The primary

Table 3 Comparison of pumping principles used in gas micropumps

	Electrostatic	*Piezoelectric*	*(Thermo)pneumatic*	*Knudsen/sputter ion*	*Electromagnetic*
References	Cabuz *et al*. (2001), Judy *et al*. (1991), Kim *et al*. (2006a), Linderman *et al*. (2005)	Gerlach (1997), Gerlach and Wurmus (1995), Kamper *et al*. (1998), Olsson *et al*. (1997), Richter *et al*. (1998), Schabmueller *et al*. (2002), Stehr *et al*. (1996), Stemme and Stemme (1993), Wijngaart *et al*. (2000, 2002), Yang *et al*. (2006)	Huang *et al*. (2006), Jeong *et al*. (2006), Rapp *et al*. (1994), Schomburg *et al*. (1994)	McNamara and Gianchandani (2005), Vargo *et al*. (1999), Wright and Gianchandani (2006), Young (1999)	Bohm *et al*. (1999)
Generic properties					
Frequency	High	High	Low	–	High
Deflection	Small	Medium	Large	–	Large
Force	Small	Medium	Large	–	Large
Fabrication	Easy	Complex	Complex	Easy	Complex
Control	Easy	Easy	Medium	Easy	Hard (due to interferences between stages)
Power	Low	Low	High	High	High
Voltage	High	High	Low	Low/high	Low
Current	Low	Low	High	High/low	High
Reported performance					
Flow rate (sccm)	<30	<35	<0.1	<0.1	N/A
Pressure (kPa)	<5	<35	<3.8	<50	N/A
Frequency	Tens of kilohertz	Tens of kilohertz	Slow	Slow	N/A
Deflection (μm)	<5	<3	<30	N/A	N/A
Power (mW)	<100	<200	<8000	<8000	

Source: Kim H 2006 An integrated electrostatic peristaltic gas micropump with active microvalves. Ph.D. thesis, University of Michigan.

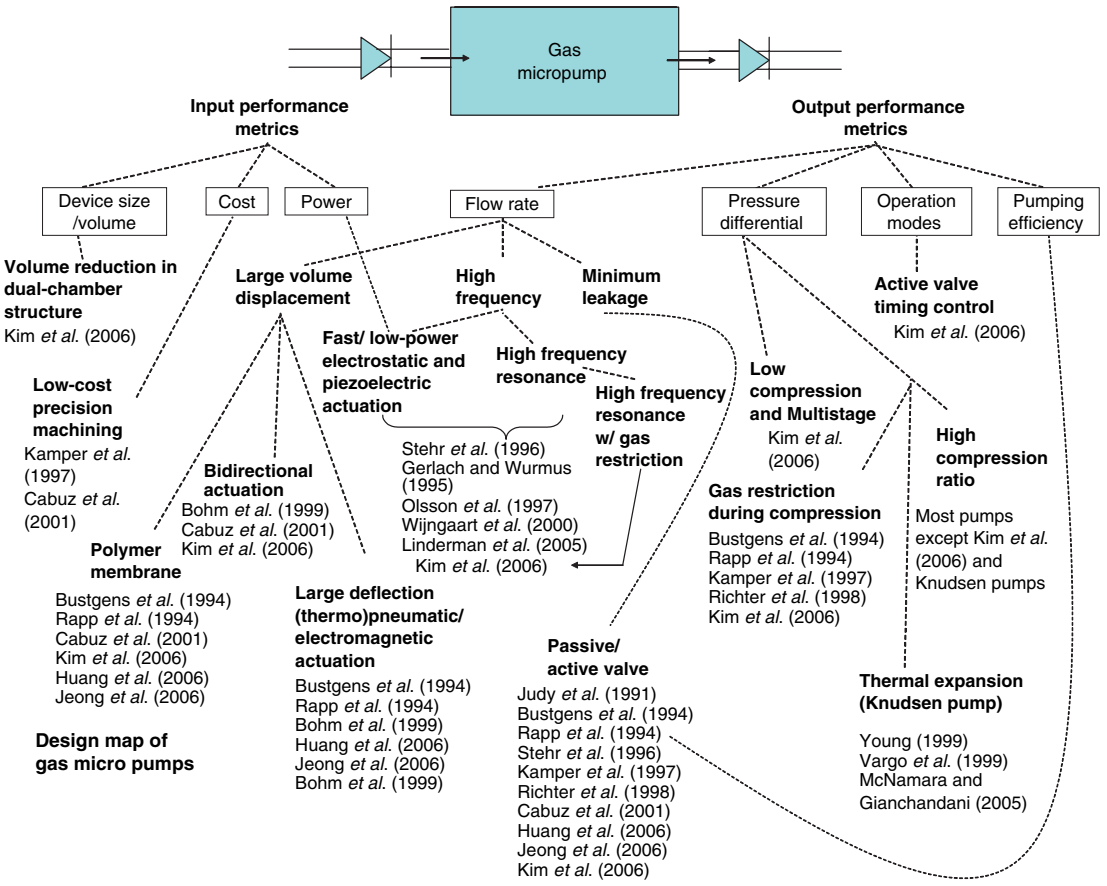


Figure 23 Design road map for gas micropumps.

shortcoming of most gas micropumps is the limited pressure difference they can produce. This is because of the limited actuation force and deflection amplitude, which are due to the limits imposed on actuators by the fabrication technology and size. Therefore, the major future developments in gas micropumps will involve better and more flexible fabrication processes.

The fabrication of a gas micropump requires three key manufacturing steps: (1) formation of a movable membrane, (2) formation of pumping chambers, and (3) assembly of the membrane and pumping chambers. The construction of thin membranes over a pumping chamber has been performed using three major manufacturing technologies: (1) bulk and surface micromachining and wafer bonding, (2) membrane transfer, and (3) mechanical machining of commercial thin films, as listed in **Table 4**. The fabrication processes used to form the pumping chambers has been relatively straightforward and include techniques such as conventional surface and bulk micromachining, injection (thermoplastic) molding, and LIGA (an acronym from German words for lithography, electroplating, and molding), as listed in **Table 5**. Gas micropumps also require suitable packaging and assembly techniques. Packaging has been typically performed by using wafer bonding and mechanical assembly, as listed in **Table 6**. Wafer bonding can be performed at either a high or a low temperature. Among high-temperature bonding techniques, anodic bonding has been frequently used. Low-temperature bonding approaches have used parylene, PDMS, and epoxy, all of which can be performed at less than 300°C. Mechanical assembly has used laser welding and screws. Although the mechanical assembly is appropriate in some applications for mass production, it does not really fit in manufacturing of gas micropumps due to the lack of precise alignment required in the microdomain.

Clearly, these and other fabrication technologies will continue to improve, and new techniques will be developed to further improve the performance of gas micropumps. Many emerging applications require further miniaturization of gas micropumps that can produce high pressure and high flow rate and consume low power (Agah *et al.* 2003, Chen *et al.* 2003,

Table 4 Comparison of three major technologies used for membrane formation

Technologies	*Bulk/surface micromachining and wafer bonding*	*Membrane transfer technology*	*Mechanical machining manufacturing*
References	Gerlach and Wurmus (1995), Huang *et al*. (2006), Judy *et al*. (1991), Richter *et al*. (1998), Schabmueller *et al*. (2002), Stehr *et al*. (1996), Wijngaart *et al*. (2002)	Bustgens *et al*. (1994), Jeong *et al*. (2006), Kim *et al*. (2006a), Linderman *et al*. (2005), Rapp *et al*. (1994)	Bohm *et al*. (1999), Cabuz *et al*. (2001), Kamper *et al*. (1998), Olsson *et al*. (1997)
Fabrication methods	Deposition and etching	Bonding and detachment	Gluing, laser welding, and screwing
Membrane			
Thickness	Thin	Thin	Thick
Materials	Limited	Less limited	Diverse
Material limitations	Etching selectivity limited	Bonding limited	Few limitations
Over shallow cavity	Hard to fabricate	Easy to fabricate	Easy to fabricate
Cavity encapsulation	Hard	Possible	Possible
Cavity dimension fabrication	Hard to control	Easy to control	Hard control
Compatibility with other microfabrication	High	Medium	Low
Precision	High	High	Low
Cost	Medium	High	Low

Table 5 Comparison of major technologies used for cavity formation

Technologies	*Bulk/surface micromachining*	*Mechanical machining manufacturing*
References	Gerlach and Wurmus (1995), Judy *et al*. (1991), Kim *et al*. (2006a), Richter *et al*. (1998), Schabmueller *et al*. (2002), Stehr *et al*. (1996), Wijngaart *et al*. (2002)	Bohm *et al*. (1999), Bustgens *et al*. (1994), Cabuz *et al*. (2001), Huang *et al*. (2006), Jeong *et al*. (2006), Kamper *et al*. (1998), Olsson *et al*. (1997), Rapp *et al*. (1994)
Fabrication methods	Deposition and etching	Injection molding and LIGA process
Cavity		
Materials	Limited	Diverse
Material limitations	Etching selectivity limited	Few limitations
Shallow cavity	Easy to fabricate	Easy to fabricate
Dimension Fabrication	Easy to control	Easy to control
Compatibility with other microfabrication	High	Low
Precision	High	Low
Mass production	Medium	High
Cost	Medium	Low

LIGA, an acronym from German words for lithography, electroplating, and molding.

Table 6 Comparison of pump assembly techniques

	Wafer bonding					
	High temperature	*Low temperature*			Mechanical assembly	
Bonding technique	Anodic	Parylene	PDMS	Epoxy	Laser welding	Screws
References	Gerlach and Wurmus (1995), Richter *et al*. (1998), Wijngaart *et al*. (2002)	Kim *et al*. (2006a)	Huang *et al*. (2006), Jeong *et al*. (2006)	Bohm *et al*. (1999), Bustgens *et al*. (1994), Schabmueller *et al*. (2002), Stehr *et al*. (1996)	Kamper *et al*. (1998)	Cabuz *et al*. (2001), Olsson *et al*. (1997)
Bonding requirements						
Temperature (°C)	400	230	120	<50	N/A	<30
Electric voltage (V)	1000	0	0	0	0	0
Surface condition	Ultra flat	Flat	Flat	Flat	Rough	Rough
Bond strength (MPa)	30–40	3.6	<5	<5	High	High
Advantages	Proven technology, compatibility with conventional microfabrication, and reasonable bond strength				Low-cost process and high bond strength	
Disadvantages	Relatively expensive process; requirements of high temperature, high voltage, or special surface quality in some cases				Noncompatibility with conventional microfabrication	

PDMS, polydimethylsiloxane.

Elwenspoek 1994, Lambertus *et al.* 2004, Lu *et al.* 2005, de Rooij 1995). Based on the rapid development of past gas micropumps over a short period of time, it is clear that these goals can and will be achieved in future micropumps.

Acknowledgments

The authors are grateful to Professor Peter Washabaugh and Dr. Aaron Astle for their valuable contributions to the development of the multistage peristaltic gas micropump reported in this chapter. The research on the multistage pump reported was supported by the Engineering Research Center Program of the National Science Foundation under Award Number EEC-9986866.

References

Agah M, Potkay J A, Driscoll J A, Sacks R D, Kaviany M, Wise K D 2003 Thermal behavior of high-performance temperature-programmed microfabricated gas chromatography columns. *Proc. 12th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '03)*, Boston, MA, USA, June 8–12, 2003, pp. 1339–42

Arana L, Schaevitz S, Franz A, Schmidt M A, Jensen K F 2003 A microfabricated suspended-tube chemical reactor for thermally efficient fuel processing. *J. Microelectromech. Syst.* **12(5)**, 600–12

Astle A, Paige A, Bernal L P, Munfakh J, Kim H, Najafi K 2002 Analysis and design of multistage electrostatically-actuated micro vacuum pumps. *Proc. ASME Int. Mechanical Engineering Congress and Exposition (ASME '02)*. New Orleans, LA, USA, IMECE 2002-39308

Astle A, Kim H, Bernal L P, Najafi K, Washabaugh P 2007 Theoretical and experimental performance of a high frequency gas micropump. *Sens. Actuators A: Phys.* **134(1)**, 245–56

Bohm S, Olthuis W, Bergveld P 1999 A plastic micropump constructed with conventional techniques and materials. *Sens. Actuators A* **77**, 223–8

Bustgens B, Bacher W, Menz W, Schomburg W K 1994 Micropump manufactured by thermoplastic molding. *Proc. 7th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '94),* Oiso, Japan, pp. 18–21

Cabuz C, Herb W R, Cabuz E I, Lu S T 2001 The dual diaphragm pump. *Proc. 14th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS 2001)*, Interlaken, Switzerland, pp. 519–22

Cellpoint Scientific Inc. http://cellpointscientific.com

Chen C-F, Kuo S-C, Chu C-C, Tseng F-G 2003 A power-free liquid driven method for micro mixing application. *Proc. Sixteenth IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '03)*, Kyoto, Japan, pp. 100–3

Chou T-K A, Najafi K, Muller M O, Bernal L P, Washabaugh P D 2001 High-density micromachined acoustic ejector array for micro propulsion. *Proc. 11th Int. Conf. Solid State Sensors and Actuators Transducers '01/Eurosensors XV,* Munich, Germany, June 10–14, 2001, pp. 890–3

Chou T-K A, Najafi K, Muller M O, Bernal L P, Washabaugh P D, Parviz B A 2002 Micromachined e-jet for IC chip cooling. *Proc. IEEE Int. Solid-State Circuits Conference (ISSCC 02)*, San Francisco, CA, USA, pp. 356–7

Eckart C 1948 Vortices and streams caused by sound waves. *Phys. Rev.* **73**, 68–76

Elwenspoek M 1994 Some selected research item of the micro mechanics department at MESA. *Proc. 5th Int. Symp. Micro Machine and Human Science*, Nagoya, Japan, pp. 8–31

Esashi M, Shoji S, Nakano A 1989 Normally close microvalve and micropump fabricated on a silicon wafer. *Proc. Micro Electro Mechanical Systems: An Investigation of Micro Structures, Sensors, Actuators, Machines and Robots*, Salt Lake City, UT, USA, pp. 29–34

Filippi P, Habault D, Lefebvre J, Bergassoli A 1999 *Acoustic: Basic Physics, Theory and Methods*. Academic Press, San Diego, CA

Frye-Mason G C, Manginell R P, Heller E J, Matzke C M, Casalnuovo S A, Hietala V M, Kottenstette R J, Lewis P R, Wong C C 1999 Microfabricated gas phase chemical analysis systems. *Proc. Microprocesses and Nanotechnology Conference,* Tokyo, Japan, pp. 60–1

Gerlach T 1997 Pumping gases by a silicon micro pump with dynamic passive valves. *Proc. Int. Conf. Solid-State Sensors and Actuators (Transducers '03)*, Chicago, IL, USA, pp. 357–60

Gerlach T, Wurmus H 1995 Working principle and performance of the dynamic micropump. *Sens. Actuators A* **50**, 135–40

Gromoll B 1994 Advanced micro air-cooling systems for high density packaging. *Proc. 10th IEEE SEMI-THERM,* San Jose, CA, USA, pp. 53–8

Huang C, Huang S, Lee G 2006 Pneumatic micropumps with serially connected actuation chambers. *J. Micromech. Microeng.* **16**, 2265–72

Jeong O, Morimoto T, Watanabe Y, Konish S 2006 Peristaltic PDMS pump with perfect dynamic valves for both gas and liquid. In: *Proc. 19th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '06)*, Istanbul, Turkey, pp. 782–5

Jiang L, Mikkelsen J, Koo J-M, Huber D, Yao S, Zhang L, Zhou P, Maeety J G, Prasher R, Santiago J G, Kenny T W, Goodson K E 2002 Closed-loop electroosmotic microchannel cooling system for VLSI circuits. *IEEE Trans. Compon. Packag. Technol.* **25(3)**, 347–55

Judy J W, Tamagawa T, Polla D L 1991 Surface-machined micromechanical membrane pump. *Proc. 4th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '91)*, Nara, Japan, pp. 182–6

Kamper K-P, Dopper J, Ehrfeld W, Oberbeck S 1998 A self-filling low-cost membrane micropump. *Proc. 11th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '98)*, Heidelberg, Germany, pp. 432–7

Kim H 2006 An integrated electrostatic peristaltic gas micropump with active microvalves. Ph.D. thesis, University of Michigan

Kim H, Astle A, Najafi K, Bernal L, Washabaugh P 2006a Integrated peristaltic 18-stage electrostatic gas micro pump with active microvalves. *Proc. 12th Solid-State Sensor, Actuator and Microsystems Workshop (Hilton Head 2006)*, Hilton Head Island, SC, USA, pp. 292–5

Kim H, Jauregui A, Morrison C, Najafi K, Bernal L P, Washabaugh P D 2006b Low power electrostatic Helmholtz-resonance microjet generator for propulsion and cooling. In: *Proc. 12th Solid-State Sensor, Actuator, and Microsystems Workshop (Hilton Head '06)*, Hilton Head Island, SC, USA, pp, 127–30

Kim H, Steinecker W H, Lambertus G R, Astle A A, Najafi K, Zellers E T, Bernal L, Washabaugh P, Wise K D 2006c Integrated high-pressure 4-stage micro pump for high speed micro chromatography. *Proc. 10th Int. Conf. Miniaturized Systems for Chemistry and Life Science (uTAS '06)*, Tokyo, Japan, pp. 1037–9

KNF Inc. http://www.knf.com

Kwon J, Yu H, Kim E S 2005 Film transfer and bonding techniques for covering single-chip ejector array with microchannels and reservoirs. *IEEE J. Microelectromech. Syst.* **14(6)**, 1399–408

Lambertus G, Elstro A, Sensenig K, Potkay J, Agah M, Scheuering S, Wise K, Dorman F, Sacks R 2004 Design, fabrication, and evaluation of microfabricated columns for gas chromatography. *Anal. Chem.* **76(9)**, 2629–37

Laser D J, Santiago J G 2004 A review of micropumps. *J. Micromech. Microeng.* **14**, R35–64

Legtenberg R, Gilbert J, Senturia S D, Elwenspoek M 1997 Electrostatic curved electrode actuators. *J. Microelectromech. Syst.* **6(3)**, 257–65

Linderman R J, Nilsen O, Bright V M 2005 Electromechanical and fluidic evaluations of the resonant microfan gas pump and aerosol collector. *Sens. Actuators A* **118**, 162–70

van Lintel H T G, van de Pol F C M, Bouwstra S 1988 A piezoelectric micropump based on micromachining of silicon. *Sens. Actuators A* **15**, 153–7

Lu C-J, Steinecker W H, Tian W-C, Oborny M C, Nichols J M, Agah M, Potkay J A, Chan H K L, Driscoll J, Sacks R D, Wise K D, Pang S W, Zellers E T 2005 First-generation hybrid MEMS gas chromatograph. *Lab-on-a-Chip* **5**, 1123–31

McNamara S, Gianchandani Y B 2005 On-chip vacuum generated by micromachined Knudsen pump. *J. Microelectromech. Syst.* **14(4)**, 741–6

Mesagas Inc. http://www.mesagas.com

Muller M, Bernal L P, Washabaugh P D, Kim H, Najafi K 2002 Resonance effects of electrostatically actuated acoustic jets. *41st Aerospace Sciences Meeting and Exhibit (AIAA)*, Reno, NV, USA, January 6–9, 2003, AIAA Paper 2003-1272

Olsson A, Enoksson P, Stemme G, Stemme E 1996 An improved valve-less pump fabricated using deep reactive ion etching. *Proc. 9th Int. Conf. Micro Electro Mechanical Systems (MEMS '96)*, San Diego, CA, USA, pp. 479–84

Olsson A, Stemme G, Stemme E 1997 The first valve-less diffuser gas pump. *Proc. 10th Int. Conf. Micro Electro Mechanical Systems (MEMS '97)*, Nagoya, Japan, pp. 108–13

Rapp R, Schomburg W K, Maas D, Schulz J, Stark W 1994 LIGA micropump for gases and liquids. *Sens. Actuators A* **40**, 57–61

Rayleigh L 1916 The theory of the HelmHoltz resonator. *Proc. Royal Soc. London, Series A, Containing Papers of a Mathematical and Physical Character* **92(638)**, 265–75

Richter M, Linnemann R, Woias P 1998 Robust design of gas and liquid micropumps. *Sens. Actuators A* **68**, 480–6

Rodgers P, Eveloy V, Pecht M 2005 Extending the limits of air-cooling in microelectronic equipment. *Proc. 6th Int. Conf. Thermal, Mechanical and Multiphysics Simulation and Experiment in Micro-electronics and Micro-systems, EuroSimE 2005,* Berlin, Germany, pp. 695–701

de Rooij N F 1995 Current applications of silicon based Microsystems. *Proc. Sixth Int. Symp. Micro Machine and Human Science (MHS '95)*, Nagoya, Japan, pp. 7–10

Schabmueller C, Koch M, Mokhtari M, Evans A G R, Brunnschweiler A, Sehr H 2002 Self-aligning gas/liquid micropump. *J. Micromech. Microeng.* **12**, 420–4

Schomburg W K, Vollmer J, Bustgens B, Fahrenberg J, Hein H, Menz W 1994 Microfluidic components in LIGA technique. *J. Micromech. Microeng.* **4**, 186–91

Seo Y H, Cho Y-H 2004 MEMS-based direct methanol fuel cells and their stacks using a common electrolyte sandwiched by reinforced microcolumn electrodes. *Proc. IEEE Int. Conf. Micro Electro Mechanical Systems,* Maastricht, Netherlands, pp. 65–8

Smits J G 1989 Piezoelectric micropump with microvalves. *Sens. Actuators A* **21–23**, 203–6

Stehr M, Messner S, Sandmaier H, Zengerle R 1996 The VAMP – A new device for handling liquid or gases. *Sens. Actuators A* **57**, 153–7

Stehr M, Gruhler H, Straatmann H, Messner S, Sandmaier H, Zengerle R 1997 The selfpriming VAMP. *Proc. Int. Conf. Solid-State Sensors and Actuators (Transducers '97)*, Chicago, IL, USA, pp. 351–2

Stemme E, Stemme G 1993 Valveless diffuser/nozzle-based fluid pump. *Sens. Actuators A* **A39(2)**, 159–67

Tuckerman D B, Pease R F W 1981 High-performance heat sinking for VLSI. *Electron Device Lett. IEEE* **2(5)**, 126–9

Vargo S E, Muntz E P, Shiflett G R, Tang W C 1999 Knudsen compressor as a micro- and macroscale vacuum pump without moving parts or fluids. *J. Vac. Sci. Technol. A* **17(4)**, 2308–13

Wang Y, Yuan G, Yoon Y-K, Allen M G, Bidstrup S A 2006 Large eddy simulation (LES) for synthetic jet thermal management. *Int. J. Heat Mass Transfer* **49(13–14)**, 2173–9

Watson T J 1997 *Introduction to Mass Spectrometry*. Lippincott-Raven, Philadelphia, PA

van der Wijngaart W, Andersson H, Enoksson P 2000 The first self-priming and bi-directional valve-less diffuser micropump for both liquid and gas. In: *Proc. 13th Int. Conf. Micro Electro Mechanical Systems (MEMS '00)*, Kyoto, Japan, pp. 674–9

van der Wijngaart W, Ask H, Enoksson P, Stemme G 2002 A high-stroke, high-pressure electrostatic actuator for valve applications. *Sens. Actuators A* **100**, 264–71

Wright S A, Gianchandani Y 2006 A micromachined titanium sputter ion pump for cavity pressure control. *Proc. 19th Int. Conf. Micro Electro Mechanical Systems (MEMS '06)*, Istanbul, Turkey, pp. 754–7

Yang E-H, Wiberg D V 2003 A wafer-scale membrane transfer: Process for the fabrication of optical quality, large continuous membranes. *J. Microelectromech. Syst.* **12(6)**, 804–15

Yang X, Zhou Z, Ye X, Xiao M 2006 Simulation and experimental studies on a piezoelectrically actuated microdiaphragm air pump. *J. Microlith. Microfab. Microsyst.* **5(2)**, 021106

Young R M 1999 Analysis of a micromachine based vacuum pump on a chip actuated by the thermal transpiration effect. *J. Vac. Sci. Technol. B* **17(2)**, 280–7

Young R, Freidhoff C, Pollar D, Schiller P 1995 Micro-miniature piezoelectric diaphragm pump for the low pressure pumping of gases USP 5466932

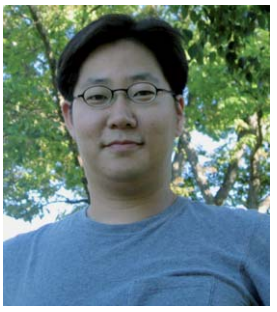
Yu H, Kim E S 2003 Micropropulsion of air and liquid jet by acoustic streaming. *Proc. IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '03)*, Kyoto, Japan, pp. 76–9

Zellers E, Steinecker W H, Lambertus G R, Agah M, Lu C-J, Chan H K L, Potkay J A, Oborny M C, Nichols J M, Astle A, Kim H S, Rowe M P, Kim J, da Silva L W, Zheng J, Whiting J J, Sacks R D, Pang S W, Kaviany M, Bergstrom P L, Matzger A J, Kurdak Ç, Bernal L P, Najafi K, Wise K D 2004 A versatile MEMS gas chromatograph for determinations of environmental vapor mixtures. In: *Proc. 11th Solid-State Sensor, Actuator, and Microsystems Workshop (Hilton Head '04)*, Hilton Head Island, SC, USA, pp. 61–6

Zengerle R, Kluge S, Richter M, Richter A 1995 A bidirectional silicon micropump. *Proc. 8th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS '95)*, Amsterdam, Netherlands, pp. 19–24

Zhang L, Kang J-M, Jiang L, Asheghi M, Goodson K E, Santiago J G, Kenny T W 2002 Measurements and modeling of two-phase flow in microchannels with nearly constant heat flux boundary conditions. *J. Microelectromech. Syst.* **11(1)**, 12–19

Biographies



Hanseup Kim received his BS degree in 1997 in Electrical Engineering from the Seoul National University, Korea, and his MS and Ph.D. degree in 2003 and 2006, respectively, in Electrical Engineering from the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. Since 2006, he has been employed as a research fellow in the Center for Wireless Integrated MicroSystems (WIMS), the Department of Electrical Engineering and Computer Science, University of Michigan. His research interests include MEMS, bio-microsystems, lab-on-a-chips, microfluidics, microsensors and actuators, micro/nano-fabrication, and polymer technology development.

He was the recipient of the 2001 Best Paper and the 1st place award in the 38th Student Design Contest of Design Automation Conference (DAC 2001) with four other coauthors. To date, he has authored and coauthored over 20 papers in refereed journals and conferences, held one patent, and served as a technical reviewer for the IEEE *Journal of Micro Electromechanical Systems (JMEMS)*, IEEE *Transactions on Electron Devices (TED)*, and *Sensors and Actuators Journal (SNA)*.

Khalil Najafi received his BS, MS, and Ph.D. degree in 1980, 1981, and 1986 respectively, all in Electrical Engineering from the University of Michigan, Ann Arbor. From 1986 to 1988 he was employed as a research fellow, from 1988 to 1990 as an assistant research scientist, from 1990 to 1993 as an assistant professor, from 1993 to 1998 as an associate professor, and since September 1998 as a professor in the Solid-State Electronics Laboratory, Department of Electrical Engineering and Computer Science, University of Michigan. His research interests include micromachining technologies, micromachined sensors, actuators, and MEMS; analog integrated circuits; implantable biomedical microsystems; micropackaging; and low-power wireless sensing/actuating systems.

Dr. Najafi was awarded a National Science Foundation Young Investigator Award from 1992 to 1997 and has been active in the field of solid-state sensors and actuators for more than 20 years. He has been involved in several conferences and workshops dealing with micro sensors, actuators, and microsystems, including the International Conference on Solid-State Sensors and Actuators, the Hilton-Head Solid-State Sensors and Actuators Workshop, and the IEEE/ASME Micro Electromechanical Systems (MEMS) Conference. Dr. Najafi has served as the editor for several journals, including the IEEE *Transaction on Electron Devices*, *Journal of Solid-State Circuits*, *Transactions on Biomedical Engineering*, and is currently an associate editor for the IEEE *Journal of Micro Electromechanical Systems*, the *Journal of Micromechanics and Microengineering*, Institute of Physics Publishing, and an editor for the *Journal of Sensors and Materials*. He is a Fellow of the IEEE and the AIBME.

Luis P. Bernal received an Aeronautical Engineering Degree in 1971 from the School of Aeronautics at the Polytechnical University of Madrid, Spain, and a Ph.D. in Aeronautics in 1981 from the California Institute of Technology. He also has a Doctorate in Aeronautical Engineering from the Polytechnic University of Madrid which he received in 1983. From 1981 to 1983, he was Member of Technical Staff in the Experimental Fluid Mechanics group at Jet Propulsion Laboratory (now JPL). In 1984 he joined the Faculty of Aerospace Engineering Department at the University of Michigan as an assistant professor, and he was promoted to his current position in 1992. Professor Bernal's research interests span a wide range of topics in Gas Dynamics, Aerodynamics, and Propulsion including the development of microsystems for propulsion, actuation, and flow control applications.

Dr. Bernal has served in numerous scientific panels from NSF, NASA, and AFOSR. He has served as a reviewer for the *Journal of Fluid Mechanics*, *Experiments in Fluids*, the *AIAA Journal*, and the *Journal of Fluids Engineering*, among other journals. He teaches undergraduate and graduate courses in Gas Dynamics and Aerodynamics.

2.09 Liquid Micropumps

Shuichi Shoji,[1] Hironobu Sato,[1] and Roland Zengerle[2], [1]Waseda University, Tokyo, Japan, [2]Department of Microsystems Engineering – IMTEK, University of Freiburg, Freiburg, Germany

Glossary

Electrohydrodynamic Fluid flow are generated by interaction of electric field and charges in the fluid. Due to the viscosity, liquid molecules around the charges are dragged; as a result, pumping force was induced. There are two types of EHD micropumps, EHD injection micropump, and traveling-wave EHD micropump.

Traveling wave Waves that are moving are called traveling waves. In the traveling-wave EHD micropumps, the electric field waves are produced by applying the phase-shifted rectangular pulses on the arrayed planar electrodes formed on the bottom of the pump channel.

Magnetohydrodynmic Lorentz force was utilized to pump the fluid. The pumping force was generated when an electric currents is applied across a channel filled with conducting solution in the presence of a perpendicular magnetic field. Both DC and AC current can be used to achieve the pumping of the liquid.

Electroosmosis The electric double layer was formed at the interface between the channel wall and the fluid. When a voltage is applied between two ends of the microchannel, the ions of the double layer are moved by the influence of the electric field. Due to the viscosity, the ions force the bulk liquid to move through the channel.

2.09.1 Introduction

Microfabrication technologies based on the photolithography have been applied to miniaturize the conventional chemical and biochemical analysis systems. A new field called micrototal analysis systems (μTAS) or lab-on-a-chip (LOC) started in the 1980s and was established in the 1990s. μTAS is now one of the most practical microelectromechanical systems (MEMS) for chemical and biochemical applications. μTAS realize very small necessary sample volume, fast response, and reduction of reagents, which are very useful for the chemical, biochemical, and medical analysis. Now this field includes chemical synthesis also, which enables efficient chemical reactions in the small-scale reactors. One of the features of μTAS is to use the efficiencies of microfluidics, which represents the behaviors of fluids in micrometer-scale channels (Gravenson *et al.* 1993). Since precise flow control on the order of microliters per minute or nanoliters per minute is required in this area, microflow devices of microvalves, micropumps, and microflow sensors play very important roles (Shoji 1998, Shoji and Esashi 1994). Many kinds of microfluidic devices have been developed over the past 20 years. Precision of the device structures has been improved with the remarkable progresses of MEMS technologies. The materials of the flow devices show large diversity. Si and glass were used in early stage while plastic is one of the trends considering the actual biochemical and medical applications. Simulation tools for microfluidic behaviors of these devices, computational fluid dynamics (CFD) software, are currently developing and are used to achieve the optimum design of the microflow devices.

Micropumps are one of the essential devices to realize precise flow control in microscale channels. Micropumps that have been developed so far are categorized into mechanical pumps and nonmechanical pumps. Mechanical micropumps have the electrically driven actuators and the mechanical structure to generate pressure and flow. Conventional pumps have been miniaturized with MEMS technologies using miniaturized or microfabricated actuators. Diaphragm-type and peristaltic-type micropumps have been developed from the early stage. On the other hand, various methods for pressure generation or flow generation that take into account the flow behaviors in microchannels have been proposed. Physical phenomena, for example, electrohydrodynamic (EHD) behavior, magnetohydrodynamic (MHD) behavior, electroosmosis, and chemical phenomena, for example, chemical and electrochemical gas generations, are used for the flow generations. Structures, principles, and features of various types of micropumps are categorized and described below.

2.09.2 Mechanical Pumps

Micropumps are one of the frontier MEMS devices and were fabricated by Si and glass micromachining in the 1980s (Shoji *et al.* 1989, Smits 1989, Van Lintel *et al.* 1988). Small disk-type and stack-type piezoelectric actuators were used to drive the pump in the early stage. Over the past 20 years, various actuators have been developed and used for driving the micropumps and active microvalves. Piezoelectric, electrostatic, electromagnetic, and thermopneumatic actuators are commonly used for these purposes. The performances of the mechanical micropump, such as maximum flow rate and output pressure, depend strongly on the features of the actuator. The actuator must be chosen considering the characteristics of the generated pressure, displacement and response time, etc. In fact, the sizes of the micropump are determined by the actuator used. Typical mechanical micropumps are a diaphragm-type (reciprocating) pump and a peristaltic-type pump (Shoji 1998). In the 1980s and the first half of the 1990s, micropumps consisting of Si and glass stacks were fabricated by anisotropic Si etching and anodic Si glass bonding. From the mid-1990s, plastic micropumps fabricated by molding have been developed considering the fabrication cost. First, the features of the actuators used for the micropump are described. Then structures and principles of the mechanical micropumps developed so far are reviewed.

2.09.2.1 Actuators

Many kinds of small-sized or microfabricated actuators have been developed since the MEMS field was established in the mid-1980s. The newly developed actuators have been applied for microflow control devices of microvalves and micropumps. Almost all possible principles of the actuators are proposed for these purposes. The performances of mechanical micropumps depend strongly on the features of the actuator. In fact, the sizes of these devices are determined by the actuator. The actuators used

for micropumps are classified into two categories: miniaturized conventional actuators and batch-fabricated microactuators (Shoji 1998, Shoji and Esashi 1994). Typical and realistic miniaturized actuators and batch-fabricated microactuators are described below.

2.09.2.1.1 Miniaturized conventional actuators

Miniaturized conventional actuators are fabricated individually and assembled on the microfabricated pump body. The pump body has to be designed considering the structure of the actuator. Assembly method of the actuator is one of the problems to obtain mass production of the micropumps or to realize the integration with other micro fluidic components.

2.09.2.1.1.(i) Piezoelectric Actuators using piezoelectric effect have been widely used from very early stage. Some kinds of miniaturized piezoelectric actuators of disk type (**Figure 1(a)**) and stack type (**Figure 1(b)**) are commercialized. Piezoelectric actuators consist of piezoelectric plates and metal electrodes. Since disk-type actuators are easily glued with resin to the pump body, the disk-type actuators of piezoelectric unimorph and piezoelectric bimorph are most commonly used. They show large deformation and fast response and reasonable generated pressure. High applied voltage of over 100 V, small necessary current, and small power consumption are the features of this actuator.

To generate large pressure, a stack structure of many piezoelectric plates and metal electrodes is proposed. Very large pressure greater than 3.5×10^5 hPa is obtained. The stroke to length of the actuator is smaller than 10^{-3}, which means that the stroke is on the order of 10 μm for the small-sized stack actuator.

2.09.2.1.1.(ii) Electromagnetic Electromagnetic actuator was realized with solenoid plunger as shown in **Figure 1(c)**. The force generated by this actuator depends on the applied current to the solenoid coil and on the number of turns. A large stroke is one of the advantages of this actuator. Since high current is necessary to generate high pressure, power consumption must be considered. A miniaturized electromagnetic actuator consists of a soft magnetic mass suspended by a spring beam, and an external coil is also fabricated. The generated pressure P_{em} is calculated using the following equation:

$$P_{em} = \frac{M_m}{A} \int \frac{\partial H_x}{\partial x}\, dx$$

where A, M_m, and H_x are the area, the magnetization of the mass, and the vertical component of the magnetic field produced by the external coil, respectively.

The other type of electromagnetic actuator is also possible as shown in **Figure 1(d)**. It consists of a deflectable membrane having a metal conductor and an external permanent magnet. A current applied to the conductor perpendicular to the magnetic field generates Lorentz force.

2.09.2.1.2 Batch-fabricated actuators

2.09.2.1.2.(i) Electrostatic Electrostatic actuators are the most widely used MEMS actuators. The principle of the electrostatic actuator is very simple as shown in **Figure 1(e)**. The electrostatically generated pressure P_{el} is calculated using the following equation:

$$P_{el} = 1/2\varepsilon_0 \left\{ \frac{V}{d_0 + \varepsilon_{0x} d_{0x}} \right\}^2$$

where ε_0, ε_{0x}, d_0, d_{0x}, and V are the dielectric constant of vacuum, relative dielectric constant of the isolating layer, gap between the electrode and isolating layer, the isolating layer thickness, and the applied voltage, respectively. The generated force is inversely proportional to the second power of the distance d_0. In actual uses, the displacement of the actuator is limited if large generated pressure is required.

2.09.2.1.2.(ii) Thermopneumatic A typical thermopneumatic actuator has a sealed pressure chamber and a movable diaphragm as shown in **Figure 1(f)**. The pressure chamber is filled with a gas or a liquid. The inner pressure is increased by electrical dissipation in a heater resistor incorporated into the chamber. Generated pressure P_{tp} in the chamber is given by the following equation:

$$P_{tp} = P_0 \exp\left(\frac{-L_0}{RT}\right)$$

where P_0, L_0, R, and T are the initial pressure, the latent heat of vaporization, a gas constant and the temperature, respectively. Large volume expansion in liquid phase to gas phase transfer is sometimes utilized for this actuator. The rising time of the actuator depends on the heat capacitance and the available power of the heater, while the relaxation time is determined by the

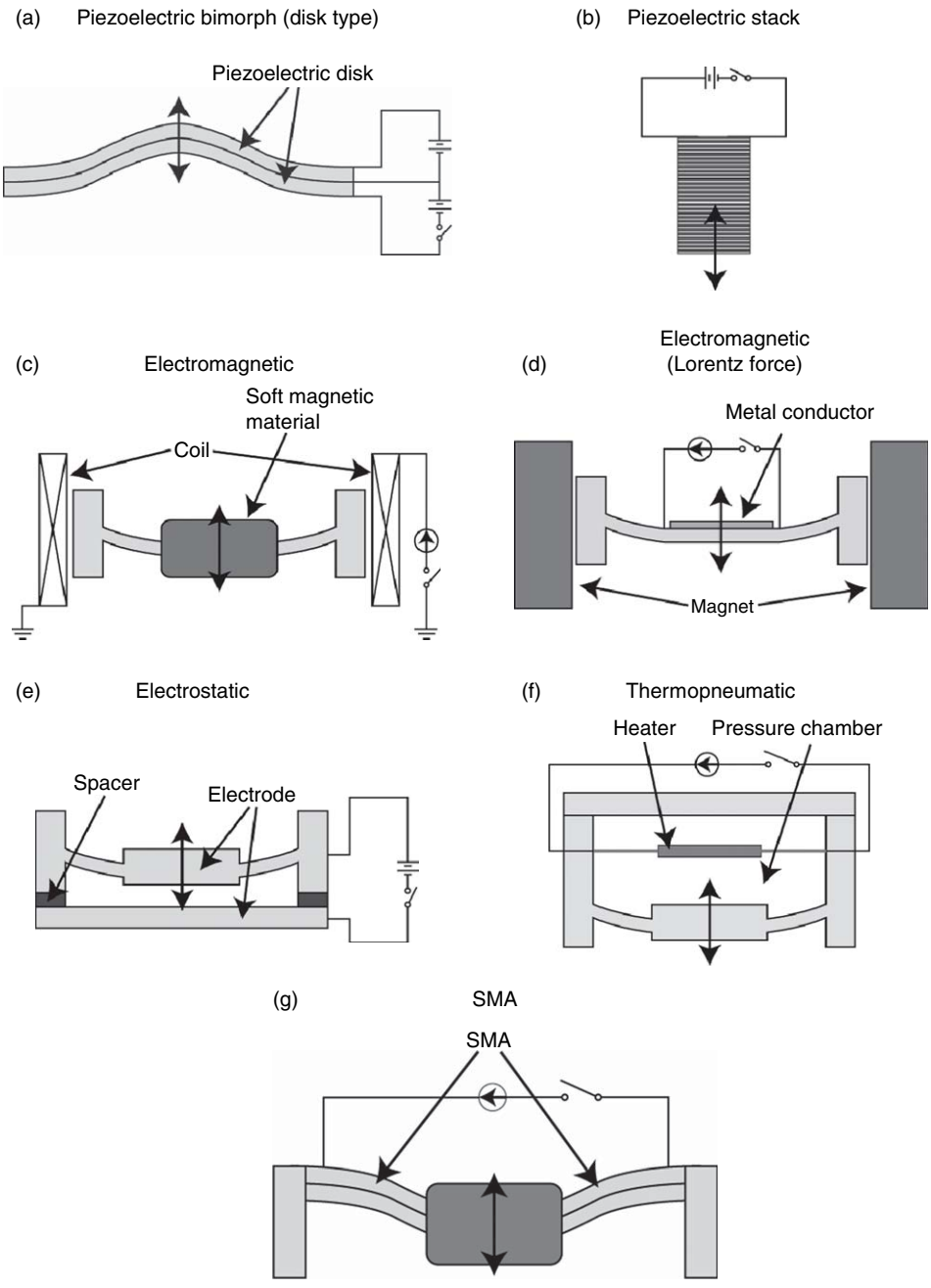


Figure 1 Schematic structure of the actuators used for the micropumps. (Source: Shoji S 1998 Topics in current chemistry. 194. Microsystem technology in chemistry and life science. In: Manz A, Becker H (eds.) *Fluids for Sensor Systems*. Springer-Verlag, Berlin, pp. 163–88.)

heat transfer to the outside. Since the small chamber volume has small heat capacitance, miniaturization of the actuator enables faster response time.

2.09.2.1.2.(iii) Shape memory alloy Small actuators are obtained by using a shape memory alloy (SMA) memorizing the expanded state. **Figure 1(g)** is the schematic of SMA actuators consisting of a SMA thin film and a diaphragm. The SMA regains its original shape when it is heated above the critical temperature. The critical temperature of Ti–Ni SMA is 40–50°C, which is achieved by the Joule

Table 1 Features of miniaturized conventional and batch-fabricated actuators

Actuators	*Pressure*	*Displacement*	*Response time*	*Figures*
Piezoelectric disk	Large	Medium	Fast	**Figure 1(a)**
Piezoelectric stack	Very large	Very small	Fast	**Figure 1(b)**
Electromagnetic (external coil)	Medium	Large	Medium	**Figure 1(c)**
Electromagnetic (Lorentz force)	Small	Medium	Medium	**Figure 1(d)**
Electrostatic	Very small	Very small	Very fast	**Figure 1(e)**
Thermopneumatic	Large	Large	Medium	**Figure 1(f)**
Shape memory alloy	Small	Large	Slow	**Figure 1(g)**

effect. To obtain a reasonable response time, a large current is required so that the power consumption should be considered. The general features of the miniaturized conventional actuators and the batch-fabricated microactuators are listed in **Table 1**.

2.09.2.2 Diaphragm-Type Micropump

2.09.2.2.1 Principle

The principle of the diaphragm-type micropump is illustrated in **Figure 2**. It consists of a thin flexible membrane (diaphragm) driven by an actuator, a pump chamber, and two passive check valves. Reciprocating the pump mode (**Figure 2(a)**) and the supply mode (**Figure 2(b)**) realizes continuous pumping. The actuator is driven by a sinusoidal wave or a rectangular wave AC voltage source. The displacement of the membrane (valuable volume, V_{dif}) causes the volume flow through the check valve (Zengerle *et al.* 1993). A pump chamber has a dead volume (V_{dead}). An important parameter that determines the pumping behavior is the ratio of dead volume to valuable volume (V_{dead}/V_{dif}) as shown in **Figure 3**. Small V_{dead}/V_{dif} realizes a self-priming and high bubble-tolerant micropump. The high-performance micropumps are obtained by minimizing the dead volume and maximizing the valuable volume (Kämper *et al.* 1998, Linnemann *et al.* 1998). The output pressure depends on the generated power of the actuator. The leakage flow of the check valve under reverse pressure applied state is also a factor for the maximum pump pressure. The flow rate of the micropump is determined by the frequency of the AC voltage as well as the displacement of the diaphragm. Usually, the flow rate is saturated with the increment of the frequency under the same applied voltage. The saturation is caused by not only the action of the diaphragm driven by the actuator but also the action of the check valve. If the dead volume exists around the check valve, it also affects the pumping behaviors.

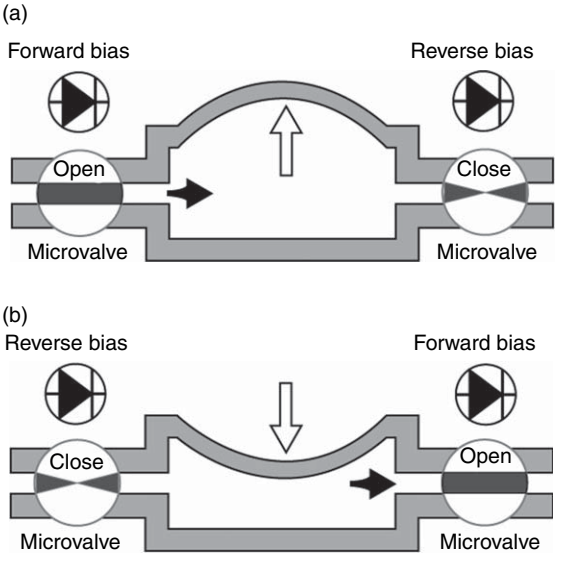


Figure 2 Principle of a diaphragm-type micropump.

2.09.2.2.2 Quasistatic simulation

One of the most common ways to model the performance of microdisplacement pumps with passive check valves is the quasistatic approach published in Zengerle *et al.* (1993) and Zengerle and Richter (1994). The approach allows the calculation of the

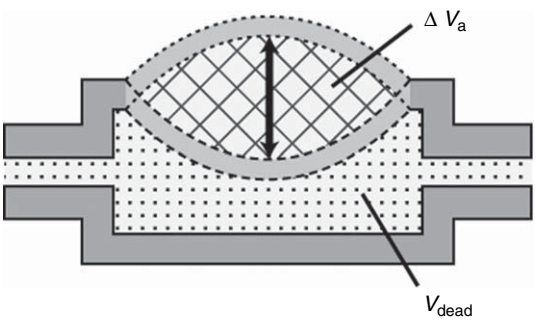


Figure 3 Schematic view of a variable pump chamber.

frequency dependence or back pressure dependence of the pump rate from the static characteristics of its building blocks, namely the displacement characteristics of the actuated membrane and the flow characteristics of the check valves. The result can be used not only to gain valuable insights for the optimization of microdisplacement pumps but also to understand malfunctions of it caused, e.g., by enclosed gas bubbles (Zengerle and Richter 1994) or potential leakage of the valves. A generalized schematic view of a microdisplacement pump with passive check vales is given in **Figure 4**. The pump consists of a flexible membrane, displaced periodically, and two passive check valves. When the membrane is displaced to enlarge the volume of the pump chamber, fluid is sucked in through the inlet valve. When the displacement decreases the chamber volume, fluid is pushed out through the outlet valve. The diode-like characteristics of the valves direct the fluid flow from the inlet to the outlet and prevent backflow.

At first sight these microdisplacement pumps just seem to be miniaturized versions of their macroscopic equivalents. Therefore, it is quite surprising that their characteristics in terms of frequency or back pressure–dependent pump rate are quite different. For instance, it is quite common for microdisplacement pumps that the pump rate is reduced linearly with the back pressure applied, a feature that is not known from macroscopic displacement pumps. The reason is that in most cases there is a strong coupling between the actuation of the displaced membrane, the back pressure between the outlet and the inlet, and the performance of the valves of miniaturized displacement pumps. Typically that is not the case in macroscopic pumps under standard operating conditions.

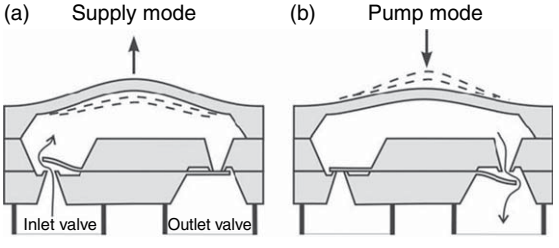


Figure 4 Schematic principle of a microdisplacement pump of the reciprocating type. (Source: Zengerle R, Richter M, Brosinger F, Richter A, Sandmaier H 1993 Performance simulation of microminiaturized membrane pumps. *7th Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, pp. 106–9; reproduced with permission.)

Many different designs of microdisplacement pumps have been presented so far. They differ in the actuation principle of the membrane and in the design of the valves. Flap valves (Oh and Ahn 2006), diaphragm valves (Shoji 1998), and even diffuser/nozzle elements without any movable parts (Oh and Ahn 2006) are typically used as flow-directing elements. On the other hand, piezoelectric actuation (Laser and Santiago 2004, Van Lintel *et al.* 1988), electrostatic actuation (Laser and Santiago 2004, Zengerle *et al.* 1992, 1995a, b), pneumatic actuation (Laser and Santiago 2004), or more specifically thermopneumatic actuation (Laser and Santiago 2004, van de Pol *et al.* 1989) have been reported among others (Laser and Santiago 2004) to drive the membrane. For generalization purpose, all actuation principles in the following are described by a characteristic actuation parameter A. The actuation parameter A could be a voltage, e.g., for electrostatic or piezoelectric actuation, a pressure in case of pneumatic or thermopneumatic actuation, or other physical parameters for other actuation principles. The parameter is in the next sections to deduce the dynamics of a microdisplacement pump in a generalized way.

2.09.2.2.2.(i) Calculation of the transient pressure within the pump chamber The time-dependent mass flow density $j(t)$ into the pump chamber is connected with the time-dependent change of the mass content $m(t)$ inside the pump chamber via the continuity in eqn [1].

$$-\int \vec{j}\,\vec{d}A = \frac{\mathrm{d}m}{\mathrm{d}t} \qquad [1]$$

The transient mass flow density can also be expressed as the difference in the rate of inlet and outlet flow through the check valves, which depend only on the time-dependent pressure drop $(p_1 - p(t))$ and $(p(t) - p_2)$ across the inlet and the outlet valve.

$$-\int \vec{j}\,\vec{d}A = \rho(\Phi_{\mathrm{inlet}}(p_1 - p(t)) - \Phi_{\mathrm{outlet}}(p(t) - p_2)) \qquad [2]$$

where p_1, p_2 are the hydrostatic pressure at the inlet/outlet side of the pump, $p(t)$ is the time-dependent hydrostatic pressure inside the pump chamber, and ρ is the density of the fluid.

On the right-hand side of eqn [1] the total mass m inside the pump chamber is given by the fluid mass m_{fluid} and the gas mass m_{gas}. If only fluid flows

through the valves and no degassing of the fluid occurs, time-dependent change in gas mass is zero.

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{\mathrm{d}m_{\mathrm{fluid}}}{\mathrm{d}t} + \frac{\mathrm{d}m_{\mathrm{gas}}}{\mathrm{d}t} = \frac{\mathrm{d}m_{\mathrm{fluid}}}{\mathrm{d}t} \qquad [3]$$

Therefore in the case of incompressible fluids, the time-dependent change in the fluid mass inside the pump chamber can be expressed as a corresponding volume change of the movable parts (**Figure 5**). This results in eqn [4]

$$\frac{\mathrm{d}m_{\mathrm{fluid}}}{\mathrm{d}t} = \rho\frac{d}{dt}(V_0 + V_{\mathrm{mem}}(p, A(t)) - V_{\mathrm{inlet}}(p, p_1) + V_{\mathrm{outlet}}(p, p_2) - V_{\mathrm{gas}}(p)) \qquad [4]$$

where V_0 is a constant offset, given by the total volume of the pump chamber, V_{inlet} and V_{outlet} are related with the volume displacements of the movable parts of the valves, and V_{gas} denotes the volume of a potential gas bubble inside the pump chamber.

The most important part of eqn [4] is given by the actuation characteristic $V_{\mathrm{mem}}(p, A)$ of the pump. $V_{\mathrm{mem}}(p, A)$ describes the coupling between the volume displacement of the membrane and the resulting pressure p inside the pump chamber at a specified actuation condition A. This function has to be derived considering a static equilibrium between all forces acting on the membrane (quasistatic model).

Combining eqns [4], [2], and [1] results in the basic differential equation [5] for the calculation of the time-dependent pressure p inside the pump chamber

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{(\Phi_{\mathrm{inlet}}(p_1 - p) - \Phi_{\mathrm{outlet}}(p - p_2)) - \frac{\partial V_{\mathrm{mem}}}{\partial A}\Big|_p \frac{\mathrm{d}A}{\mathrm{d}t}}{\frac{\partial V_{\mathrm{mem}}}{\partial p}\Big|_A - \frac{\mathrm{d}V_{\mathrm{inlet}}}{\mathrm{d}p} + \frac{\mathrm{d}V_{\mathrm{outlet}}}{\mathrm{d}p} - \frac{\mathrm{d}V_{\mathrm{gas}}}{\mathrm{d}p}} \qquad [5]$$

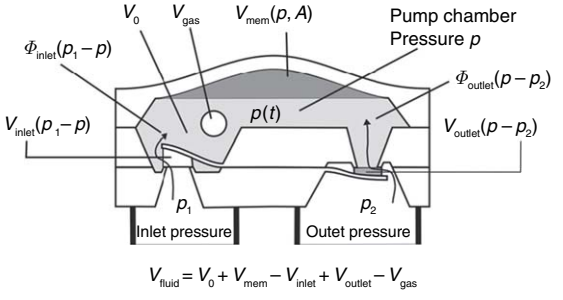


Figure 5 Schematic view of a microdisplacement pump. The characteristic functions of its building blocks are illustrated. (Source: Zengerle R, Richter M, Brosinger F, Richter A, Sandmaier H 1993 Performance simulation of microminiaturized membrane pumps. *7th Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, pp. 106–9; reproduced with permission.)

The characteristic functions of the basic building blocks of the microdisplacement pump can be determined by measurements or simulations, whichever is easier. For instance, the valve characteristics $\Phi_{\mathrm{inlet}}(p_1 - p)$, $\Phi_{\mathrm{outlet}}(p - p_2)$ typically can be easily determined by static measurements, whereas the actuation characteristic $V_{\mathrm{mem}}(p, A)$, the volume displacement of the valves $V_{\mathrm{inlet}}(p_1 - p)$, $V_{\mathrm{outlet}}(p - p_2)$, and the expansion characteristic of a gas bubble are more easily calculated.

The derivation of eqn [5] does not allow for compressibility and inertia of fluids. A more detailed study can be done within the same formalism, considering both effects that are important only at high pump frequencies (Zengerle and Richter 1994).

Eqn [5] can be easily integrated numerically by standard mathematical tools like Mathematica (Mathematica®, Wolfram Research, Inc., Champaign, IL). Once the transient pressure inside a pump chamber is known, all other physical parameters can be easily calculated. For instance, the transient volume flow through the inlet or the outlet valve can be easily calculated from the corresponding valve characteristics. The net flow can be calculated from the difference of outlet flow and inlet flow according to eqn [2], and the pumped fluid volume per pump stroke can be derived by numerical integration of the flow through one of the check valves during one pump cycle.

An alternative to the quasistatic approach presented before is the use of common tools for network analysis such as Saber (Saber®, Synopsys Inc., Mountain View, CA). In general, these tools are ideal to understand the behavior of complex machines that consist of several subsystems, which interact with each other. The basic principle behind network simulation is not very different from the quasistatic approach discussed before, but the differential equations should not be derived by hand. In contrast, a network model is set up just by combining different building blocks in a logical way. All the building blocks are represented by numerical functions – similar to the characteristic functions discussed before – and the network simulator creates the differential equation internally on its own. Once a model is implemented for a building block, it can be connected with other building blocks to build more complex systems. A fluidic channel, for instance, can be modeled as a flow resistance and a membrane can be modeled as a fluidic capacitance. All these building blocks can be connected to each other and the behavior of the total system (Litterst *et al.* 2005,

Voigt 2002) can be simulated without the need of the user to explicitly set up differential equations.

2.09.2.2.2.(ii) Example of a pneumatically actuated micropump The performance of the quasistatic approach can most easily be demonstrated by modeling the pneumatically operated microdisplacement pump depicted in **Figure 6**. In **Figure 7(a)** a typical measurement of the flow characteristic of passive check valves is depicted. The forward flow rate is much higher compared to the backward flow rate (leakage). The pressure-dependent volume displacement of the valve (**Figure 7(b)**) is assumed to be zero for the backward direction and increases linearly with the pressure difference in the forward direction.

The most important term is the actuation characteristic $V_{\text{mem}}(p, A)$. For pneumatic pumping the actuation parameter A is given by the time-dependent pneumatic pressure $p_a(t)$ used to actuate the membrane. In case of small displacements, the volume displacement of the membrane depends only on the net pressure $(p - p_a)$ acting on the membrane.

$$V_{\text{mem}}(p, A(t)) = C(p - p_a(t)) \qquad [6]$$

The constant factor C includes elastic and geometrical parameters of the membrane. **Figure 7(c)** shows the actuation characteristic for a pneumatically actuated microdisplacement pump, with a lateral membrane dimension of $5 \times 5\,\text{mm}^2$ and a thickness of a silicon membrane of 40 µm. An external system generates pressure pulses with an amplitude of $p_a^0 = -20\,000$ Pa (e.g., vacuum) relative to the atmosphere. Now, there are two characteristic functions specifying the equilibrium at the membrane. A first function when the external pressure is applied ($p_a(t) = p_a^0$) and a second function when it is not ($p_a(t) = 0$).

Starting at any pressure p_{start} within the pump chamber and at any volume displacement $V_{\text{mem}}(p_{\text{start}})$, $V_{\text{inlet}}(p_{\text{start}})$, ..., with a user-defined time characteristic of $p_a(t)$, solving the eqn [5] results in the time-dependent pump chamber pressure $p(t)$ after some cycles of iteration (**Figure 8**). As mentioned before, when $p(t)$ is known, other transient quantities can be easily derived.

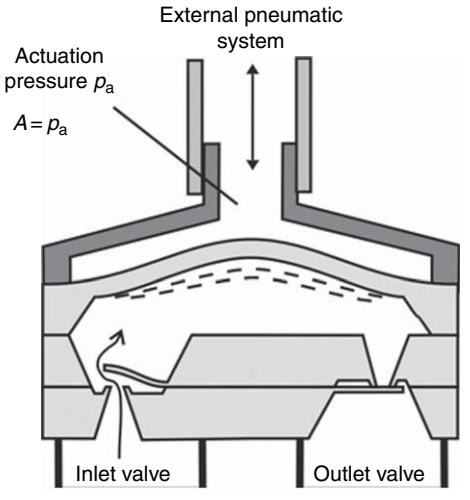


Figure 6 Schematic of a pneumatically actuated microdisplacement pump. The external pneumatic system applies a pneumatic pressure (here a vacuum) on top of the membrane. (Source: Zengerle R, Richter M, Brosinger F, Richter A, Sandmaier H 1993 Performance simulation of microminiaturized membrane pumps. *7th Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, pp. 106–9; reproduced with permission.)

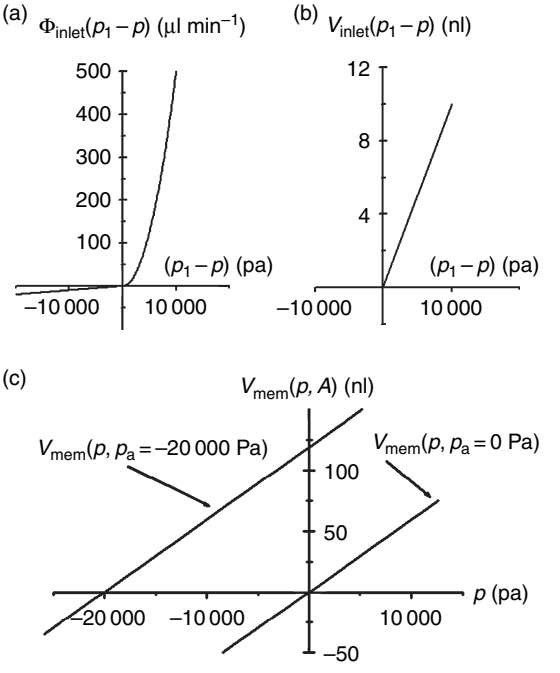


Figure 7 Characteristic functions of the building blocks of a pneumatically actuated microdisplacement pump: (a) characteristic function of the flow rate through a check valve; (b) characteristic function of the volume displacement of a check valve; (c) characteristic function of the actuated membrane. (Source: Zengerle R, Richter M, Brosinger F, Richter A, Sandmaier H 1993 Performance simulation of microminiaturized membrane pumps. *7th Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, pp. 106–9; reproduced with permission.)

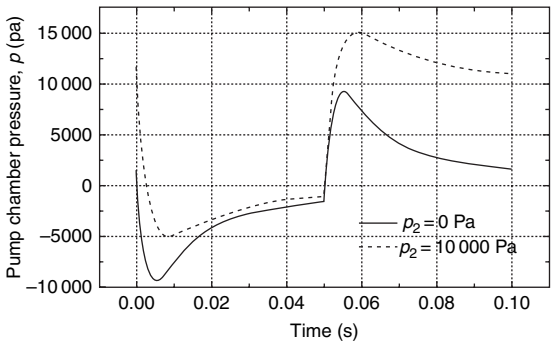


Figure 8 Transient simulation of the pump chamber pressure p for pneumatic actuation. (Source: Zengerle R, Richter M, Brosinger F, Richter A, Sandmaier H 1993 Performance simulation of microminiaturized membrane pumps. *7th Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, pp. 106–9; reproduced with permission.)

Very interesting features of the pump cycles can be understood by the diagram of the volume displacement of the membrane via the pump chamber pressure p (**Figure 9**). A parallelogram (graph a) encloses the possible states of the system. Two significant points of the parallelogram are given by the volume displacement of the nonactuated membrane at the outlet pressure p_2 ($V_{\mathrm{mem}}(p_2, A=0)$) and the volume displacement of the actuated membrane at the inlet pressure p_1 ($V_{\mathrm{mem}}(p_1, A=p_{\mathrm{a}}^0)$).

First, the easiest case (negligible volume displacement of the valves and low pump frequencies) is explained in order to demonstrate the usefulness of this diagram. The pump cycle starts with a nonactuated membrane at a point $V_{\mathrm{mem}}(p_2, A=0)$. Switching on the actuation pressure p_{a}^0 within a very short time constant ε ($\Delta V_{\mathrm{mem}}=0$) the system changes to a point α at the actuated membrane characteristic resulting in a low pressure $p=-p_{\mathrm{a}}^0$. Now the supply mode (**Figure 4(a)**) of the pump starts. Fluid flows through the inlet valve according to the valve characteristic until the inlet pressure p_1 is reached (point $V_{\mathrm{mem}}(p_1, A=p_{\mathrm{a}}^0)$). If the actuation is switched off the restoring force in the deflected membrane drives the system to a point β resulting in an overpressure in the pump chamber. The pump mode (**Figure 4(b)**) starts and fluid flows through the outlet valve until outlet pressure p_2 is reached (point $V_{\mathrm{mem}}(p_2, A=0)$). The volume stroke of the membrane within one cycle can be easily determined from the curve by $V_{\mathrm{mem}}(p_1, p_{\mathrm{a}}^0) - V_{\mathrm{mem}}(p_2, 0)$. From the same diagram the amplitudes of the pump chamber pressure p can also be extracted (see also **Figure 8**).

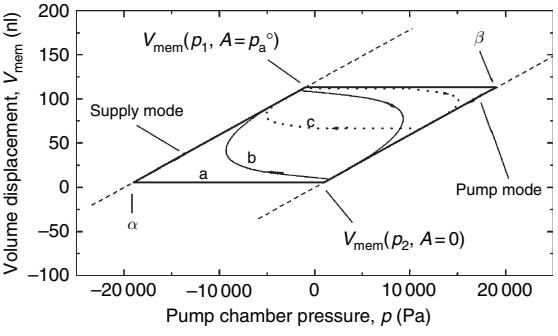


Figure 9 Pressure-dependent volume displacement curves of the pump membrane for pneumatic actuation: (a) simulation for negligible volume displacement of the valves, low pump frequencies and zero back pressure; (b) volume displacement of the valves are switched on; (c) back pressure increased to 10 000 Pa. (Source: Zengerle R, Richter M, Brosinger F, Richter A, Sandmaier H 1993 Performance simulation of microminiaturized membrane pumps. *7th Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, pp. 106–9; reproduced with permission.)

Operating with more realistic volume displacements of the valves and time-dependent actuation pressure $p_{\mathrm{a}}(t)$ the pump cycle is represented by the graph b, resulting especially in smaller amplitudes of pump chamber pressure p. By varying the outlet pressure p_2 from 0 to 10 000 Pa the trajectory changes to graph c, which corresponds to a displacement of the significant point $V_{\mathrm{mem}}(p_2, 0)$. It can easily be deduced from the diagram that the volume stroke of the membrane becomes smaller. This way the linear reduction of the volume stroke or the pump rate with the increasing back pressure, a feature that is quite common with micro displacement pumps, can be easily understood.

A comparison between the simulated- and the measured frequency-dependent pump rates is shown in **Figure 10**. The pneumatic actuated microdisplacement pump had outer dimensions of 7 mm × 7 mm × 2 mm, an active membrane with the lateral dimension of 4 mm × 4 mm, and a thickness of 30 µm. The pressure pulses for the actuation of the membrane had an amplitude of $p_{\mathrm{a}}^0 = -12\,500$ Pa (vacuum) relative to atmosphere. The time constant for switching on and off the external actuation pressure was in the range of 3 ms. Very good agreement between measurement and simulation is given for actuation frequencies of up to 50 Hz. At higher frequencies, the above-mentioned effects of inertia of the fluid have to be considered. The same formalism can also be applied to other actuation principles. For instance, it has been done for electrostatic actuation in Zengerle *et al.* (1993).

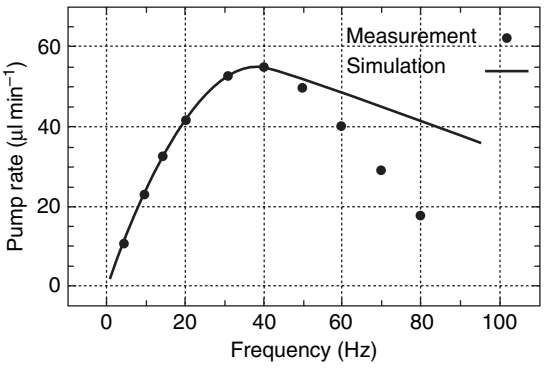


Figure 10 Pump rate of a pneumatically operated microdisplacement pump as a function of the driving frequency: simulation and experimental results according to Zengerle *et al.* (1993). At frequencies above 50 Hz, inertial effects result in divergences between experiments and simulation that are studied by an adopted model in more details in Zengerle and Richter (1994); reproduced with permission.

In summary, the performance of microdisplacement pumps can be easily derived by the quasistatic approach described here. All building blocks of the pumps (actuation unit, valves, etc.) are modeled by characteristic functions (V_{mem}, V_{inlet}, Φ_{inlet}, ...) measured or simulated in the case of its static equilibrium. The dynamic interaction of the building blocks can easily be modeled by applying the continuity equation [1]. A more detailed study as described here has to be used when the compressibility of the fluid plays a major role, and the inertia of the fluid cannot be neglected anymore at frequencies above 50 Hz (Zengerle and Richter 1994). In that case network simulation tools are more adequate to simulate the performance of microdisplacement pumps due to the fact that they are more user-friendly (Litterst *et al.* 2005, Voigt 2002).

2.09.2.2.3 Check valves

2.09.2.2.3.(i) Mechanical check valve Pumping behaviors are determined by the action of the check valves. Behavior of the check valve has symmetry of that of the semiconductor diode. Many types of mechanical check valves have been developed over the past 20 years. **Figure 11** illustrates the structure and the principle of the typical microcheck valves. Small leakage under reverse applied pressure and large reverse-to-forward flow resistance ratio are required for high-performance check valves. Tangent time during open-to-close or close-to-open is also an important factor of the check valve (Shoji 1998). **Figure 11(a)** shows a bulk Si ring-type check valve, which is the first developed check valve for the diaphragm micropump (Van Lintel *et al.* 1988). This check valve has a complicated structure and some dead volumes. The response time is slow compared to the other valves. A bulk Si cantilever is also used as the check valve (**Figure 11(d)**) (Zengerle *et al.* 1992, 1995a). A square ridge structure is formed under the cantilever. This valve shows high performance in large reverse-to-forward flow resistance ratio and short response time. This type of check valve has been widely used due to its simple structure and easy fabrication. A disk suspended by four thin beams made of poly-Si or SU-8 photoresist as shown in **Figure 11(b)** is the other commonly used microvalve (Shoji *et al.* 1989, Truong and Nguyen 2004). Stable actions by the symmetrical structure are the feature of this type. For simplifying the valve structure, polymer membrane–type check valve as shown in **Figure 11(e)** was developed (Rapp *et al.* 1994). This type of valve has been used for plastic micropumps. Reasonable performances are obtained even for its simple structure. Mass production is available with the plastic molding. A ball valve as shown in **Figure 11(c)** is also utilized as a check valve (Yamahata *et al.* 2004, 2005a). A freestanding ball is used as a plug for the through-hole. All micropumps using mechanical check valves are free from the backflows to the inlet and from the outlet. For high-pressure micropump, active microvalves are used instead of the passive check valve (Bodén *et al.* 2005). Thermopneumatic actuators are used for two active check valves and one valuable chamber. This type of micropump has similar flow behaviors to those of the peristaltic micropumps and are sometimes categorized to the peristaltic micropump.

2.09.2.2.3.(ii) Nonmechanical check valve A microfabricated nozzle/diffuser element is used instead of the mechanical check valves used in some micropumps (Stemme and Stemme 1993). A small-angle 2D flow channel, a diffuser (**Figure 12(a)**), and a large-angle 2D flow channel, a nozzle (**Figure 12(b)**), are used as the check valves. One-way motion of this valve is obtained only under the dynamic fluidic actuations. The geometry of these elements is simple but their flow behaviors are very complicated. Simulation studies are carried out for optimizing the nozzle/diffuser design (Olsson *et al.* 2000). Forward flow (small flow resistance) and reverse flow (large flow resistance) of the diffuser element and the nozzle element are opposite as

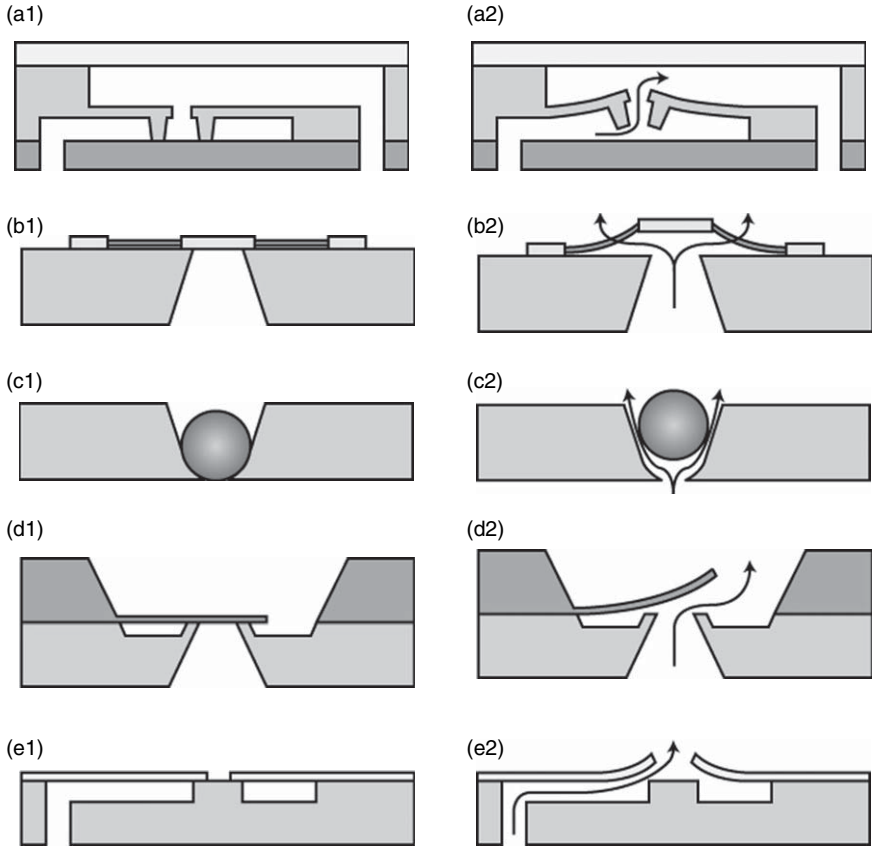


Figure 11 Structures and principles of the micro check valves. (Sources: Van Lintel H T G, van de Pol F C M, Bouwstra 1988 A piezoelectric micropump based on micromachining of silicon. *Sens. Actuators* **15**, 153–67; Shoji S, Nakagawa S, Esashi M 1989 Micropump and sample-injector for integrated chemical analyzing systems. *Sens. Actuators* **A21**–**A23**, 189–92; Zengerle R, Ulrich J, Kluge S, Richter M, Richter A 1995 Bidirectional silicon micropump. *Sens. Actuators* **A50**, 81–6; Rapp R, Schomburg W K, Maas D, Schulz J, Stark W 1994 LIGA micropump for gas and liquid. *Sens. Actuators A* **40**, 57–61; Yamahata C, Lacharme F, Burri Y, Gijs M A M 2005 A ball valve micropump in glass fabricated by powder blasting. *Sens. Actuators B* **110**, 1–7.

shown in **Figure 12**. For a microfabricated diffuser element, typical optimum angle realizing high performances is around 7°, while that of a nozzle element is around 70° for a microfabricated nozzle element.

Liquid viscosity change in narrow microchannels, orifices, caused by the temperature change is also used for the micropump (Matsumoto *et al.* 1999). Low viscosity at a high temperature generates larger flow as illustrated in **Figure 13**. Microheaters are formed at the inlet and outlet orifices of the variable pressure chamber of the micropump.

2.09.2.2.4 Examples of the diaphragm micropumps

2.09.2.2.4.(i) Piezoelectric actuators In practice, piezoelectric actuators have been used in high-performance micropumps. A disk-type piezoactuator is commonly used for its easy assembly and large displacement. Structures of the self-priming and bubble-tolerant micropumps using piezoelectric disk as the actuator are illustrated in **Figure 14** (Linnemann *et al.* 1998). The stacked silicon-type micropump (**Figure 14(a)**) has silicon cantilever-type check valves. A maximum pump rate of 1 ml min^{-1} under high-frequency operation of up to 200 Hz and a large back pressure of 1000 hPa are obtained. Its simple structure enables wafer-level batch fabrication, which is suitable for mass production (Richter *et al.* 2001). A low-cost plastic micropump (**Figure 14(b)**) is also realized by a molding process (Kämper *et al.* 1998). The use of polycarbonate membrane check valves enables a very simple structure, which is suitable for mass

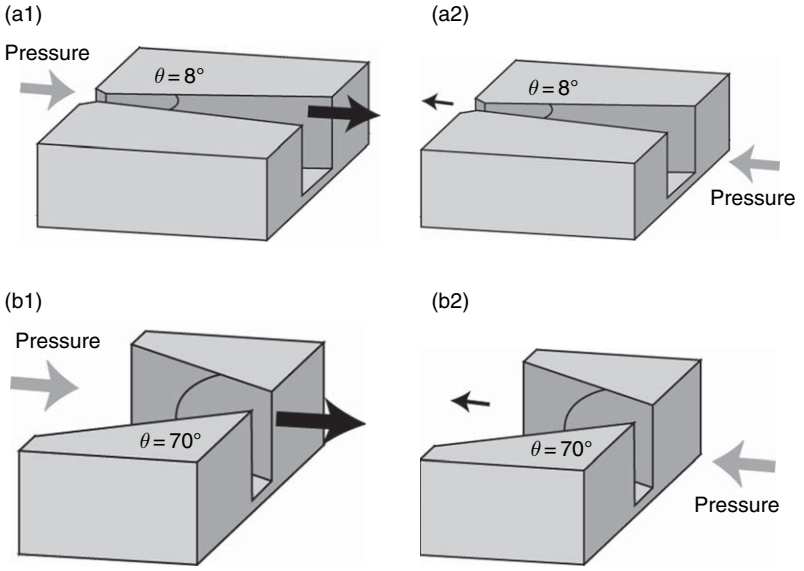


Figure 12 Flow behaviors of the diffuser element and the nozzle element. (Sources: Stemme E, Stemme G 1993 A novel piezoelectric valve-less fluid pump. *7th Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, pp. 110–13; Olsson A, Stemme G, Stemme E 2000 Numerical and experimental studies of flat-walled diffuser elements for valve-less micropumps. *Sens. Actuators A* **84**, 165–75.)

production. The ratio of low dead volume to valuable volume enables self-priming pumping. A maximum pump rate of 400 µl min^{-1} and a maximum back pressure of 2100 hPa are obtained at a driving frequency of 70 Hz. Fully polymeric micropumps use SU-8 disk-type check valves and poly(methyl methacrylate) PMMA stacks (Nguyen and Truong 2004). This structure was designed to achieve easy assembly and integration with other fluidic devices. Parylene dome diaphragm driven by a sputter-deposited piezoelectric ZnO film actuator is also applied to the micropump (Feng and Kim 2005). In this case, the piezoelectric actuator is fabricated by sputter deposition. Maximum output pressure of 120 Pa and maximum flow rate of 1 µl/min was obtained under driving frequency of 30 kHz.

Many types of valveless micropumps utilizing micro diffuser/nozzles have been developed. These types of micropumps using diffuser elements were fabricated by Si–glass micromachining and thermoplastic molding in the mid-1990s (Olsson *et al.* 1998, van der Wijngaart *et al.* 2000). A valveless structure prevents stacking of the particles observed at the mechanical check valve, and a double chamber–type micropump as shown in **Figure 15** was also reported. These micropumps use the 2D diffuser/nozzle structure. Precise numerical and experimental studies for these types of micropumps were reported (Olsson *et al.* 2000). A maximum back pressure of 160 hPa and a maximum pump rate of 1.2 ml min^{-1} were obtained. Truncated, pyramid through-holes realized by Si anisotropic wet etching are also used for the nozzle elements (Gerlach *et al.* 1995). These nozzle structures are easily fabricated by silicon anisotropic wet etching. The angle of the nozzle structure in this case is around 35.3°. A maximum pump rate of 1.5 ml min^{-1} is realized. A micropump consists of a screen-printed PZT actuator and these nozzle elements are fabricated (Schabmueller *et al.* 2002). A pump rate of 155 µl min^{-1} and a maximum back pressure of 10 hPa were obtained at a driving voltage of 600 V.

2.09.2.2.4.(ii) Thermopneumatic actuators

Thermo-pneumatic actuators are also commonly used in micropumps. Silicon and glass stacks having a hermetically sealed chamber with microheater and bulk were fabricated (Mathematica, Wolfram Research, Inc., Champaign, IL). Plastic micropumps using membrane-type check valves were also reported (Büstgens *et al.* 1994). A low-voltage and high-current power supply is required compared to the piezoelectric actuators. A maximum back pressure of 38 hPa and a maximum pump rate of 44 µl min^{-1} were realized under an average power consumption of 0.45 W. A high-performance stacked silicon micropump fabricated by sophisticated

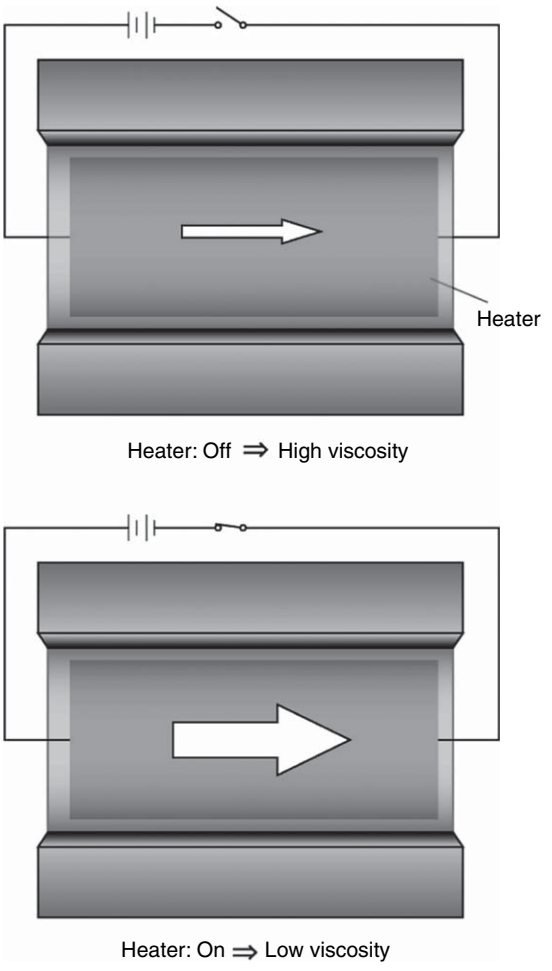


Figure 13 Flow behaviors of the thermal viscosity change. (Source: Matsumoto S, Klein A, Maeda R 1999 Development of bi-directional valve-less micropump for liquid. *12th IEEE Int. Conf. Micro Electro Mechanical Systems*, Orlando, FL, USA, pp. 141–6.)

MEMS processes with silicon-on-insulator (SOI) substrates was reported (Maillefer *et al.* 2001). Precisely fabricated microstructure enables linear and accurate pumping characteristics. The pumping behaviors are insensitive to external conditions. High reliability of this micropump will satisfy the specifications required in a drug delivery system. A valveless micropump using diffuser element was fabricated with a polydimethylsiloxane (PDMS) structure (Kim *et al.* 2005). A simple structure and a low-cost fabrication are suitable for disposable uses. Thermal bubble generation in a pump chamber was also used as the actuator (Tsai and Lin 2001). A diffuser–nozzle micropump was realized with very simple silicon–glass stack in this case.

2.09.2.2.4.(iii) Electromagnetic actuators

Micropumps using electromagnetic actuators were also reported. One is used as a NdFeB magnet and a miniaturized coil actuator. The membrane-type check valves are used (Böhm *et al.* 1999a). A maximum pump rate of 2 ml min^{-1} and a back pressure of 125 hPa were obtained. A micropump consists of a similar electromagnetic actuator, and two ball-type check valves were also developed (Yamahata *et al.* 2005a). Housing of the ball valve is fabricated by powder blasting. A maximum pump rate around 5 ml min^{-1} and a maximum back pressure about 280 hPa were demonstrated.

2.09.2.2.4.(iv) Other types of actuators

A stacked silicon micropump consists of an electrostatic actuator, and two cantilever-type check valves were fabricated (Zengerle *et al.* 1992). Low power consumption and high-frequency operation are the features of this micropump. A maximum pump rate of 350 µl min^{-1} and a maximum back pressure of 310 hPa were obtained under a driving voltage of 200 V at 400 Hz. A micropump using micromachined TiNi shape memory alloy as the actuator was also reported (Makino *et al.* 2001). Si cantilever-type check valves and Si stuck structure were utilized in this case. A low-power and low-voltage-driven micropump was realized by continuous electrowetting (CEW) as an actuator (Yun *et al.* 2002). The surface tension–induced motion of a mercury drop is used for the actuator. This pump has two diaphragms located at the pump inlet and outlet where two Cu cantilever-type check valves are positioned. Input and output pressures inside the micropump cavity are controlled by the CEW actuator. A pump rate of 70 µl min^{-1} pump rate with a driving voltage of 2.3 V and a power consumption of 170 µW was demonstrated. A back pressure of about 800 Pa at an applied voltage of 2.3 V was also obtained.

2.09.2.3 Peristaltic Micropumps

One of the oldest micropump is peristaltic-type pump, which consists of piezoelectric disk actuator microvalves connected in series (Smits 1989). **Figure 16** shows the structure and the actuation principle of the thermopneumatically driven micropump (Grosjean and Tai 1999). An Au microheater is incorporated in the sealed cavity with 100% liquid fill. Considering the trade-off between power consumption and response time, a kind of Fluorinert™ is chosen as the liquid (Grosjean *et al.* 1999). Pumping is

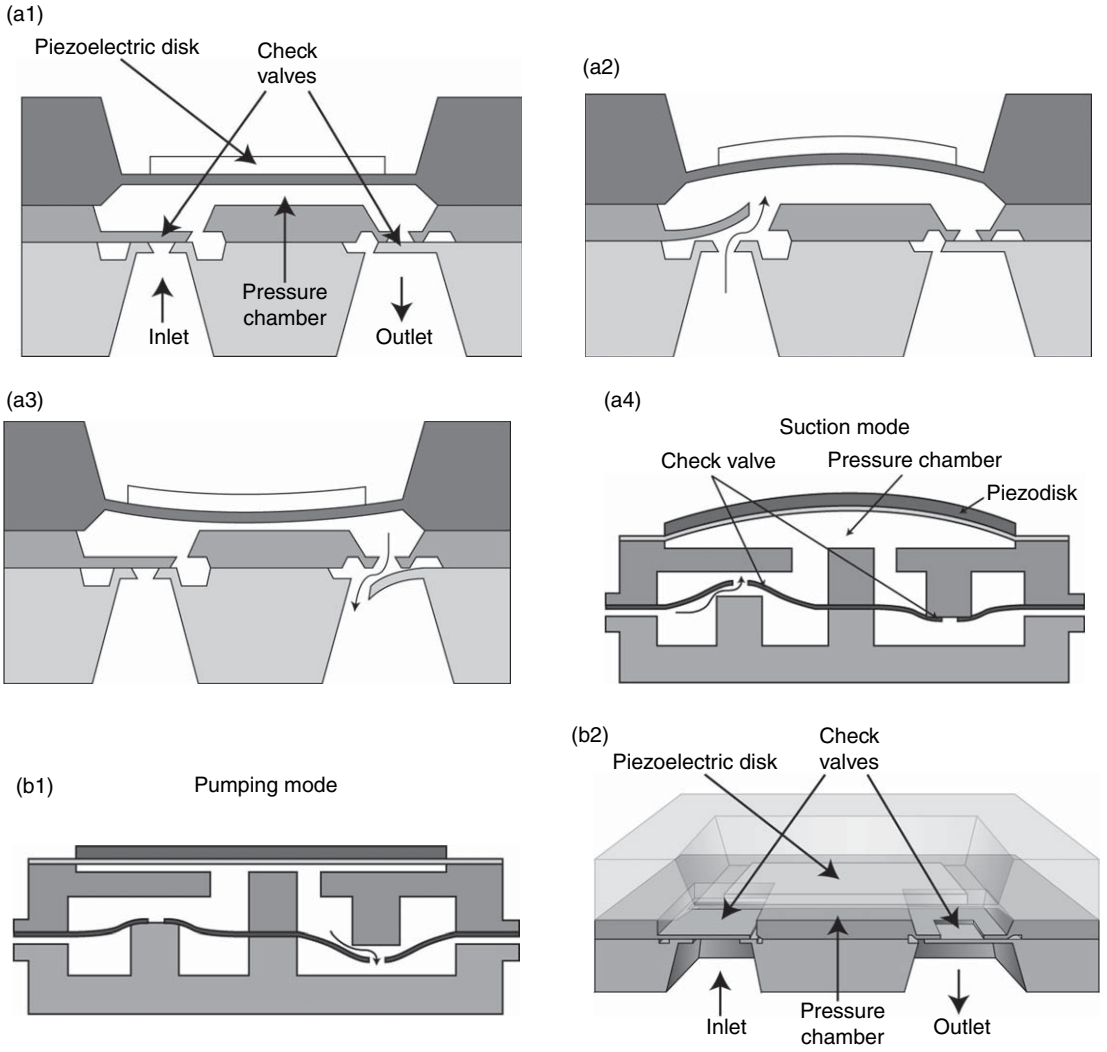


Figure 14 Typical diaphragm micropumps using a piezoelectric actuator. (Sources: Linnemann R, Woias P, Senfft C-D, Ditterich J A 1998 A self-priming and bubble-tolerant piezoelectric silicon micropump for liquids and gases. *11th Annu. Int. Workshop Micro Electro Mechanical Systems*, Heidelberg, Germany, pp. 532–7; Kämper K-P, Döpper J, Ehrfeld W, Oberbech S 1998 A self-filling low-cost membrane micropump. *11th Annu. Int. Workshop Micro Electro Mechanical Systems*, Heidelberg, Germany, pp. 432–7.)

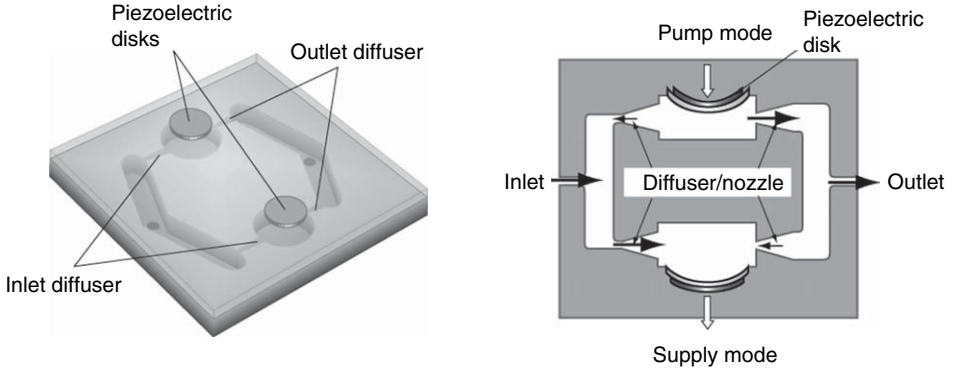


Figure 15 Valveless diaphragm micropump using diffuser elements. (Source: Olsson A, Larsson O, Holm J, Lundbladh L, Öhman O, Stemme G 1998 Valve-less diffuser micropumps fabricated using thermoplastic replication. *Sens. Actuators A* **64**, 63–8.)

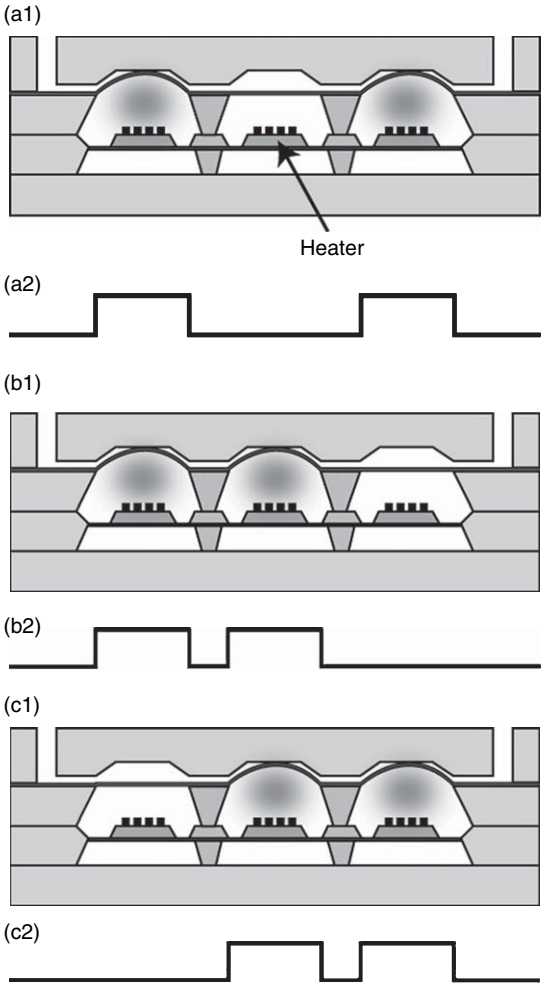


Figure 16 Peristaltic micropumps using thermopneumatic actuators. (Source: Grosjean C, Tai Y C 1999 A thermopneumatic peristaltic micropump. *10th Int. Conf. Solid-State Sensors and Actuators*, Sendai, Japan, Vol. 2, pp. 1776–9.)

performed by the phase-sifted sequential rectangular voltage. Since the pump has a symmetrical structure, bidirectional pumping is easily obtained. Self-priming operation with a flow rate of up to 6.3 µl min^{-1} was demonstrated with a power consumption of 291 mW. Three thermopneumatic actuator–type PDMS micropumps and three electrostatic actuator–type micropumps were developed (Jeong *et al.* 2005, Teymoori and Abbaspour-Sani 2005). These micropumps were designed for the disposable applications like drug delivery. A unique micropump using a magnetic field rotary pneumatic actuator was also proposed (Kim *et al.* 2006). For peristaltic actuation, magnetic fluids are generated in round-shaped channel by magnetic force. The magnetic force is applied by the stepping motor–controlled permanent magnet. Maximum pump rates of 2.8 µl min^{-1} at 4 rpm and 3.8 µl min^{-1} at 8 rpm were obtained.

2.09.2.4 Other Micropumps

A bidirectional micropump consisting of an active valve and an elastic buffer element driven by a piezoelectric actuator as shown in **Figure 17** was proposed (Zengerle *et al.* 1995a). Pumping directions are changed with the driving frequency. Gear pumps are one of the common pumping systems. Microgear pumps were fabricated by LIGA process and precision engineering (Döpper *et al.* 1997). In this case, an outside rotary motor is used to drive the gear. This type of micropump is useful to pump high-viscosity liquids.

2.09.3 Nonmechanical Micropumps

2.09.3.1 EHD Micropumps

In EHD pumping, fluid forces are generated by the interaction of electric fields and charges in the fluid. There are two types of the EHD micropumps: the high-voltage DC charge EHD injection micropump and the traveling wave–driven EHD micropump using induced charges in the bulk liquid.

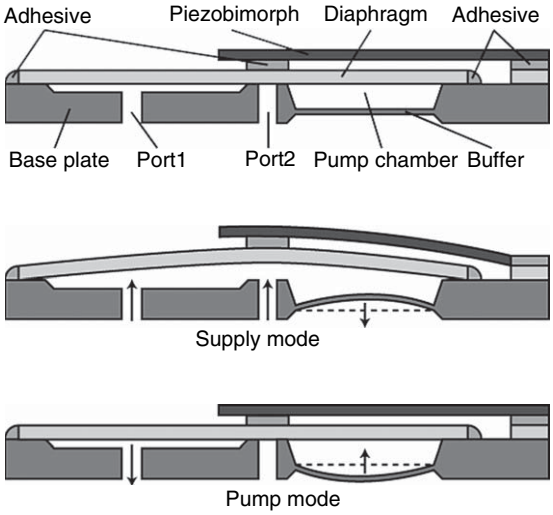


Figure 17 Bidirectional micropump using an active microvalve and an elastic buffer element. (Source: Zengerle R, Ulrich J, Kluge S, Richter M, Richter A 1995a Bidirectional silicon micropump. *Sens. Actuators* **A50**, 81–6.)

2.09.3.1.1 EHD injection micropumps

In EHD injection pumps, the motive force is the Coulomb force on ions injected from the electrode into the fluid by electrochemical reactions. EHD injection pumps required two facing electrodes (**Figure 18**) (Richter and Sandmaier 1990). High DC voltage was applied between the two electrodes. Due to the electrochemical reactions, the ions were injected from the emitter electrode into the liquid. The injected ions are subjected to an electric field and drag the liquid molecules around them. As a result, the bulk liquid motion is induced. Pumping is not restricted to DC fields. AC fields of relatively low frequencies are also available. Recently, 3D triangular bumps of solder were utilized as the electrode to maximize the electric field gradient (Darabi *et al.* 2002). The pumped fluids are restricted to the liquids of extremely low conductivity such as organic solvents.

2.09.3.1.2 Traveling wave micropumps

In traveling wave EHD pump, the free charges induced in the bulk liquid were utilized to drive the fluid. In general, application of the electric field in the presence of the gradient in permittivity or conductivity induces free charge in the volume of a material. The electrical conductivity of slightly conducting liquids depends on the temperature. Therefore, the gradient of the electrical conductivity can be achieved by developing the temperature gradient in the liquid. **Figure 19** illustrates an EHD pump using a temperature-induced conductivity gradient (Bart *et al.* 1990). The temperature and the conductivity gradient allow free charges to develop in the fluid volume. These charges are subjected to electric field and drag the liquid molecules around them. As a result, pumping force is induced. **Figure 20** shows the schematic diagram of the traveling wave EHD pump. The planer electrodes are formed on the bottom of the pump channel. The electric field waves are produced by applying the phase-shifted rectangular pulses on the electrodes. The temperature gradient is produced by the traveling wave itself and not by additional heating (Fuhr *et al.* 1992). In contrast to the EHD injection pump, relatively conductive liquids such as water solutions could be pumped.

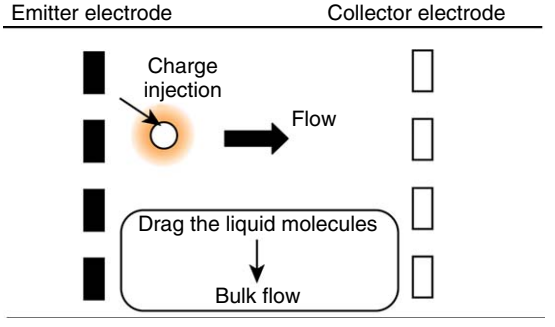


Figure 18 Schematic diagram of the electrohydrodynamic (EHD) injection pumps.

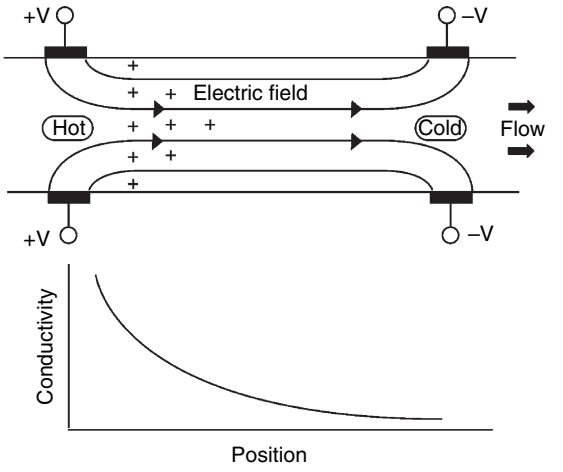


Figure 19 An electrohydrodynamic (EHD) pump employing a temperature-induced conductivity gradient. (Source: Bart S F, Tavrow L S, Mehran M, Lang J H 1990 Microfabricated electrohyrdrodynamic pumps. *Sens. Actuators* **A21–A23**, 193–7.)

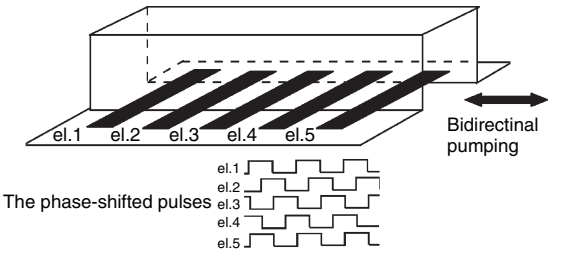


Figure 20 Schematic drawing of the traveling wave electrohydrodynamic (EHD) micropump. (Source: Fuhr G, Hagedorn R, Muller T, Benecke W, Wagner B 1992 Microfabricated electrohydrodynamic (EHD) pumps for liquids of higher conductivity. *J. Microelectromech. Syst.* **1(3)**, 141–6.)

2.09.3.2 MHD Micropumps

The pumping mechanism for a MHD pump results from the Lorentz force. This force is produced when an electric current is applied across a channel filled with conducting solution in the presence of a perpendicular

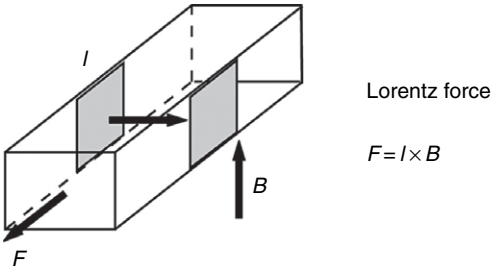


Figure 21 Vector diagram of a magnetohydrodynamic (MHD) pump. (Source: Lemoff A V, Lee A P 2000 An AC magnetohydrodynamic micropump. *Sens. Actuators B* **63**, 178–85.)

magnetic field. The Lorentz force is perpendicular to the current both in the channel and in the magnetic field. **Figure 21** shows the vector diagram of the MHD pump (Lemoff and Lee 2000). The flow along the length of the channel resulting from the Lorentz force has a profile similar to that of pressure-driven flow. Both the DC and the AC current can be utilized to produce the Lorentz force.

2.09.3.2.1 DC MHD micropumps

In a DC MHD micropump, a DC current is applied across the channel in the presence of a uniform magnetic field from a permanent magnet. Due to the electrolysis of the conductive solution, the bubble was generated and the pumping performance was degraded (Jang and Lee 2000). In order to avoid disruption of flow due to bubble formation, a frit-like structure was reported (Homsy *et al.* 2005). In this structure, the electrode and the pumping channel were separated; as a result, the high current density DC operation was achieved without gas bubble fluctuation.

2.09.3.2.2 AC MHD micropumps

In an AC MHD micropump, a sinusoidal electric current is applied across the microchannel in the presence of a perpendicular, synchronous AC magnetic field induced by an electromagnet (Lemoff and Lee 2000). Since the fields are synchronous, the Lorentz force point in the same direction throughout the cycle. The AC MHD micropump produces a continuous flow and is compatible with solutions containing biological samples. Using the high-frequency AC current, electrolysis of the conductive liquid can be avoided. As a result, the bubble formation and electrode degradation do not occur.

2.09.3.3 Ultrasonic Micropumps

Ultrasonic pumping is based on the phenomenon of acoustic streaming. When a flexural wave propagates in a thin membrane, a high-intensity acoustic field is generated in the fluid near the membrane. This acoustic field causes fluid flow in the direction of wave propagation (**Figure 22**) (Bradly *et al.* 1995). Since the acoustic field in the channel decays exponentially with the distance from the vibrating plate, the channel should be shallow to drive fluid flow efficiently. All kinds of liquids can be pumped. The flow rate can be controlled by changing the amplitude of the driving signal of the vibrating device. The flexural wave was generated by using a thin film of piezoelectric actuator (Miyazaki *et al.* 1991, Nguyen *et al.* 2000). **Figure 23** shows an acoustic streaming micropump using interdigitated transducers formed by a sputtered ZnO film and a sputtered aluminum electrode (Black *et al.* 1999, Moroney *et al.* 1991, Nguyen *et al.* 1999). The integration of the thermal flow sensor was also achieved (Nguyen *et al.* 2000).

2.09.3.4 Electrochemical Micropumps

In electrochemical micropumps, the gas bubble generated by the electrochemical reaction was utilized to drive the fluid flow. **Figure 24** shows the schematic diagram of an electrochemical micropump. The bidirectional pumping was achieved by using reversible chemical reactions. The precise regulation of the fluid flow can be realized by controlling the total amount of electric charge for the electrochemical reactions. The oxygen and hydrogen gas bubbles formed by the electrolysis reaction of water were used to drive the liquid (Böhm *et al.* 1999b). If the direction of the current is reversed, the previously produced gas can be reacted back to water and, as a

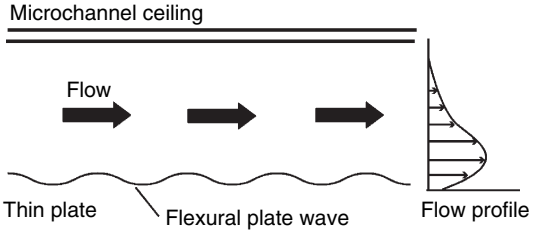


Figure 22 A flow channel with acoustic pumping. (Source: Bradly C E, Bustillo J M, White R M 1995 Flow measurements in a micromachined flow system with integrated acoustic pump. *Proceedings of the IEEE Ultrasonic Symposium*, pp. 505–10.)

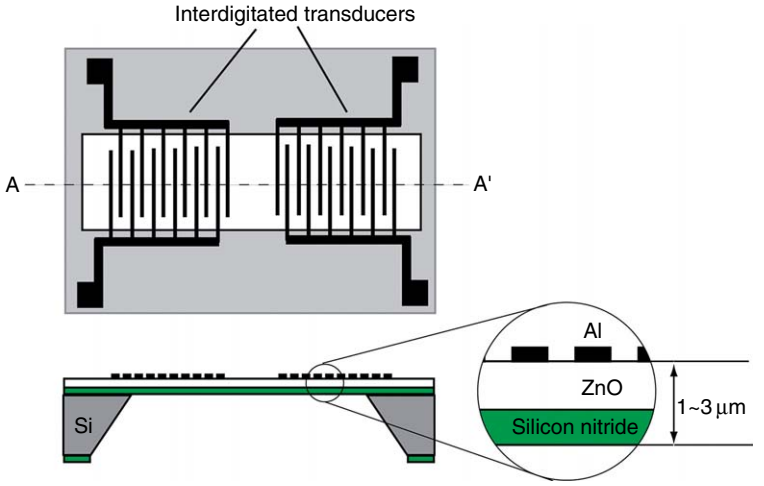


Figure 23 An acoustic streaming pump using interdigitated transducers. (Source: Black J P, White R M 1999 Microfluidic applications of ultrasonic flexural plate waves. *10th Int. Conf. Solid-State Sensors and Actuators*, Sendai, Japan, Vol. 2, pp. 1134–7.)

result, the liquid can be drawn back. The redox reaction of an acidic electrolyte solution was used for pumping (Suzuki and Yoneyama 2003). The generation and shrinkage of the hydrogen bubble was utilized to realize the bidirectional pumping of the fluid. The chemical reaction was controlled using a potentiostat. Since the rates of growth and shrinkage were reproducible, accurate control of pumping was realized.

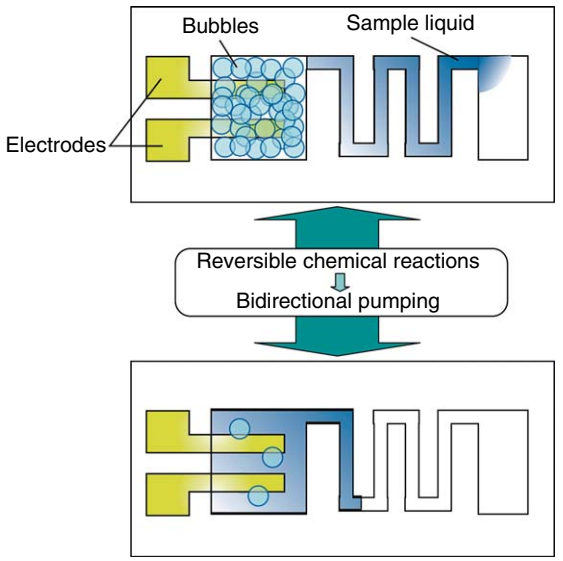


Figure 24 Schematic diagram of the electrochemical micropump. (Source: Böhm S, Olthuis W, Bergveld P 1999b An electrochemically actuated micropump for use in a push-pull microdialysis based *in-vivo* monitoring system. *10th Int. Conf. Solid-State Sensors and Actuators*, Sendai, Japan, Vol. 2, pp. 880–1.)

2.09.3.5 Other Nonmechanical Micropumps

2.09.3.5.1 Electroosmotic micropumps

Electroosmosis is one of the electrokinetic phenomena and is used as a precise liquid-handling technique. The electrical double layer at the interface between a channel wall and an electrolyte plays an important role to produce electroosmotic flow. When a voltage is applied between the two ends of the microchannel, the ions of the double layer are moved by the influence of the electric field. Due to the shear viscosity, the ions force the bulk liquid to move through the channel (**Figure 25**). The microchannel packed with the silica spheres (Takemori *et al.* 2005, Zeng *et al.* 2001) and closely spaced parallel plate microchannels (Chen and Santiago 2002) were utilized to achieve the high-pressure and high-flow-rate electroosmotic micropump.

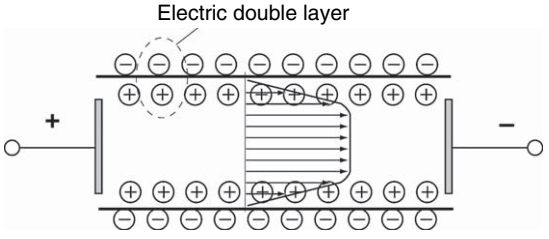


Figure 25 Pumping mechanism of electroosmotic flow. (Source: Zeng S, Chen C H, Mikkelsen J C, Santiago J G 2001 Fabrication and characterization of electroosmotic micropump. *Sens. Actuators B* **79**, 107–14.)

2.09.3.5.2 Chemical gas generating pumps

The chemically produced gases were used as the pressure source for liquid pumping. The generated gases drive the liquid sample from the reservoir to the outlet along the microchannel (**Figure 26**). The amount of chemical reactant affected working time and pumping volume. Because of its simple structure and easy fabrication, this type of micropump is suitable for the disposable lab-on-a chip and biomedical analysis systems. Nitrogen gas was generated by heat decomposition of azobis-isobutyronitrile (Hong *et al.* 2003). Oxygen gas was produced by the decomposition of hydrogen peroxide by catalyst (Choi *et al.* 2004).

2.09.4 Applications of the Micropumps

Recently, some micropumps designed to make their structure suitable for actual applications have been reported. Two types of micropumps using piezoelectric disk actuators for drug delivery applications were developed (Geipel *et al.* 2006, Maillefer *et al.* 1999). Stable and constant flow injection is required for this purpose. A plastic drug delivery system without using any electric power was developed (Su *et al.* 2002). An osmotic actuator consisting of a water semipermeable membrane, actuation membrane, and cavity was employed. A constant delivery rate of 0.2 $\mu l\,h^{-1}$ for 10 h is obtained. A diaphragm micropump with active microvalves driven by piezoelectric actuators was developed for an artificial sphincter system (Doll *et al.* 2005). A nozzle/diffuser micropump with an electromagnetic actuator was reported for cellomics applications (Yamahata *et al.* 2005b). The valveless structure is suitable for biological cell transport. High-pressure operation is required for actual use. For electronic applications, silicon electroosmotic micropump for integrated circuit thermal management was also reported (Laser *et al.* 2003).

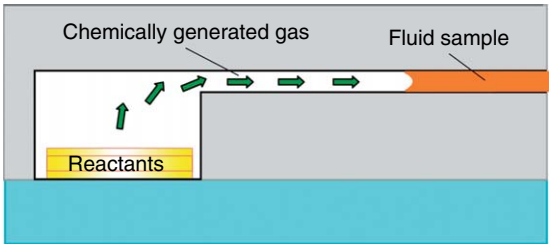


Figure 26 Schematic illustration of the chemically generated gas micropump. (Source: Hong C C, Murugesan S, Kim S, Beaucage G, Choi J W, Ahn C H 2003 A functional on-chip pressure generator using solid chemical propellant for disposable lab-on-a-chip. *IEEE 16th Int. Conf. Micro Electro Mechanical Systems*, Kyoto, Japan, pp. 16–19.)

2.09.5 Summary

Various types of micropumps have been developed over the past 20 years. Self-priming and bubble-tolerant micropumps are realized by optimizing the pump structure. Low-cost plastic micropumps are also developed considering biochemical and medical uses. Mass production of micropumps are already obtained by developing the large-scale batch fabrication technology. High-performance micropumps are realized by sophisticated silicon micromachining. Since the performances of the micropump are strongly dependent on the working principle, we have to choose the structure and the material of the micropump considering its application. To realize optimize design of the pump structure, CFD simulation is now getting a powerful tool. Integration and assembly with other fluidic elements are the most important issues to realize μTAS and LOC.

References

Bart S F, Tavrow L S, Mehran M, Lang J H 1990 Microfabricated electrohyrdrodynamic pumps. *Sens. Actuators* **A21–A23**, 193–7

Black J P, White R M 1999 Microfluidic applications of ultrasonic flexural plate waves.*10th Int. Conf. Solid-State Sensors and Actuators*, Sendai, Japan, Vol. 2, pp. 1134–7

Bodén R, Lehto M, Simu U, Thornell G, Hjort K, Schweitz JÅ 2005 A polymeric paraffin micropump with active valves for high-pressure microfluidics.*13th Int. Conf. Solid-State Sensors, Actuators and Microsystems*, Seoul, Korea, Vol. 1, pp. 201–4

Böhm S, Olthuis W, Bergveld P 1999a A plastic micropump constructed with conventional techniques and materials. *Sens. Actuators A* **77**, 223–8

Böhm S, Olthuis W, Bergveld P 1999b An electrochemically actuated micropump for use in a push-pull microdialysis based *in-vivo* monitoring system.*10th Int. Conf. Solid-State Sensors and Actuators*, Sendai, Japan, Vol. 2, pp. 880–1

Bradly C E, Bustillo J M, White R M 1995 Flow measurements in a micromachined flow system with integrated acoustic pump. *Proceedings of the IEEE Ultrasonic Symposium*, Seattle, USA, pp. 505–10

Büstgens B, Bacher W, Menz W, Schomburg W K 1994 Micropump manufactured by thermoplastic molding. *Micro Electro Mechanical Systems*, Oiso, Japan, pp. 18–21

Chen C H, Santiago J G 2002 A planer electroosmotic micropump. *J. Microelectromech. Syst*, **11(6)**, 672–83

Choi Y H, Son S U, Lee S S 2004 A micropump operating with chemically produced oxygen gas. *Sens. Actuators A* **111**, 8–13

Darabi J, Rada M, Ohadi M, Lawler J 2002 Design, fabrication and testing of an electrohydrodynamic ion-drag micropump. *J. Microelectromech. Syst.* **11(6)**, 684–90

Doll A, Reimers S, Heinrich M, Goldschmidtboeing F, Schrag H-J, Hopt U T, Woias P 2005 A high performance bidirectional micropump for a novel artificial sphincter system.*13th Int. Conf. Solid-State Sensors, Actuators and Microsystems*, Seoul, Korea, Vol. 1, pp. 188–91

Döpper J, Clemens M, Ehrfeld W, Jung S, Kämper K-P, Lehr H 1997 Micro gear pumps for dosing of viscous fluids. *J. Micromech. Microeng.* **7**, 230–2

Feng G-H, Kim E S 2005 Piezoelectrically actuated dome-shaped diaphragm micropump. *J. Microelectromech. Syst.* **14(2)**, 192–9

Fuhr G, Hagedorn R, Muller T, Benecke W, Wagner B 1992 Microfabricated electrohydrodynamic (EHD) pumps for liquids of higher conductivity. *J. Microelectromech. Syst.* **1(3)**, 141–6

Geipel A, Doll A, Goldschmidtböing F, Jantscheff P, Esser N, Massing U, Woias P 2006 Pressure-independent micropump with piezoelectric valves for low flow drug delivery systems. *19th IEEE Int. Conf. Micro Electro Mechanical Systems*, Istanbul, Turkey, pp. 786–9

Gerlach T, Schuenemann M, Wurmus H 1995 A new micropump principle of the reciprocating type using pyramidic micro flow channels as passive valves. *J. Micromech. Microeng.* **5**, 199–201

Gravenson P, Branebjerg J, Jensen O S 1993 Microfluidics – Review. *J Micromech Microeng.* **3**, 169–82

Grosjean C, Tai Y C 1999 A thermopneumatic peristaltic micropump.*10th Int. Conf. Solid-State Sensors and Actuators*, Sendai, Japan, Vol. 2, pp. 1776–9

Grosjean C, Yang X, Tai Y C 1999 A practical thermopneumatic valve. *9th Int. Conf. Solid-State Sensors and Actuators*, Orland, CA, USA, pp. 147–52

Homsy A, Koster S, Eijkel J C, van den Berg A, Lucklum F, Verpoorte E, de Rooij N F 2005 A high current density DC magnetohydrodynamic (MHD) micropump. *Lab Chip* **5**, 466–71

Hong C C, Murugesan S, Kim S, Beaucage G, Choi J W, Ahn C H 2003 A functional on-chip pressure generator using solid chemical propellant for disposable lab-on-a-chip. *IEEE 16th Int. Conf. Micro Electro Mechanical Systems*, Kyoto, Japan, pp. 16–19

Jang J, Lee S S 2000 Theoretical and experimental study of MHD (magnetohydrodynamic) micropump. *Sens. Actuators A* **80**, 84–9

Jeong O C, Park S W, Yang S S, Pak J J 2005 Fabrication of a peristaltic PDMS micropump. *Sens. Actuators A* **123–124**, 453–8

Kämper K-P, Döpper J, Ehrfeld W, Oberbech S 1998 A self-filling low-cost membrane micropump. *11th Annu. Int. Workshop Micro Electro Mechanical Systems*, Heidelberg, Germany, pp. 432–7

Kim J-H, Na K-H, Kang C J, Kim Y-S 2005 A disposable thermopneumatic-actuated micropump stacked with PDMS layers and ITO-coated glass. *Sens. Actuators A* **120**, 365–9

Kim E-G, Oh J-G, Choi B 2006 A study on the development of a continuous peristaltic micropump using magnetic fluids. *Sens. Actuators A* **128**, 43–51

Laser D J, Santiago J G 2004 A review of micropumps. *J. Micromech. Microeng.* **14(6)**, R35–64

Laser D J, Myers A M, Yao S, Bell K F, Goodson K E, Santiago J G, Kenny T W 2003 Silicon electroosmotic micropumps for integrated circuit thermal management.*12th Int. Conf. Solid-State Sensors, Actuators and Microsystems*, Boston, MA, USA, Vol. 1, pp. 151–4

Lemoff A V, Lee A P 2000 An AC magnetohydrodynamic micropump. *Sens. Actuators B* **63**, 178–85

Linnemann R, Woias P, Senfft C-D, Ditterich J A 1998 A self-priming and bubble-tolerant piezoelectric silicon micropump for liquids and gases. *11th Annu. Int. Workshop Micro Electro Mechanical Systems*, Heidelberg, Germany, pp. 532–7

Litterst C, Streule C W, Koltay P, Zengerle R 2005 Simulation toolkit for micro-fluidic pumps using lumped models. *2005 Nanotechnology Conference and Trade Show*, Anaheim, CA, USA, pp. 736–9

Maillefer D, van Lintel H, Rey-Mermet G, Hirschi R 1999 A high-performance silicon micropump for an implantable drug delivery system. *12th IEEE Int. Conf. Micro Electro Mechanical Systems*, Orlando, FL, USA, pp. 541–6

Maillefer D, Gamper S, Frehner B, Balmer P 2001 A high-performance silicon micropump for disposable drug delivery systems. *14th IEEE Int. Conf. Micro Electro Mechanical Systems*, Interlaken, Switzerland, pp. 413–17

Makino E, Mitsuya T, Shibata T 2001 Fabrication of TiNi shape memory micropump. *Sens. Actuators A* **88**, 256–62

Matsumoto S, Klein A, Maeda R 1999 Development of bi-directional valve-less micropump for liquid. *12th IEEE Int. Conf. Micro Electro Mechanical Systems*, Orlando, FL, USA, pp. 141–6

Miyazaki S, Kawai T, Araragi M 1991 A piezo-electric pump driven by a flexural progressive wave. *Micro Electro Mechanical Systems*, Nara, Japan, pp. 283–8

Moroney R M, White R M, Howe R T 1991 Ultrasonically induced microtransport. *Micro Electro Mechanical Systems*, Nara, Japan, pp. 277–82

Nguyen N-T, Truong T-Q 2004 A fully polymeric micropump with piezoelectric actuator. *Sens. Actuators B* **97**, 137–143

Nguyen N T, White R M 1999 Design and optimization of an ultrasonic flexural plate wave micropump using numerical simulation. *Sens. Actuators A* **77**, 229–36

Nguyen N T, Meng A H, Black J, White R M 2000 Integrated flow sensor for in situ measurement and control of acoustic streaming in flexural plate wave micropumps. *Sens. Actuators A* **79**, 115–21

Oh K W, Ahn C H 2006 A review of microvalves. *J. Micromech. Microeng.* **16(5)**, R13–39

Olsson A, Larsson O, Holm J, Lundbladh L, Öhman O, Stemme G 1998 Valve-less diffuser micropumps fabricated using thermoplastic replication. *Sens. Actuators A* **64**, 63–8

Olsson A, Stemme G, Stemme E 2000 Numerical and experimental studies of flat-walled diffuser elements for valve-less micropumps. *Sens. Actuators A* **84**, 165–75

van de Pol F C M., van Lintel H T G., Elwenspoek M, Fluitman J H J 1989 A thermopneumatic micropump based on micro-engineering techniques. *Sens. Actuators* **A21–23**, 198–202

Rapp R, Schomburg W K, Maas D, Schulz J, Stark W 1994 LIGA micropump for gas and liquid. *Sens. Actuators A* **40**, 57–61

Richter A, Sandmaier H 1990 An electrohydrodynamic micropump. *Proceedings of the IEEE-MEMS Workshop*, Napa Valley, USA, pp. 99–104

Richter M, Kruckow J, Weidhaas J, Wackerle M, Drost A, Schaber U, Schwan M, Kühl K 2001 Batch fabrication of silicon micropumps.*11th Int. Conf. Solid-State Sensors and Actuators*, Munich, Germany, Vol. 2, pp. 936–9

Schabmueller C G J, Koch M, Mokhtari M E, Evans A G R 2002 Self-aligning gas/liquid micropump. *J. Micromech. Microeng.* **12**, 420–4

Shoji S 1998 Topics in current chemistry. 194. Microsystem technology in chemistry and life science. In: Manz A and Becker H (eds.) *Fluids for Sensor Systems*. Springer-Verlag, Berlin, pp. 163–88

Shoji S, Esashi M 1994 Micro flow devices. *J. Micromech. Microeng*, **4**, 157–71

Shoji S, Nakagawa S, Esashi M 1989 Micropump and sample-injector for integrated chemical analyzing systems. *Sens. Actuators* **A21–A23**, 189–92

Smits J G 1989 Piezoelectric micropump with three valves working peristaltically. *Sens. Actuators* **A21–A23**, 203–6

Stemme E, Stemme G 1993 A novel piezoelectric valve-less fluid pump. *7th Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, pp. 110–13

Su Y C, Lin L, Pisano A P 2002 A water-powered micro drug delivery system. *Solid-State Sensors, Actuators and Microsystems Workshop*, Hilton Head Island, SC, USA, pp. 69–72

Suzuki H, Yoneyama R 2003 Integrated microfluidic system with electrochemically actuated on-chip pumps and valves. *Sens. Actuators B* **96**, 38–45

Takemori Y, Horiike S, Nishimoto T, Nakanishi H, Yoshida T 2005 High pressure electroosmotic pump packed with uniform silica nanospheres.*13th Int. Conf. Solid-State Sensors, Actuators and Microsystems*, Seoul, Korea, Vol. 2, pp. 1573–6

Teymoori M M, Abbaspour-Sani E 2005 Design and simulation of a novel electrostatic micromachined pump for drug delivery applications. *Sens. Actuators A* **117**, 222–9

Truong T-Q, Nguyen N-T 2004 A polymeric piezoelectric micropump based on lamination technology. *J. Micromech. Microeng.* **14**, 632–8

Tsai H Jr., Lin L 2001 A thermal bubble actuated micro nozzle-diffuser pump. *14th IEEE Int. Conf. Micro Electro Mechanical Systems*, Interlaken, Switzerland, pp. 409–12

Van Lintel H T G, van de Pol F C M, Bouwstra 1988 A piezoelectric micropump based on micromachining of silicon. *Sens. Actuators* **15**, 153–67

Voigt P 2002 Compact modeling of microsystems. Ph.D. dissertation, Technische Universität München, Lehrstuhl für Technische Elektrophysik

van der Wijngaart W, Andersson H, Enoksson P, Noren K, Stemme G 2000 The first self-priming and bi-directional valve-less diffuser micropump for both liquid and gas. *13th Annu Int. Conf. Micro Electro Mechanical Systems*, Miyazaki, Japan, pp. 674–9

Yamahata C, Gijs M A M 2004 Plastic micropumps using ferrofluid and membrane actuation. *17th IEEE Int. Conf. Micro Electro Mechanical Systems*, Maastricht, The Netherlands, pp. 458–61

Yamahata C, Lacharme F, Burri Y, Gijs M A M 2005a A ball valve micropump in glass fabricated by powder blasting. *Sens. Actuators B* **110**, 1–7

Yamahata C, Vandevyver C, Lacharme F, Izewska P, Vogel H, Freitag R, Gijs M A M 2005b Pumping of mammalian cells with a nozzle-diffuser micropump. *Lab Chip* **5**, 1083–8

Yun K-S, Cho I-J, Bu J-U, Kim C-J, Yoon E 2002 A surface-tension driven micropump for low-voltage and low-power operations. *J. Microelectromech. Syst.* **11(5)**, 454–61

Zeng S, Chen C H, Mikkelsen J C, Santiago J G 2001 Fabrication and characterization of electroosmotic micropump. *Sens. Actuators B* **79**, 107–14

Zengerle R, Richter M 1994 Simulation of microfluid systems. *J. Micromech. Microeng.* **4**, 192–204

Zengerle R, Richter A, Sandmaier 1992 A micro membrane pump with electric actuation. *Int. Workshop Micro Electro Mechanical Systems*, Travemunde, Germany, pp. 19–24

Zengerle R, Richter M, Brosinger F, Richter A, Sandmaier H 1993 Performance simulation of microminiaturized membrane pumps. *7th Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, pp. 106–9

Zengerle R, Ulrich J, Kluge S, Richter M, Richter A 1995a Bidirectional silicon micropump. *Sens. Actuators* **A50**, 81–6

Zengerle R, Geiger W, Richter M, Ulrich J, Kluge S, Richter A 1995b Transient measurements on miniaturized diaphragm pumps in microfluidic systems. *Sens. Actuators* **A46–47**, 557–61

Biographies

Shuichi Shoji received his BS, MS, and Ph.D. degrees in Electronic Engineering from Tohoku University in 1979, 1981, and 1984, respectively. He was with the Tohoku University as a research associate and an associate professor from 1984 to 1992. In 1994 he moved to Waseda University as an associate professor and he is currently a professor in the Department of Electrical Engineering and Bioscience, and a Major in Nano-science and Nano-engineering, at Waseda University. His current interests are micro/nanodevices and systems for chemical/bio applications.

Hironobu Sato received his Ph.D. degree in Engineering from Waseda University, Tokyo, Japan, in 2005. From 2002 to 2005, he was a research associate at Waseda University. He is a postdoctoral fellow of the Japan Science and Technology Agency. He is interested in microfabrication and microfluidics.

Prof. Dr. Roland Zengerle is Head of the Laboratory for MEMS Applications at the Department of Microsystems Engineering (IMTEK) at the University of Freiburg, Germany. In addition he is a director at the Institute for Micro- and Information Technology of the Hahn-Schickard-Gesellschaft (HSG-IMIT). HSG-IMIT is a nonprofit organization supporting industries in the development of new products based on MEMS technologies.

The research of Dr Zengerle and his team of 50 engineers is focused on microfluidics and covers topics like miniaturized and autonomous dosage systems, implantable drug delivery systems, nanoliter and picoliter dispensing, lab-on-a-chip systems, thermal sensors, miniaturized fuel cells as well as micro- and nanofluidics simulation. Dr. Zengerle coauthored more than 200 technical publications and 25 patents. He is the European editor of the *Springer Journal of Microfluidics and Nanofluidics*. Dr. Zengerle serves on the International Steering Committee of the IEEE-MEMS Conference as well as on the Technical Program Committees of several other international conferences.

2.10 Mixers

Shizhi Qian[1] **and Jérôme F. L. Duval**[2], [1]University of Nevada Las Vegas, Las Vegas, NV, USA, [2]Centre National de la Recherche Scientifique, Nancy, France

Glossary

AC Alternating Current
CCD Charge-Coupled Device
DC Direct Current
DEP Dielectrophoresis
DNA Deoxyribonucleic Acid
EDL Electrical Double Layer
EKI Electrokinetic Instability
EOF Electroosmotic Flow
ER Electrorheological
LOC Lab-on-a-Chip
MHD Magnetohydrodynamics
PCR Polymerase Chain Reaction
PDMS Polydimethylsiloxane
PZT Piezoelectric Transducer
SAR Split-and-Recombine
SEM Scanning Electron Microscope
SHM Staggered Herringbone Mixer

2.10.1 Introduction

Over the past decade, there has been a growing interest in developing microsystems or lab-on-a-chip (LOC) technologies for biodetection, biotechnology, chemical and biological reactors, or even medical, pharmaceutical, and environmental monitoring (Chow 2002, Dittrich and Manz 2006, Dittrich *et al.* 2006, Gardeniers and Van den Berg 2004, Huikko *et al.* 2003, Jain 2000, Jensen 1999, Langer 2000, Stone and

Kim 2001, Thilmany 2005, Verpoorte 2002). Lab on a chip is a minute chemical processing plant, where common laboratory procedures ranging from filtration and mixing to separation and detection are carried out in the palm of the hand. Resembling electronic circuit boards, these integrated devices contain a network of microconduits, which dilutes the sample, separates it into multiple channels for parallel analysis, mixes it with target-specific antibodies or reagents, propels it from one part of the device to the other, and detects the presence of chemical or biological targets. This whole set of precise and reproducible operations is done automatically in a single platform, which ultimately results in high data quality and reduces the need for trained personnel. This technology therefore has the potential of revolutionalizing various bioanalytical applications, the success of which is subjected to the possibility of handling tiny amounts of samples in a rapid and cheap manner, and analyzing them according to automatized and multiple parallel procedures that lead to minimal cross-contamination. To meet these very demanding criteria, a rapid mixing of the fluids to be manipulated must be achieved within the microsystems, as those used in biochemical analysis, drug delivery, point-of-care testing, crystallization, protein and RNA folding, and sequencing or synthesis of nucleic acids. Biological processes like cell activation, enzyme reactions, and protein folding often involve reactions that require mixing of various reagents and chemicals for initiation. In the case of large molecules, although the characteristic lengths associated with microdevices are small – typically on the order of 100 μm – diffusion alone does not allow sufficiently fast mixing. For example, at room temperature, the diffusion coefficient of myosin in water is about $10^{-11}\,m^2\,s^{-1}$, and the time constant for the diffusion along a length of 100 μm is thus intolerably large, i.e., about 10^3 s. Therefore, mixing several fluids in chambers or channels at the micrometer scale is not as easy as it might seem at first sight. Since the Reynolds numbers of flows in microdevices are usually very small (i.e., $Re \sim O(1)$), the flows are laminar and mixing enhancement cannot be reached by making use of turbulence-like flow patterns. **Figure 1** depicts the multifluid laminar flow scheme that is typically encountered in microfluidic channels (Stone and Kim 2001). Various streams colored with different dyes are brought together through Y-junctions. Parallel, side-by-side, laminar flow pattern is formed with diffusion that facilitates only very slow mixing process. In order to achieve reasonable speed and yield of chemical reactions and bioassays, micromixers must be necessarily integrated into the chips.

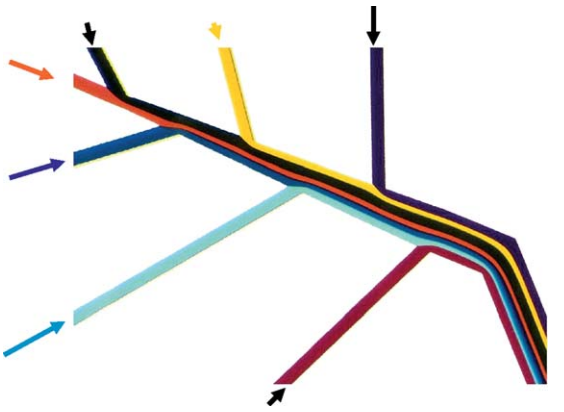


Figure 1 Various streams colored with different dyes are brought together through Y-junctions and flow downstream side by side with poor diffusion-controlled mixing. (Source: Stone H A, Kim S 2001 Microfluidics: Basic issues, applications, and challenges. *AICHE J.* **47**, 1250–4.)

In recent years, numerous micromixers based on various principles have been developed for various LOC applications. Generally speaking, mixing in microdevices is enhanced through chaotic advection that allows passive particles transported by a periodic velocity field to exhibit chaotic trajectories (Aref 1984, 1990, 2002, Ottino 1989). More specifically, the developed microscale mixing schemes can be divided into two categories, according to their so-called active or passive natures. A mixer is said to be active when its functioning relies on the use of a time-variant external energy source such as that provided by applied pressure, the electric field, and/or the magnetic field, to quote only a few of them. This external energy source excludes that which is employed for propelling the fluid itself, regardless of its state of mixing. Unlike active mixers, passive mixers use the same power source (which is in most of the cases time-independent) as that needed to propel the fluid, and typically take the benefit of irregular or asymmetric channel geometry to laminate the flowing fluids either in-plane or out-of-plane so as to promote chaotic advection. Each of these two types of micromixers has different capacity, mixing speed, and operating requirements. The lack of moving components makes passive mixers free of additional friction and wear effects even if their intricate channel topologies are often hard to microfabricate, and generally not switchable. Once a passive mixer is incorporated into a microdevice, it fulfills its function when fluids flow through it. In contrast, active mixers can be controlled externally, which makes them

suitable and attractive for reconfigurable microfluidic systems that can perform several different functions given different states of external controls. Various techniques used in microsystems for mixing enhancement are summarized below, starting first with the passive mixers before dealing with the active ones.

2.10.2 Passive Mixers

Due to the dominating laminar flow pattern in microfluidic systems, mixing in passive micromixers relies mainly on molecular diffusion and/or chaotic advection. Parallel lamination, split-and-recombine (SAR), injection, droplet, and chaotic mixers are among the most commonly used passive micromixers (Nguyen and Wu 2005).

2.10.2.1 Parallel Lamination Mixer

A parallel lamination mixer splits the inlet fluid streams into several substreams and subsequently joins those into one single stream. This process can be achieved by means of T-mixers, Y-mixers, and hydrodynamic focusing mixers, which essentially differ by their geometry.

2.10.2.1.1 T-mixer

The T-shaped micromixer was one of the first introduced micromixers for mixing in a microfluidic channel. Two fluid streams are directed into two different inlet arms of a T-channel, which then flow into a common long microchannel, as depicted in Figure 2. Although the flow in the mixing microchannel is in the laminar flow range, there may be developments of vortices depending on the magnitude of the Reynolds number in the mixing channel – denoted as $Re = ud/v$ where u is the average velocity in the mixing channel, d is the hydraulic diameter, and v is the kinematic viscosity of the fluid.

Figure 3 depicts the velocity profiles and concentration fields in the middle part of a T-mixer (in relation to the channel depth). The geometry of the micromixer is 200 μm (width of the mixing channel) by 100 μm (width of the inlet arms) by 100 μm (depth of the mixer). For low Re values, the velocity profile is parabolic, the two fluid streams flow side by side and there is only a small amount of mass transfer perpendicular to the main flow (Figure 3(a)). In this flow region, the mixing of the two fluid solutions relies only on molecular diffusion from one side of the channel to the other, which explains why long microchannels with small width are required. For higher Re values, a double vortex pair occurs, which can stir fluid and improve mixing. When Re exceeds a certain threshold value, an unsteady engulfment flow characterized by the development of a fine lamella takes place in the mixing channel (Figure 3(b)). In the engulfment flow region, the intertwinement of both input streams leads to an enlarged interfacial surface area resulting in a rapid mixing.

Figure 4 depicts the mixing performance of a T-mixer in which HCl solution is introduced into one of the two inlet arms, and NaOH solution with disodium fluorescein dye is injected into the other. The micromixer size is 600 μm (width of the mixing channel) by 300 μm (width of the inlet arms) by 270 μm (depth of the mixer). The observed fluorescence intensity is proportional to the product of the concentration of the HCl–NaOH reactive system. Obviously, when $Re = 150$, the product of the reactive system is very limited due to poor mixing. When $Re = 250$, rapid mixing occurs and the product is formed as the fluid streams flow downstream.

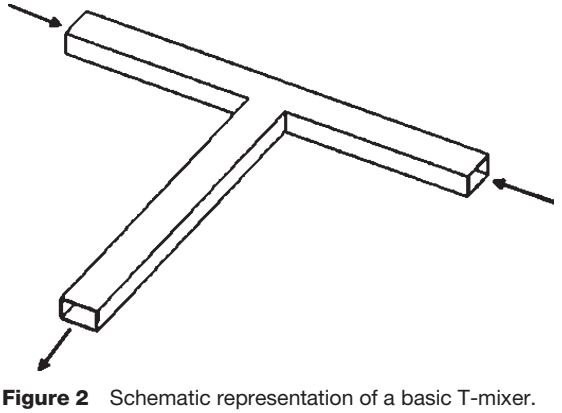


Figure 2 Schematic representation of a basic T-mixer.

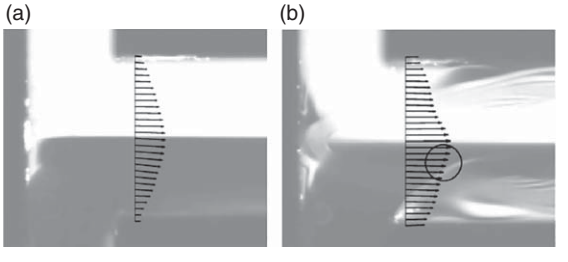


Figure 3 Velocity profiles and concentration fields in the middle part of a T-mixer at $Re = 120$ (a) and $Re = 186$ (b). (Source: Hoffmann M, Schluter M, Rabiger N 2006 Experimental investigation of liquid–liquid mixing in T-shaped micro-mixers using μ-LIF and μ-PIV. *Chem. Eng. Sci.* **61**, 2968–76.)

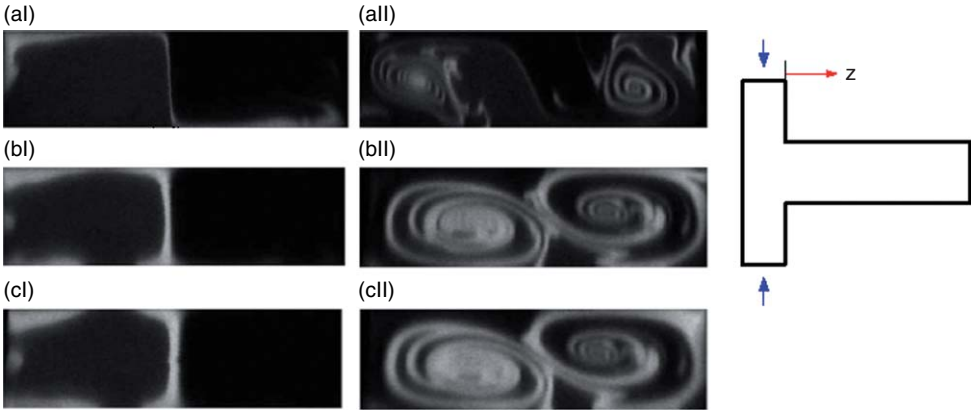


Figure 4 Cross-section areas of the microsystem containing HCl and NaOH with the dye disodium fluorescein at different *Re* (*left column*: *Re* = 150, *right column*: *Re* = 250) and various positions along the mixing channel. (a) *z* = 0.5 mm; (b) *z* = 5 mm; and (c) *z* = 10 mm. (Source: Hoffmann M, Schluter M, Rabiger N 2006 Experimental investigation of liquid–liquid mixing in T-shaped micro-mixers using μ-LIF and μ-PIV. *Chem. Eng. Sci.* **61**, 2968–76.)

Due to its simplicity, the basic T-mixer has been integrated in microsystems for various chemical and biological applications. For the sake of illustration, we mention its use for performing enzyme assays (Hadd *et al.* 1997), for measuring target analyte concentrations (Kamholz *et al.* 1999), for detecting ammonia in aqueous solution (Veenstra 1999), for studying rapid chemical reactions (Hinsmann *et al.* 2001), for crystallizing benzoic acid by mixing sodium benzoate and hydrochloric acid (Stahl *et al.* 2004), for nanoparticle precipitation (Schwarzer and Peukert 2004), and also its use in sol–gel reactors (Rivallin *et al.* 2005).

To reduce the mixing channel length of the basic T-shaped micromixer, several variations of the design/geometry have been proposed: square-wave, inclined, oblique, and wavelike microchannels (Jen *et al.* 2003) are some illustrations reported in **Figure 5**. The flow sways around these varying structures within the twisted microchannels, and secondary flows are induced, which significantly improve the mixing propensity of the system. Numerical calculations indicate that the inclined channel yields the most efficient mixing. Other variants, like those proposed by Jen *et al.* (2003), include substructures/subsections that may lead to optimum length of the mixing channel and consequently to significant enhancement of the mixing process.

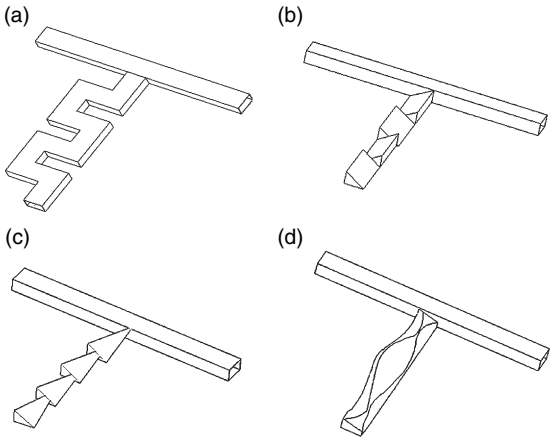


Figure 5 Schematics of (a) square-wave T-mixer; (b) inclined T-mixer; (c) oblique T-mixer; and (d) wave-like T-mixer.

2.10.2.1.2 Y-mixer

A Y-shaped micromixer is generally viewed as a subset of the T-shaped micromixer described in Section 2.10.2.1.1. The basic Y-mixer works on the same principles that govern the mixing quality of the basic T-mixer. **Figure 6** depicts the mixing performance of a typical Y-shaped micromixer. Two fluid streams, as marked by red and blue colors, are brought into contact at *Re* = 30 through the Y-junction. The two fluids begin to diffuse, which results in the formation of a mixed fluid layer. Diffusive mixing is the dominant mechanism in the straight mixing channel, and hence mixing remains very poor.

Figure 6 Mixing performance in a basic Y-mixer where two fluid streams (red and blue colors) meet at a Y-junction and enter a straight mixing microchannel at *Re* = 30. Scale bar, 0.2 mm.

As for the T-mixer, different geometries have been developed to improve the mixing performance of the basic Y-mixer. Figure 7 schematically depicts some of them. A method to reduce the mixing length is based on the introduction of a throttle immediately after the Y-junction, as represented in Figure 7(b) (Gobby *et al.* 2001). The mixing length decreases as the throttle size decreases, but at the expense of a higher pressure drop. Despite the reduction in the length of the mixing channel as generated by the throttle, the mixing mechanism therein still relies on molecular diffusion, which prevents fast and efficient mixing.

Another method sometimes followed for improving the mixing quality of a given Y-shaped micromixer is the replacement of the straight mixing microchannel by a square-wave microchannel, as sketched in Figure 7(c). The presence of the bends induces secondary flows, which can stir fluids, enhance mixing, and under certain conditions can even induce chaotic advection (Yi and Bau 2003). Unfortunately, for sufficiently low Reynolds numbers, the bend-induced secondary flows decay well before they have an opportunity to significantly stir the fluid, so that diffusive mixing remains the dominant process. Figure 8 depicts the mixing performance in a square-wave microchannel with a series of C-turns at $Re = 30$. Two fluid streams are introduced in the legs A and B of the Y part of the microsystem, and remain well separated after flowing through the bends.

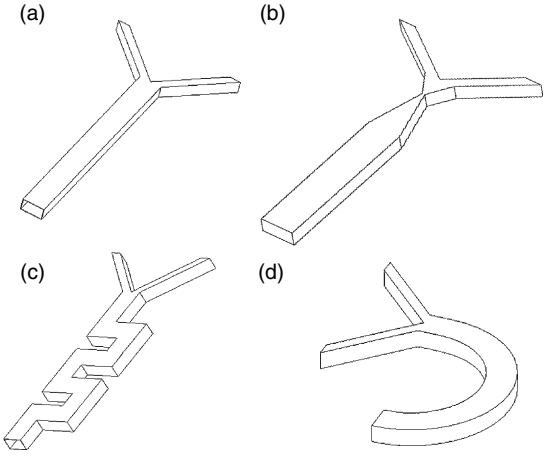


Figure 7 Schematics of (a) basic Y-mixer; (b) throttle Y-mixer; (c) square-wave Y-mixer; and (d) curved Y-mixer.

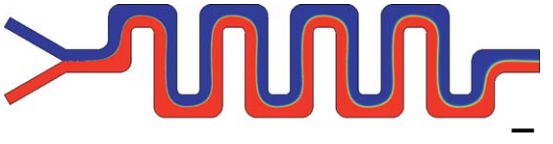


Figure 8 Mixing performance in a square-wave microchannel with a series of C-turns. Two fluid streams (red and blue colors) meet at a Y-junction and enter a square-wave microchannel at $Re = 30$. Scale bar, 0.2 mm.

At any flow rate, the flow through a curved microfluidic channel is obviously different from that through a straight channel. By simply introducing curvature in the mixing channel, vortices, also called Dean vortices, arise in the vertical plane of the curved channel as a result of an interplay between inertial, centrifugal, and viscous effects. Under appropriate and judiciously chosen conditions, these effects establish a radial pressure gradient whose magnitude, if sufficiently large, may lead to the generation of a transverse flow field. This transversal secondary flow associated with the Dean effects can be characterized in terms of a dimensionless number, the Dean number denoted as κ and defined as follows:

$$\kappa = \left(\frac{d}{R}\right)^{0.5} Re$$

where R is the flow path radius of curvature and d is the hydraulic diameter of the channel, i.e., $d = 4A/P$, where A is the cross-sectional area and P is the wetted perimeter. This dimensionless Dean number expresses the relative magnitudes of inertial and centrifugal forces to viscous forces. Figure 9 illustrates the Dean flow phenomena in curved microchannels. Two fluid streams are brought together through the Y-junction and enter a curved microchannel segment. At low κ ($\kappa \sim 1$), the secondary flow is not strong enough to perturb the axial laminar flow profile, and the two fluid streams remain well separated at the exit of the curved channel (Figure 9, *upper*). As κ increases ($\kappa \sim 10$), the induced transverse flow component acts to transport fluid from the inner wall of the channel radially toward the outer wall and the interfacial area between the two fluids significantly increases (Sudarsan and Ugaz 2006).

2.10.2.1.3 Hydrodynamic focusing mixer

Another way to reduce the mixing length in parallel lamination micromixers is by creating a thin jet containing the macromolecules of interest by hydrodynamic focusing (Hertzog *et al.* 2004, Knight *et al.* 1998, Park *et al.* 2006). Figure 10 depicts the three-inlet hydrodynamic focusing micromixer developed by Knight *et al.* (1998). The edges of the microchannels are marked with dashed lines, and the variables W_c and W_f represent the width of the channel and the width of the

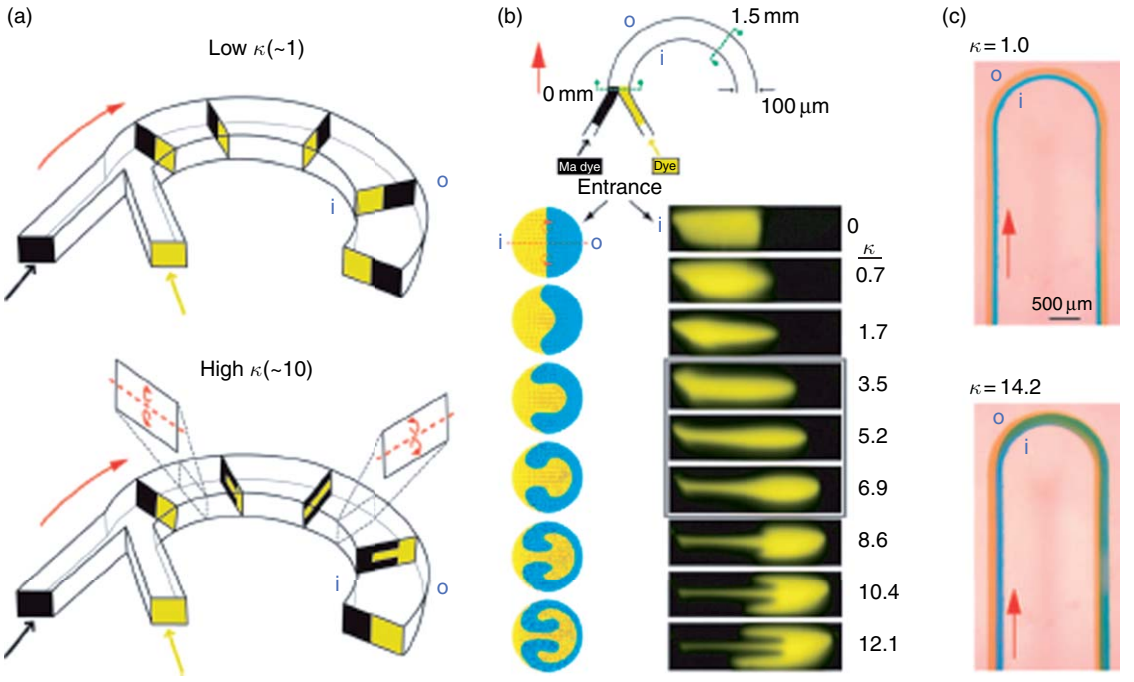


Figure 9 Dean flow phenomena in curved microchannels. (a) Idealized Dean flow-mediated rotation sequence (i and o denote the inner and outer channel walls); (b, *upper*) schematic of the curved microchannel geometry investigated (100 μm wide; 29 μm high; and 630 μm radius of curvature). The transverse flow field is examined at the entrance to the curved segment and at a location 1.5 mm downstream. Analytically computed velocity and concentration profiles are shown (*left lower*) beside confocal cross-sectional images of the transverse flow in the microchannel (*Right lower*) at flow rates in the range of $2.6 < Re < 45.1$ ($0.7 < \kappa < 12.1$). The boxed area represents conditions under which the transverse flow induces ≈90° rotation in the upper and lower halves of the channel. (c) Top-view images of aqueous streams labeled with blue and yellow dyes in a curved microchannel segment (200 μm wide; 29 μm high; and 630 μm radius of curvature). At $\kappa \sim 1.0$ (*upper*) the streams flow parallel to each other along the entire length, whereas at $\kappa \sim 14.2$ (*lower*) the blue stream is transported from the inner to the outer wall. (Source: Sudarsan A P, Ugaz V M 2006 Multivortex micromixing. *Proc. Natl. Acad. Sci. U S A* **103**, 7228–33.)

focused inlet stream, respectively. A solution containing given macromolecules is confined into a thin jet by a second solution without macromolecules flowing from two side channels perpendicular to the long axis of the mixing system. Low molecular weight reactants in the second solution rapidly diffuse into the thin jet with the macromolecules. The width of the focused stream does not depend on the magnitude of the applied pressure, but rather on the ratio, α, of the side pressure P_s to the inlet pressure P_i: $\alpha = P_s/P_i$. Focus occurs in a range of α that depends on the details of the channel geometry. Once the stream is focused, the width remains constant unless broadened by diffusion into the side flow or by increase in channel width. One particular problem associated with this type of mixer is the occurrence of premature mixing prior to the formation of the focused jet.

In order to reduce the premixing, a five-inlet hydrodynamic focusing mixer has recently been developed by Park *et al.* (2006) and is shown in **Figure 11**. Macromolecules in solution A are injected into the center of the channel, solution B flows into two side channels, and solution A without macromolecules flows in diagonal channels. The angle between the center and the diagonal channels is 45°. The sheath flow from the diagonal channels serves as a barrier between solutions flowing from the center and the two side channels during the focusing process. With minuscule amounts of sheath flow, the mixing uniformity, while maintaining rapid diffusive mixing, has improved by more than an order of magnitude.

2.10.2.2 Split-and-Recombine Mixer

Since the diffusion timescale is proportional to the square of the diffusion distance, splitting the streams to be mixed into multiple smaller streams and later rearranging them in alternating thin laminae will significantly decrease the mixing time (Bessoth *et al.* 1999, Erbacher *et al.* 1999, Jensen 1998, Sudarsan and

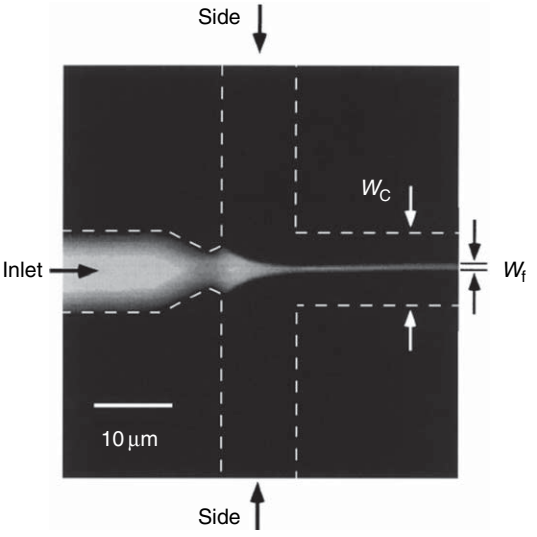


Figure 10 Three-inlet hydrodynamic focusing mixer. The edges of the microchannels are outlined with dashed lines. (Source: Knight J B, Vishwanath A, Brody J P, Austin R H 1998 Hydrodynamic focusing on a silicon chip: Mixing nanoliters in microseconds. *Phys. Rev. Lett.* **80**, 3863–6.)

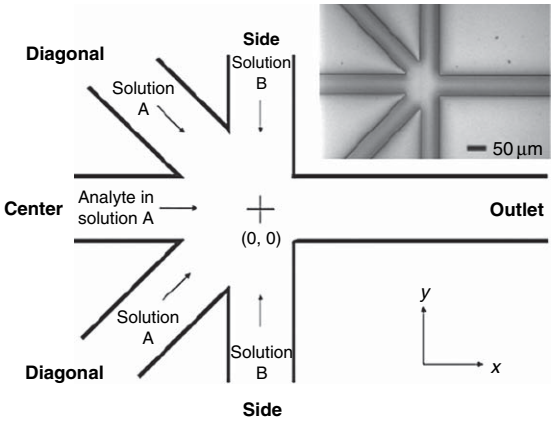


Figure 11 Five-inlet hydrodynamic focusing mixer. (Source: Park H Y, Qiu X, Rhoades E, Korlach J, Kwok L W, Zipfel W R, Webb W W, Pollack L 2006 Achieving uniform mixing in a microfluidic device: Hydrodynamic focusing prior to mixing. *Anal. Chem.* **78**, 4465–73.)

Ugaz 2006, Yamaguchi *et al.* 2004). Based on the channel configurations, SAR micromixers are called either 3D SAR or planar SAR (P-SAR) mixers. Let us examine the peculiarities of these.

2.10.2.2.1 3D SAR mixer

Figure 12(a) schematically depicts the SAR micromixer developed by Manz and coworkers (Bessoth *et al.* 1999, Erbacher *et al.* 1999). The microchip is made from a glass/silicon/glass sandwich. The silicon wafer is etched from both sides and contains a number of wafer-through holes whereas the Pyrex glass serves as a cover and bottom plate and further contains holes for inlet and outlet. On one side of the silicon wafer (layer 1), the inlet channel of liquid A is split into 16 microchannels with a width 1/16 of the total inlet channel width. This is achieved by repeated splitting of the channels in such a way that an array of symmetrical elements results. On the backside of the silicon wafer (layer 2), the channel for liquid B is split in the same way as that for the liquid A. In order to bring liquids A and B together, liquid B is introduced into layer 1 via a number of wafer-through nozzles. When liquid B comes out of a nozzle, it is allowed to develop into the full vertical height of the channel before it enters a channel with liquid A. Subsequently two neighboring channels are combined into one and this is repeated until all partial flows are united into one broad outlet channel. Figure 12(b) depicts the experimental result using fluorescein and rhodamine B at a total flow rate of 50 $\mu L\ min^{-1}$, and mixing in the millisecond regime has been successfully demonstrated.

Similar SAR micromixers using polymethylmethacrylate (PMMA) plates instead of silicon wafer were developed by Yamaguchi *et al.* (2004). Each of the two fluids to be mixed is divided into 16 transverse streams. Subsequently, two different adjacent flow streams are converged five times at a merge angle of 60°. The results demonstrated that about 90% mixing was achieved within 1 s after the streams had been merged.

The above designs are based on the splitting of the fluids to be mixed before they come into contact. A mandatory prerequisite to do so is the fabrication of successive small channels, which may lead to undesired plugging if particulate solutions are used, to higher pressure drops along the channel length, and to substantial dead volumes. To avoid the need for small channels, Munson and Yager (2004) designed a compact SAR mixer by splitting and then combining each fluid along a plane perpendicular to the splitting plane, which generates striations in the fluids to be mixed. By keeping the channel dimensions constant and increasing the number of striations, the characteristic dimension above which diffusion must occur is reduced. Figure 13 depicts the channel geometry of this SAR mixer. The mixer operates by splitting the fluid on each layer in the $y-z$ plane and merging it with the fluid from the other layer in the $x-z$ plane. A 1.8- and 3.0-fold enhancement in the rate of mixing as compared to that in a straight channel was obtained after the first and second mixing units, respectively.

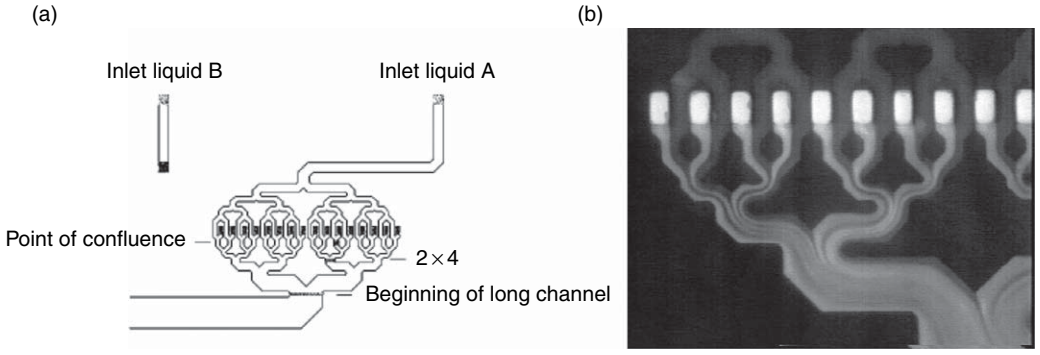


Figure 12 (a) A schematic representation of the split-and-recombine (SAR) mixer developed by Manz and coworkers. (b) The flow visualization using fluorescein and rhodamine B at a total flow rate of 50 μl min^{-1}. (Source: Bessoth F G, deMello A J, Manz A 1999 Microstructure for efficient continuous flow mixing. *Anal. Commun.* **36**, 213–15.)

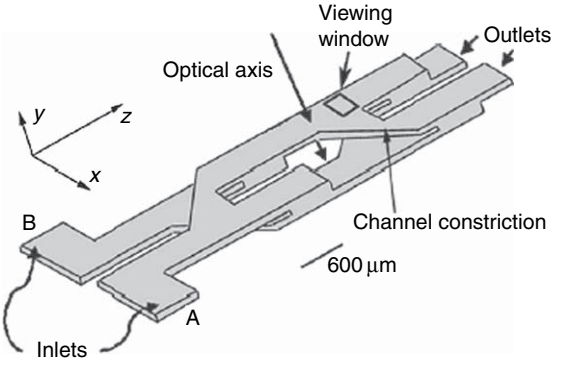


Figure 13 Channel geometry of the split-and-recombine (SAR) mixer developed by Munson and Yager (2004). The mixer operates by splitting the fluid on each layer in the $y-z$ plane and merging it with the fluid from the other layer in the $x-z$ plane. (Source: Munson M S, Yager P 2004 Simple quantitative optical method for monitoring the extent of mixing applied to a novel microfluidic mixer. *Anal. Chim. Acta* **507**, 63–71.)

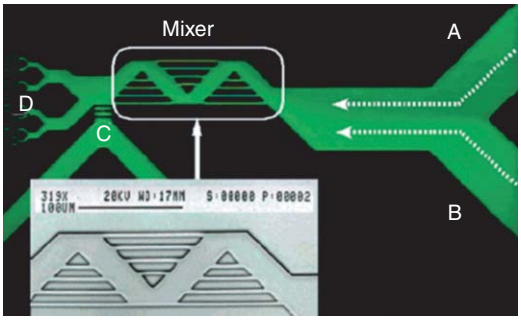


Figure 14 Photomicrograph of the planar split-and-recombine (P-SAR) mixer. The small image in the lower left corner is the scanning electron microscope (SEM) of the mixer. This mixer is about 100 μm × 200 μm wide and 10 μm deep. (Source: He B, Burke B J, Zhang X, Zhang R, Regnier F E 2001 A picoliter-volume mixer for microfluidic analytical systems. *Anal. Chem.* **73**, 1942–7.)

2.10.2.2.2 P-SAR mixer

In contrast to the 3D SAR mixers, all the axes of the channels in P-SAR mixers are located on the same plane, and the design is therefore simpler and does not require the construction of complex 3D geometries.

Figure 14 depicts the picoliter-volume P-SAR mixer developed by He *et al.* (2001). The major features of the design are the following:

(1) The mixing channels have two different channel sizes, one is 5 μm wide and the other is 27 μm wide.
(2) The smaller 5-μm-wide channels are parallel to the flow axis and vary in length; and
(3) The larger 27-μm-wide channels run back and forth through the parallel channel network at an angle of 45°.

The fluid motion is driven by electro-osmotic flow (EOF) instead of pressure-driven flow. It was observed that little mixing of the confluent solvent streams occurred in the upstream of the mixer where mixing would be achieved almost exclusively by diffusion. In contrast, after passage through the mixer, complete mixing was accomplished over a length of about 200 μm as determined by confocal microscopy and charge-coupled device (CCD) detection.

Another simple design has been proposed by Sudarsan and Ugaz (2006) and is depicted in **Figure 15**. Mixing is achieved by the secondary flow induced by Dean vortices as described in Section 2.10.1.1.2. In the P-SAR design, two fluid streams of different species are brought together through the Y-junction. The fluids enter the curved microchannel (marked i in the figure) and experience a transverse flow generated by the counterrotating vortices above and below the channel midplane that induce a

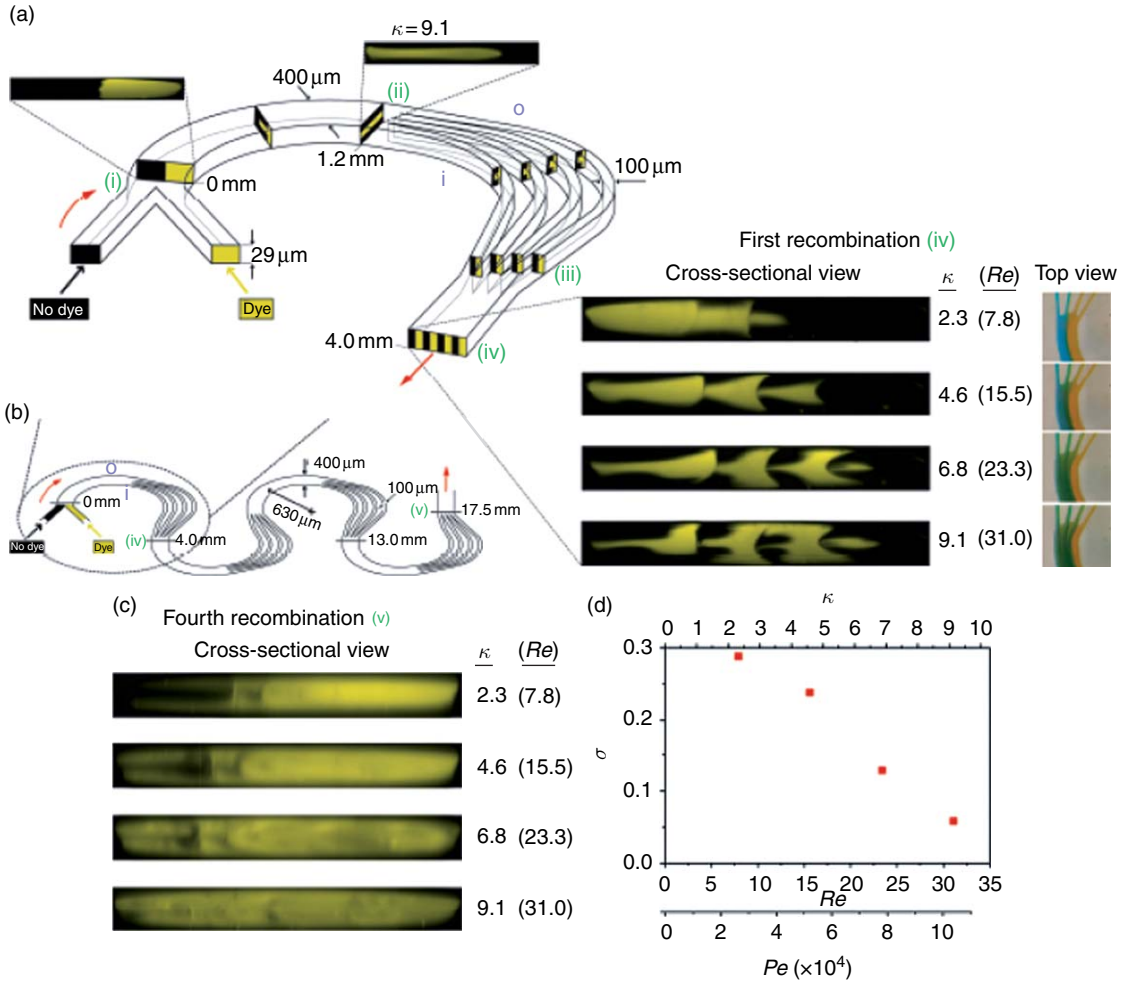


Figure 15 Planar split-and-recombine (P-SAR) micromixer incorporating four split streams. (a) Planar 2D microchannel geometry capable of generating alternating lamellae of individual fluid species in a split-and-recombine (SAR) arrangement (400 μm wide; 29 μm high; 630 μm radius of curvature; *Re* and κ are computed based on the 400-μm-wide segment; i and o denote the inner and outer channel walls, respectively). Flow schematics are shown inside the channel; corresponding confocal images are shown outside. (b) Schematic of a microchannel incorporating a series of successive P-SAR mixing elements. (c) Confocal cross-sectional images taken after the fourth recombination (position (v) in b). (d) Plot of σ (standard deviation of the intensity distribution over each image) evaluated from the confocal image sequence in C as a function of the Dean, Reynolds, and Péclet numbers (*Pe* was computed using $D = 3 \times 10^{-6}\,\mathrm{cm^2\,s^{-1}}$ for Rhodamine 6G). (Source: Sudarsan A P, Ugaz V M 2006 Multivortex micromixing. *Proc. Natl. Acad. Sci. U S A* **103**, 7228–33.)

corresponding pair of 90° rotations in the fluid (ii). At this point (1.2 mm downstream from the entrance i), the flow is split into four parallel streams that develop along curved trajectories inducing a second pair of 90° fluid rotations in each stream (between parts ii and iii of the mixer). Alternating lamellae of the two species are generated when the streams are rejoined 4 mm downstream from the entrance (iv). By employing a channel that implements a series of four successive P-SAR elements (**Figure 15(b)**), a level of 90% mixing is achieved at 17.5 mm downstream position for a Dean number $\kappa = 9.1$ (**Figure 15(c)** and **15(d)**).

2.10.2.3 Injection Mixer

Figure 16 schematically depicts the injection micromixer developed by Miyake *et al.* (1993), which has 400 micronozzles and each of them is 15 μm × 15 μm large at the bottom of a wide shallow channel on a silicon substrate. The sample fluid is introduced into the mixing chamber and the reagent is injected through the micronozzles into the sample at high speed. Because the micronozzles convert the reagent into microplumes, the effective contact surfaces with the sample are significantly increased and, as a result,

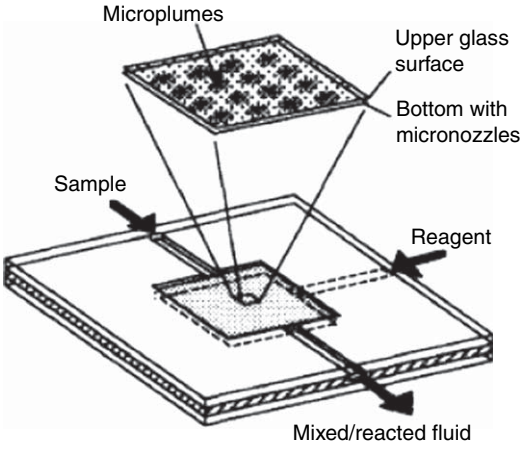


Figure 16 Schematic of an injection micromixer. (Source: Miyake R, Lammerinkz T S J, Elwenspoek M, Fluitman J H J 1993 Micro mixer with fast diffusion. *Proc. MEMS'93, 6th IEEE Int. Workshop Micro Electromechanical System*, San Diego, CA, USA, pp. 248–53.)

the mixing is greatly improved. The experimental results demonstrate that a homogeneous state of mixing is achieved within a few seconds.

The limitation of the above design is that the sample itself is not converted into microplumes. To solve for this limitation, Yang *et al.* (2004) proposed a new design shown in **Figure 17**. The mixer consists of two arrays of micronozzles that are located in parallel directions, and the nozzles are further parallel with respect to the substrate plane. The sample and the reagent are delivered to chambers A and B, respectively. They are then driven into the mixing chamber via the two arrays of micronozzles, which convert both the sample and the reagent into plumes of streams. The experimental results show that very fast mixing with a low pressure drop can be achieved.

2.10.2.4 Droplet Mixer

Another solution to reduce the mixing length is by mixing the liquids within a droplet moving through winding microchannels (Bringer *et al.* 2004, Günther *et al.* 2005, Liau *et al.* 2005, Song *et al.* 2003, Tice *et al.* 2003).

Figure 18 depicts the droplet micromixer developed by Song *et al.* (2003). In the micromixer, three separate aqueous streams converge into a single microchannel where they pinch off into droplets (called plugs) suspended in a perfluorinated oil containing a surfactant. As the plugs move through a serpentine microfluidic channel, the fluid inside the plug is folded, stretched, reoriented, and chaotic advection is induced. Under favorable conditions, submillisecond mixing in the plugs has been demonstrated. The critical advantage of this mixer is that the reactants are confined in each plug as the plugs steadily move down the channel, which prevents axial dispersion. Thus, the progress of the reactions occurring within the plugs is suitable for kinetic analysis, and the mixer can be used to perform millisecond kinetic measurements.

Despite its quality-controlled mixing, the above mixer is ineffective for mixing highly concentrated solutions (CS) of crowding agents such as bovine serum albumin and hemoglobin that are viscous and sticky, and thus extremely difficult to mix rapidly (Liau *et al.* 2005). To do so, the mixer was modified by introducing protrusions along the microfluidic channel walls as shown in **Figure 19**. The protrusions or bumps along the outer wall of the serpentine channel

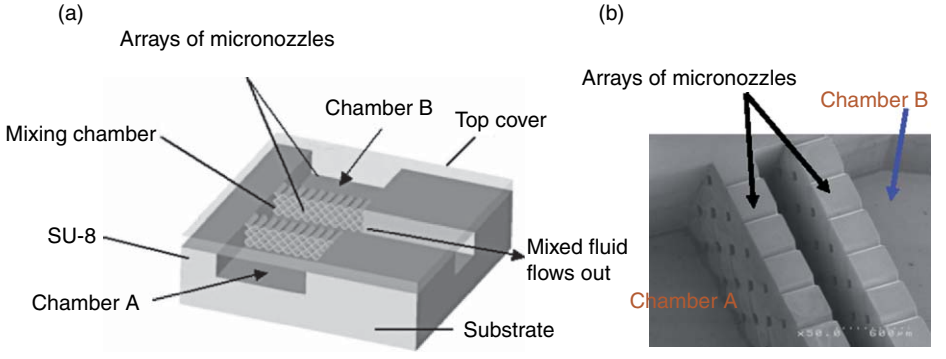


Figure 17 (a) Schematic design of the injection mixer with two arrays of micronozzles located in parallel directions. (b) Scanning electron microscope (SEM) image of the mixing chambers and nozzles. (Source: Yang R, Williams J D, Wang W 2004 A rapid micro-mixer/reactor based on arrays of spatially impinging micro-jets. *J. Micromech. Microeng.* **14**, 1345–51.)

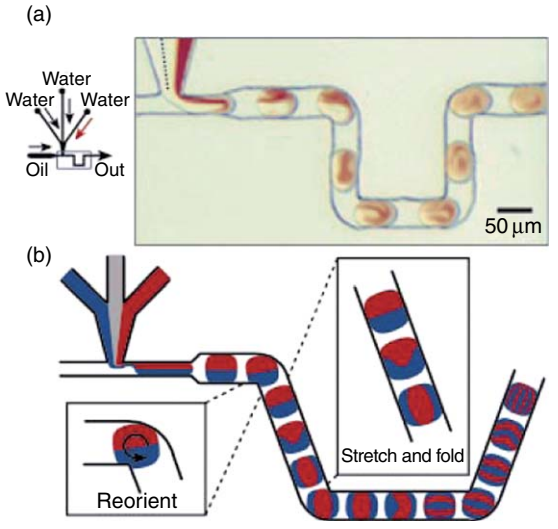


Figure 18 Mixing in plugs that move through winding microfluidic channels. (a) *Left*, a scheme of the microfluidic network; *right*, microphotograph of plugs. (b) Schematic of the droplet mixer. Three separate aqueous streams converge, then intersect with an oil stream, and pinch off into droplets (plugs) suspended in the oil carrier fluid. (Source: Song H, Bringer M R, Tice J D, Gerdts C J, Ismagilov R F 2003 Experimental test of scaling of mixing by chaotic advection in droplets moving through microfluidic channels. *Appl. Phys. Lett.* **83**, 4664–6.)

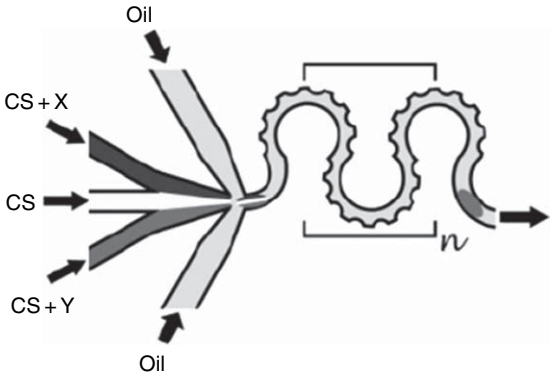


Figure 19 Schematic of a bumpy serpentine droplet mixer. (Source: Liau A, Karnik R, Majumdar A, Cate J H D 2005 Mixing crowded biological solutions in milliseconds. *Anal. Chem.* **77**, 7618–25.)

induce oscillatory flows within the plugs that fold and refold the flow very rapidly. This, in turn, induces rapid mixing of highly concentrated protein solutions, for example. Two streams of CS containing reactants X and Y separated by a third stream of concentrated solution intersect with two oil streams to form droplets suspended in oil (plugs). The plugs then move through n cycles of the bumpy serpentine microfluidic channel until the plug contents are fully mixed. The obtained experimental results show that chaotic mixing of CS can be achieved in the millisecond range. The developed mixer can be used for kinetic studies of biological reactions with millisecond time resolution under conditions of macromolecular crowding like those encountered for cells.

Besides the experimental work, theoretical investigations of chaotic mixing in a liquid droplet moving through a serpentine microfluidic channel have been performed (Muradoglu and Stone 2005). Their theoretical results indicate that the mixing quality depends on the droplet size, the capillary number, the viscosity ratio of the droplet phase to the ambient fluid. The best mixing is obtained when the droplet size is comparable with the channel width. The smaller the viscosity ratio, the better is the quality of the mixing. When the viscosity ratio is below a certain threshold value, the effects of the viscosity ratio on the mixing quality decrease rapidly. The Reynolds number has no significant impact on the mixing even if the quality of mixing somewhat increases with decreasing Reynolds numbers.

2.10.2.5 Chaotic Mixer

Besides diffusion, chaotic advection provides another important mechanism to enhance mass transfer in flows with low Reynolds numbers. The idea underlying chaotic advection is the observation that certain regular velocity fields can produce fluidic pathlines that uniformly fill the volume in an ergodic way. In such velocity fields, fluid elements that are originally close to one another, trace paths that diverge rapidly (exponentially fast in the ideal case), so that the material is dispersed throughout the volume very efficiently. It is important to note that chaotic advection cannot occur in steady 2D flows, but only in 2D time-dependent and 3D flows. Typically, passive chaotic mixers consist of complex 3D twisted conduits fabricated in various substrates such as silicon (Liu *et al.* 2000), polydimethylsiloxane (PDMS) (Stremler *et al.* 2000), ceramic tape (Yi and Bau 2003), or glass (Liu *et al.* 2000). There are several designs of chaotic passive mixers that cause transverse components of flow velocity to stretch, fold the material lines, and exponentially elongate the interface between two fluids.

In one configuration, all the conduits, such as the zigzag conduit and square-wave conduit, lay in the same plane, as represented in **Figure 20**. In these arrangements, the symmetry of the flow field is preserved, and chaotic advection occurs only at high Reynolds numbers (i.e., $Re > 80$). When the Reynolds number is below a certain threshold value, the mixing process remains governed by molecular diffusion (Mengeaud *et al.* 2002, Therriault *et al.* 2003).

The second arrangement consists of twisted pairs of bends with each pair forming a C-shape (**Figure 20(c)**) or a L-shape (**Figure 20(d)**). Previous results have shown that these mixers work well at moderate Reynolds numbers (Yi and Bau 2003). Unfortunately, in these mixers, mixing is not efficient at low Reynolds numbers because the bend-induced vortices decay well before they may significantly stir the fluid. **Figure 21** depicts the mixing result in a 3D C-shaped serpentine passive mixer at $Re = 0.2$. Nearly no mixing is observed close to the mixer outlet even after 15 bends, which indicates that no chaotic advection is induced. Combining the L-shaped serpentine mixer with the SAR technique described in Section 2.10.2.2, an out-of-plane L-shaped micromixer (**Figure 22**) has been developed and shown to work satisfactorily at low Reynolds numbers (Chen and Meiners 2004). In contrast to the basic L-shaped conduit, as illustrated in **Figure 20(d)**, the fluid streams are split out-of-plane, rotated in opposite directions, and recombined in the out-of-plane L-shaped mixer.

For generating chaotic advection at very low Reynolds numbers (i.e., $Re < 1$), Xia *et al.* (2005, 2006) proposed very complicated X-shaped crossing channels (**Figure 23**). The slanted channels are perpendicular to each other and exhibit an angle of 45° with the principal axis of the mixer. The channels are arranged in a periodic manner. In model a (**Figure 23(a)**), the two-layer channels first go across each other at a cross section A. Then at section B, the base-layer channel drives the fluid to the top layer through rotation and the top-layer channel turns the fluid by 90° in the same plane. Subsequently, the two channels meet symmetrically at section C and guide the fluid to the base layer. The structure at section D is similar to that at section B, and the only difference is that the two side channels are interconverted. Section C to D is the first half cycle. The second half cycle has structures similar to the first one except that the slanted channels turn in the opposite direction in an attempt to reduce the isolated fluid

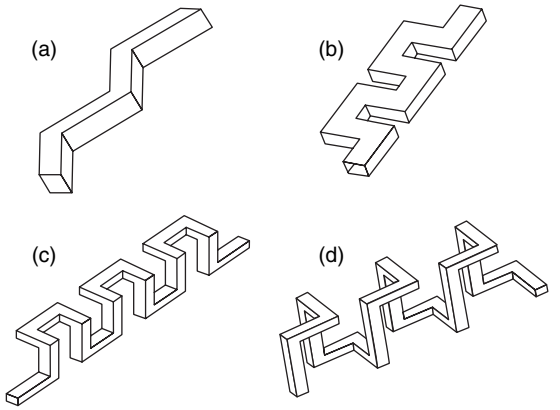


Figure 20 Samples of various arrangements of conduits to produce a 3D chaotic mixing. (a) Zigzag conduit; (b) Square-wave conduit; (c) C-shaped conduit; and (d) L-shaped conduit.

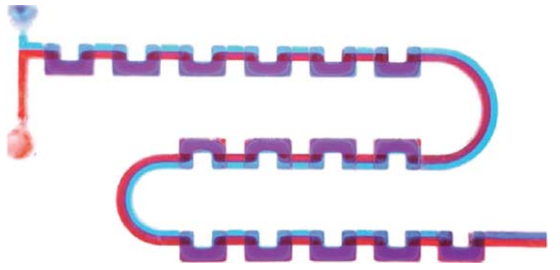


Figure 21 Experimental mixing result of a 3D C-shaped serpentine mixer at $Re = 0.2$. (Source: Xia H M, Wan S Y M, Shu C, Chew Y T 2005 Chaotic micromixers using two-layer crossing channels to exhibit fast mixing at low Reynolds numbers. *Lab Chip* **5**, 748–55.)

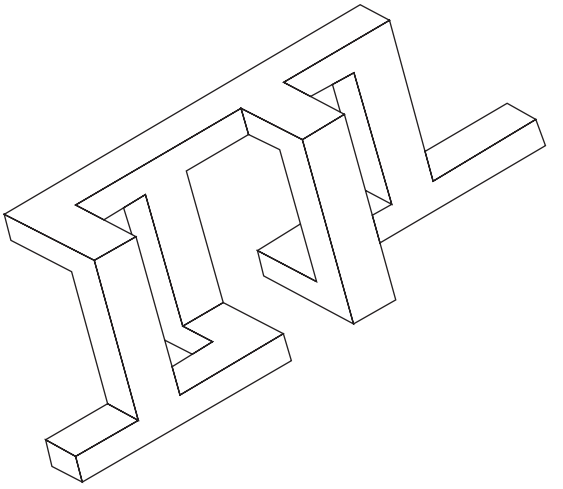


Figure 22 Schematic of an out-of-plane L-shaped conduit to produce chaotic advection at low Reynolds numbers.

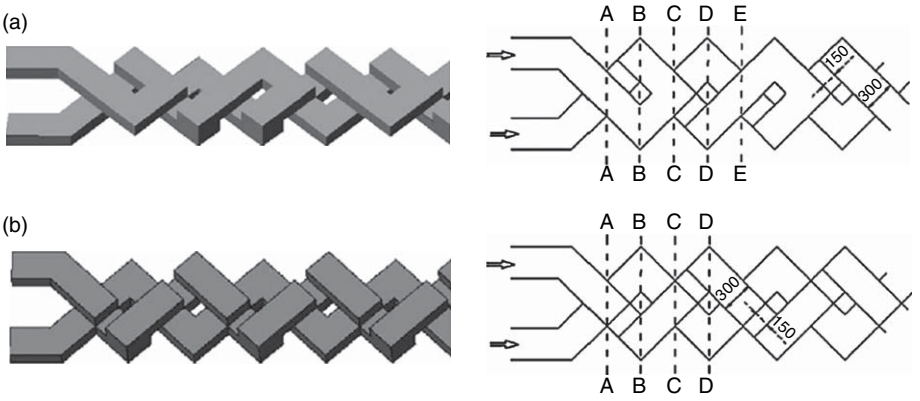


Figure 23 Configurations of the micromixers using two-layer crossing channels. (a) Model a and (b) model b. (Source: Xia H M, Wan S Y M, Shu C, Chew Y T 2005 Chaotic micromixers using two-layer crossing channels to exhibit fast mixing at low Reynolds numbers. *Lab Chip* **5**, 748–55.)

volumes. The model b (**Figure 23(b)**) has similar but simpler structures in which the fluids are alternately driven from the base-layer channel to the top-layer channel from two sides of the mixer and rejoin the main flow at the knot positions. For both types of mixers, the channel depth in different layers is 150 μm, and the channel width is 300 μm. One can fabricate such micromixers by means of lithography-based microfabrication techniques. Using polymer materials, they can also be fabricated with the laser direct writing method and thermal bonding techniques. **Figures 24** and **25** depict the experimental mixing results of the mixer models a and b, respectively, at $Re = 0.01$, and the subfigures show the detailed fluid dispersion at different positions along the mixers. The results demonstrate that chaotic advection occurs at Re of $O(10^{-2})$, and rapid mixing can be achieved at extremely low Reynolds numbers with these mixers. Compared with model a, model b exhibits faster mixing.

Following a different approach, ribs or grooves on the channel walls can lead to the formation of transverse velocity components at low Reynolds numbers (Howell *et al.* 2005, Johnson *et al.* 2002, Stroock *et al.* 2002). **Figure 26** depicts a microconduit with grooves of certain depths positioned at the bottom of the channel. These are arranged in such a way that their direction forms a certain angle with respect to the main flow direction. For a pressure-driven flow, these grooves introduce anisotropic flow resistances into an

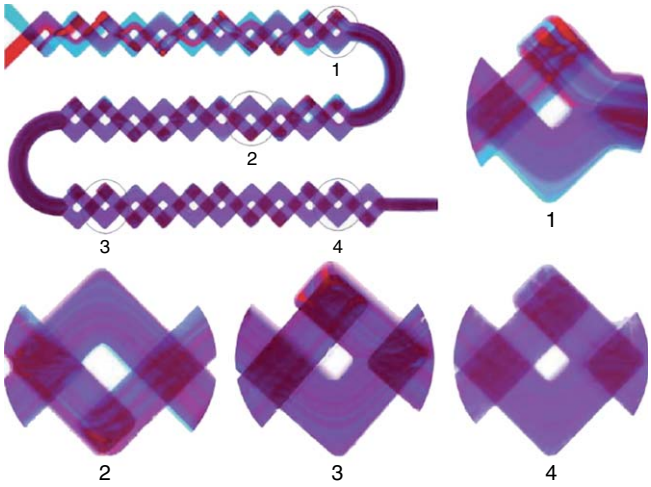


Figure 24 Experimental mixing result of the mixer (model a in **Figure 23**) at $Re = 0.01$. (Source: Xia H M, Wan S Y M, Shu C, Chew Y T 2005 Chaotic micromixers using two-layer crossing channels to exhibit fast mixing at low Reynolds numbers. *Lab Chip* **5**, 748–55.)

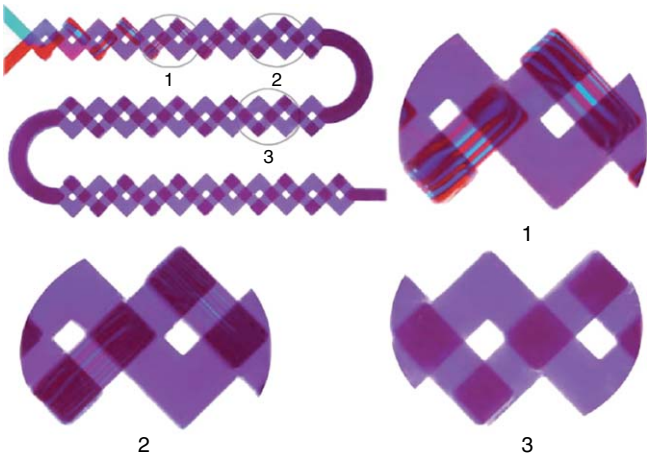


Figure 25 Experimental mixing result of the mixer (model b in **Figure 23**) at $Re = 0.01$. (Source: Xia H M, Wan S Y M, Shu C, Chew Y T 2005 Chaotic micromixers using two-layer crossing channels to exhibit fast mixing at low Reynolds numbers. *Lab Chip* **5**, 748–55.)

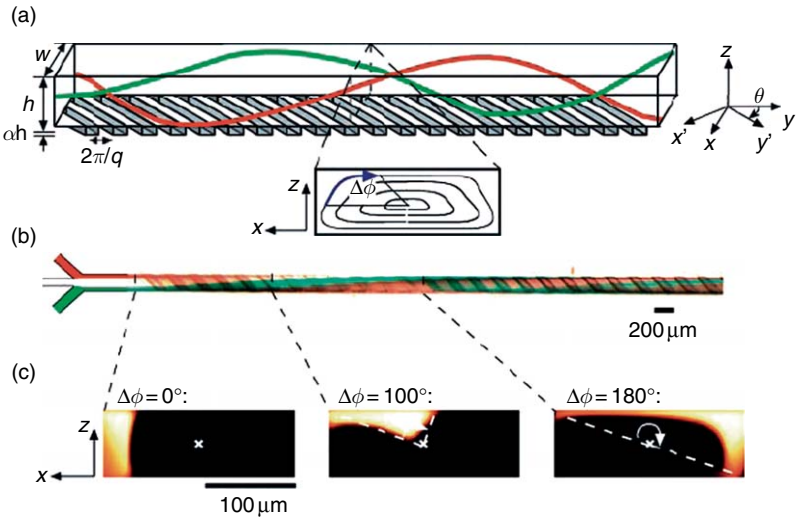


Figure 26 Transverse flows obtained with grooved surfaces at the bottom of the channel. (a) Schematic diagram of the channel with ridges and streamlines; (b) top view of experimental device and optical micrograph showing a red stream and a green stream flowing on either side of a clear stream in a channel; and (c) cross-sectional view of one fluorescent fluid. (Source: Stroock A D, Dertinger S K W, Ajdari A, Mezic I, Stone H A, Whitesides G M 2002 Chaotic mixer for microchannels. *Science* **295**, 647–51.)

otherwise isotropic system. Fluids experience a lower resistance when flowing along the ridges and valleys constituting the grooves than when flowing perpendicular to them. In this way, a rotational element is introduced, resulting in corkscrew-like fluidic streamlines. **Figure 26(a)** schematically depicts two trajectories of passive particles. The frames in **Figure 26(c)** are confocal micrographs of the vertical cross section of a channel depicted in **Figure 26(b)**. The leading edge of the fluorescent fluid clearly shows the rotation and distortion of a stream of solution that was injected along one side of the channel. Under appropriate arrangement of different types of grooves (mainly differing with respect to their orientation), the induced transverse flow can produce chaotic flows. This design is also called staggered herringbone mixer (SHM). **Figure 27** depicts the performance of the SHM mixer. **Figure 27** (*left frame*) represents the channel with staggered herringbone structure. **Figure 27** (*right frame*) depicts the distribution of fluorescent molecules in the cross sections at distances of 0.2, 0.4, 0.6, 0.8, 1.0, and 3 cm down the channel at

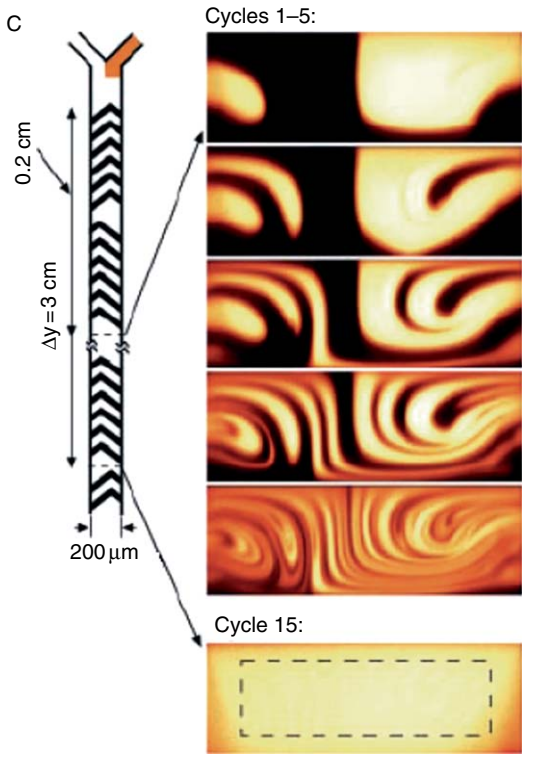


Figure 27 Performance of the staggered herringbone mixer (SHM) at $Re \sim 10^{-2}$. (*Left*) schematic diagram of a channel with a staggered herringbone structure at the bottom; (*right*) confocal micrographs that show the distribution of fluorescent molecules in the cross sections at distances 0.2, 0.4, 0.6, 0.8, 1.0, and 3 cm down the channel. (Source: Stroock A D, Dertinger S K W, Ajdari A, Mezic I, Stone H A, Whitesides G M 2002 Chaotic mixer for microchannels. *Science* **295**, 647–51.)

$Re \sim 10^{-2}$. The results demonstrate that SHM works well in the range $0 < Re < 100$. The SHM mixing is more efficient than that achieved with similar microfluidic channels devoid of internal structures like ribs and grooves. For example, the basic T-mixer requires mixing lengths of about 1 and 10 m at Péclet numbers (Pe) $= 10^4$ and 10^5, and the SHM mixer performs the same task within 1 and 1.5 cm, respectively. The mixing efficiency of the SHM mixer can be further greatly improved with ribs or grooves placed at both the top and the bottom of the channel (Howell *et al.* 2005).

Taking advantage of the induced Dean vortices in alternately curved microfluidic channels, a much simpler channel design without any internal structure has been proposed (Jiang *et al.* 2004, Schonfeld and Hardt 2004, Sudarsan and Ugaz 2006). **Figure 28** schematically depicts a four-element meander mixer, which consists of a channel with rectangular cross

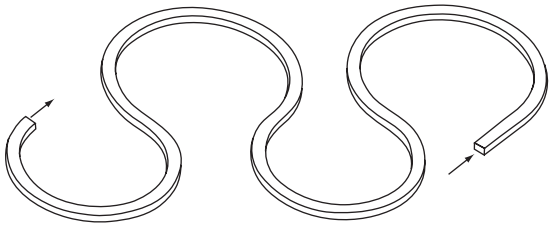


Figure 28 Schematic of four-element meander mixer.

section comprising four circular arcs and two straight inlet and outlet sections. Because of the periodic arrangement of the geometry, a periodic velocity field develops. The important design parameter is the dimensionless Dean number, κ. It has been demonstrated that when the Dean number exceeds a certain threshold value, four vortices develop, inducing a large interfacial area. **Figure 29** depicts the evolution of the species concentration (encoded in gray) at the inlet, the outlet, and at two intermediate positions within a circular arc segment for $\kappa = 150$ (*left column*) and $\kappa = 450$ (*right column*). At the inlet cross section, the species are located only in the right side of the conduit. As they flow downstream, at the cross section corresponding to $\varphi = \pi/2$, a pair of counterrotating vortices develops leading to the advection of the species into the left side of the channel, but a given fraction remains on the right side. At the cross section $\varphi = \pi$, most of the species are advected into the left side of the channel, and four vortices develop for $\kappa = 450$. At the outlet cross section $\varphi = 3\pi/2$, a steady velocity profile with four vortices is well developed for $\kappa = 450$. The interface between the two fluids is significantly elongated after the fluids flow through one circular arc segment. When the Dean number exceeds a particular threshold value, the interfacial stretching factor, denoted as λ and defined as the interface length at a certain position divided by the initial interface length, exponentially increases with the number of mixing elements. **Figure 30** depicts the experimental results of a scaled-up version of the meander mixer having 20 mixing elements with a length of 322 mm and a curved channel of a cross section of 1 mm × 1 mm. The mixing visualization was realized by employing a very fast ionic dye reaction, more precisely an aqueous solution of $Fe(NO_3)_3$ and NaSCN forming the intensely reddish Fe(III)-rhodanide according to a 1:1 stoichiometric reaction. The position 1 in **Figure 30** shows the mixing in the first and second mixing elements. The position 2 is located at 118 mm downstream and corresponds to the Nos. 7 and 8

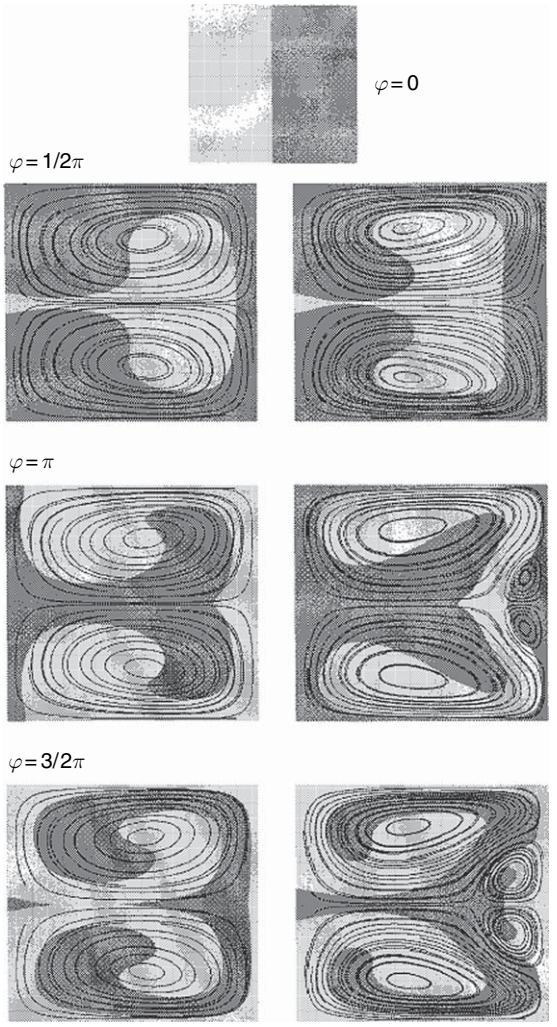


Figure 29 Evolution of species concentration within a circular arc segment at the inlet ($\varphi = 0$), two intermediate positions ($\varphi = \pi/2$ and $\varphi = \pi$), and outlet ($\varphi = 3\pi/2$) cross sections for Dean numbers $\kappa = 150$ (*left column*) and $\kappa = 450$ (*right column*). (Source: Schonfeld F, Hardt S 2004 Simulation of helical flows in microchannels. *AICHE J.* **50**, 771–8.)

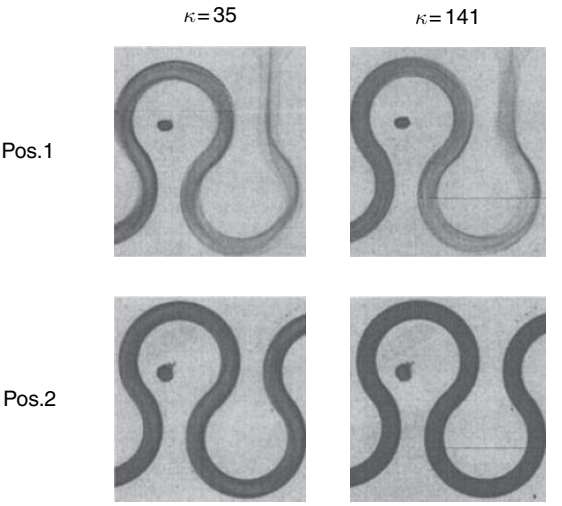


Figure 30 Experimental mixing results in the meander mixer for Dean numbers $\kappa = 35$ and141 at two different positions. (Source: Jiang F, Drese K S, Hardt S, Kupper M, Schonfeld F 2004 Helical flows and chaotic mixing in curved micro channels. *AICHE J.* **50**, 2297–305.)

mixing elements, where the two fluids are well mixed. Although the meander mixer has no internal structure within its alternately curved microfluidic channels, it requires a large space and only works under strict conditions of high Dean numbers or high Reynolds numbers.

Instead of introducing the internal ribs and grooves structure as done for the SHM mixer or introducing curved channels as in the meander mixer, another possibility consists in the patterning of materials characterized by different surface chemistry or equivalently different values of zeta potentials or surface charge densities so that recirculating transverse velocity components may arise in electroosmotically driven flows (Biddiss *et al.* 2004, Lee *et al.* 2005, Ng *et al.* 2004, Stroock *et al.* 2000). Before explaining the generation of these recirculation flows in some more detail, let us recall a few fundamental elements in surface chemistry science. Most of the solid surfaces are likely to carry electrostatic charges due to broken bonds and surface charge traps. For example, silicon-based materials are usually negatively charged, and Al_2O_3-based materials are positively charged. When a liquid containing a small amount of ions is brought into contact with such a solid boundary, counterions in the liquid will be strongly attracted to the charged solid surface forming an immobile compact layer at the solid–liquid interface. This layer is called the Stern layer. Outside the Stern layer, counterions and co-ions are accumulated following a Boltzmann statistics in a thin liquid layer, which is called the diffuse part of the electrical double layer (EDL). The thickness of the EDL is typically on the order of a few nanometers depending on the electrolyte concentration. Away from the EDL, the electrolyte is neutral and local ionic concentrations equal bulk values. This ionic charge separation within the EDL next to the solid wall (as compared to the electroneutral bulk electrolyte solution) causes either a positive or a negative potential difference across the EDL, denoted as zeta potential, sometimes also called electrokinetic potential. The magnitude of the zeta

potential depends, among other things, on the characteristics of the solid (surface chemistry) and the liquid (dielectric permittivity, viscosity, etc.). In the presence of an external electric field, the counterions in the EDL are attracted toward the oppositely charged electrode and drag the liquid along. In other words, the electric field, through its action on the counterions, creates a body force that, in turn, induces fluid motion, which is called EOF. When the thickness of the EDL is much smaller than the dimensions of the conduit, the EOF can be described by specifying a slip velocity at the wall, so the body force is nearly coincident with the bounding surface. EOFs have several important advantages over pressure-driven flows. They do not require any moving components like pumps and valves, and the fluid speeds in EOFs are independent of the transverse tube or channel dimension over a wide range of conditions, making this technique extensible to extremely small physical scales.

For steady, 1D and fully developed flow in a microfluidic channel, the EOF velocity, noted u, is simply given by the Smoluchowski equation written as below:

$$u = -\frac{\varepsilon_o \varepsilon_r \zeta}{\mu} E \qquad [1]$$

where ζ is the zeta potential; $\varepsilon_o \varepsilon_r$ and μ are the dielectric permittivity and the dynamic viscosity of the medium, respectively and E denotes the magnitude of the applied electric field. Relationship [1] is valid for sufficiently thin double layers, and is further strictly applicable for low values of the zeta potential where the Debye–Hückel approximation is correct. From expression [1], it is readily understood that the magnitude and the direction of the electroosmotic velocity can be controlled by the magnitude and the polarity of the zeta potential. Therefore, by selectively patterning negatively, neutrally, or positively charged material patches on the channel walls, a wide range of flow patterns in microfluidic channels can be created. For example, **Figure 31** schematically depicts the EOF field in the vicinity of the channel wall containing both negatively and positively charged materials. Since the negatively charged surfaces attract positive ions that drag the bulk fluid moving toward the cathode and the positively charged surfaces attract negative ions dragging the liquid toward the anode, the interaction between the two flows results in a local circulating flow, which can be used to accelerate and enhance mixing.

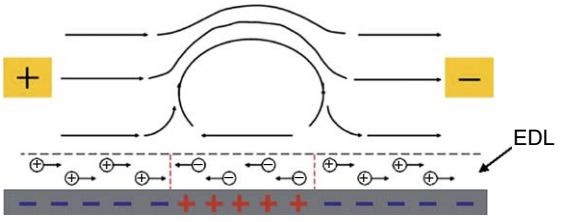


Figure 31 Illustration of the electroosmotic flow (EOF) field in the vicinity of a channel wall patterned with alternately negative and positive surface charges.

Using the rapid prototyping/soft lithography technique, the channel walls can be patterned with various surface charge configurations (Biddiss *et al.* 2004). **Figure 32** depicts some configuration examples as proposed by Biddiss *et al.* (2004). The gray zones of the channel walls have a positive surface charge, and the rest has a negative surface charge. Thus, the local electric double layer, or ionic concentration fields, and the local flow fields are different from patch to patch. **Figure 33** depicts the numerical and experimental mixing results in a microchannel with a uniform surface charge distribution (**Figure 33(a)**) and in a heterogeneous microchannel with six offset staggered patches (**Figure 33(b)**). Mixing experiments were conducted at applied electric fields ranging between 70 and 555 V cm^{-1}, which corresponds to Reynolds numbers of 0.08 and 0.7 and Péclet numbers of 190 and 1500, respectively. The numerical predictions agree very well with the experimental results. In a microfluidic channel with homogeneous distribution of the surface charge on the channel walls, the mixing depends solely on molecular diffusion process and requires 22-mm channel length to achieve 95% mixing at an electric field of 280 V cm^{-1}. For the staggered configuration, the formation of unsymmetrical concentration gradients indicates the presence of local

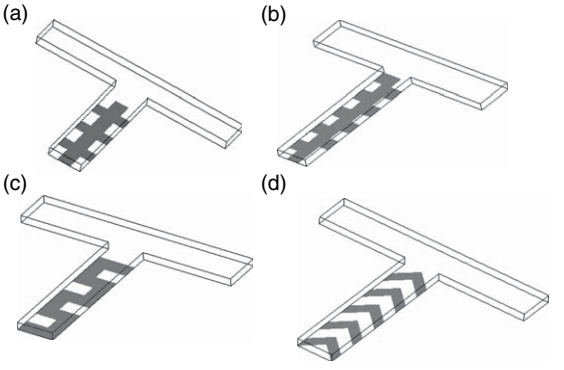


Figure 32 Schematic representation of various surface charge pattern configurations. (a) In-plane pattern; (b) staggered pattern; (c) serpentine pattern; and (d) herringbone pattern.

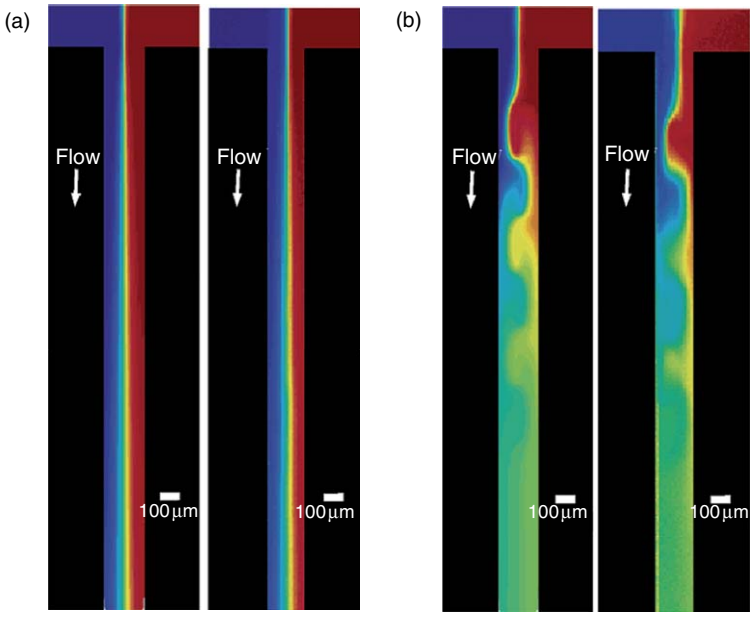


Figure 33 Mixing results (a) in a homogeneous microchannel and (b) in a heterogeneous microchannel with six offset staggered patches. The applied electric field is 280 V cm^{-1}. (Source: Biddiss E, Erickson D, Li D 2004 Heterogeneous surface charge enhanced micromixing for electrokinetic flows. *Anal. Chem.* **76**, 3208–13.)

flow circulations, which advect a portion of the mixed downstream fluid back to the unmixed upstream region, and it only requires a channel length of 2.6 mm to meet 95% mixing. In contrast to previous passive chaotic mixers, this passive electroosmotic mixer can easily create a complex flow field in a straight microchannel without the necessity for complex channel geometry and internal structure.

2.10.3 Active Mixers

To facilitate rapid mixing, some active mixers have been developed and can be typically classified according to their stirring mechanisms, namely pressure perturbation mixers, thermosiphon-based mixers, magnetic mixers, dielectrophoretic mixers, peristaltic mixers, acoustic mixers, magnetohydrodynamic (MHD) mixers, and electroosmotic mixers. In the following, the basic functioning and the performance of each of these mixers are reviewed.

2.10.3.1 Pressure Perturbation Mixer

By using a standard MEMS technology, a micromixer with one main channel and one pair of side channels, as shown in **Figure 34(a)** (Lee 2002, Lee *et al.* 2000, 2001), can be fabricated. Two fluids of different natures are injected from the inlet with the same flow rate. The pair of side channels is connected to a pair of high- and low-pressure sources via a computer-controlled high-speed solenoid valve to achieve a kind of source–sink mechanism. Time-periodic pressure perturbations are applied perpendicular to the main stream. The time-averaged mass flow rate from the side channels is kept to zero. At small perturbation amplitudes, the material line undergoes weak oscillations as it passes through the intersection region. As the perturbation amplitude increases, the oscillation amplitude also increases. Chaotic advection appears for sufficiently large perturbation amplitudes. **Figure 34(a)** depicts the evolution of an interface between two fluids as obtained by numerical simulations (Lee *et al.* 2001). In the upstream region, the two fluids are well separated. Once the pressure perturbation is applied, lobe-like distortions of the interface are induced, which facilitates rapid mixing. **Figure 34(b)** depicts the experimental result and the lobe-like distortions of the interface, as observed in practice. **Figure 34(c)** depicts the chaotic mixing by the micromixer developed by Lee *et al.* (2001) for particular values of perturbation amplitude and frequency. The mixing performance can be significantly improved with

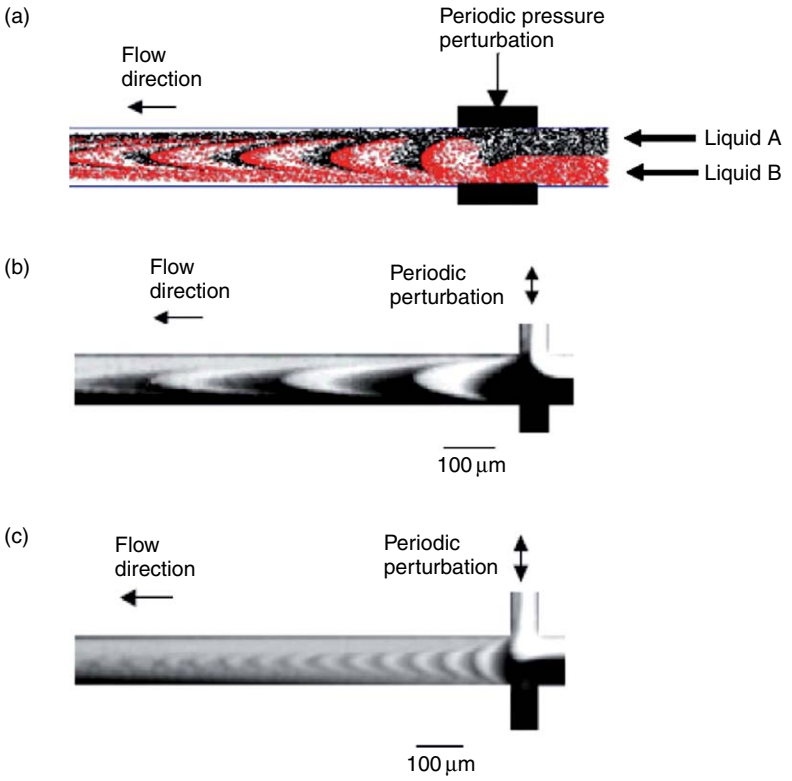


Figure 34 (a) Lobe-like distortions of the interface as induced by periodic pressure perturbations applied from side channels (simulation results); (b) experimental snapshot showing a highly convoluted interface; and (c) chaotic mixing as a result of pressure perturbation in a micromixer. (Source: Lee Y K, Deval J, Tabeling P, Ho C M 2001 Chaotic mixing in electrokinetically and pressure driven micro flows. *IEEE 14th Int. Conf. Micro Electro Mechanical System (MEMS 2001)*, Interlaken, Switzerland, pp. 483–6.)

pressure perturbations applied to multiple side channels (Niu and Lee 2003).

Recently, Niu *et al.* (2006) reported the design and fabrication of a novel microfluidic mixer chip, whose mixing mechanism is based on pressure perturbations applied to side channels orthogonal to the mixing channel. Different from the mixer evoked before and developed by Niu and Lee (2003) where the flow in the mixing channel is perturbed by means of mechanical pulsations, the pulsating pressure perturbation is now generated by controlling a smart material, an electrorheological (ER) fluid. **Figure 35(a)** schematically depicts the mixer chip, which consists of four PDMS layers fabricated using the soft lithography techniques. The fluid mixing channels (38 μm in depth and 200 μm in width) together with the main channel and six pairs of orthogonal side channels are located on layer I. Each of the six pairs of orthogonal side channels is separated by 200 μm from the neighboring pair. Layer II is a thin membrane layer with two pairs of electrodes. An ER fluid channel 200 μm in depth and 400 μm in width is located on layer III. The ER fluid channel consists of two branches connected to two ER fluid reservoirs through inlet and outlet tubes. Layer IV is a cover layer with two pairs of electrodes. The electrodes on layers II and IV are located along each branch of the ER fluid channels. The flow of the ER fluid can be slowed or stopped if a sufficient DC electric field is applied to any pair of the electrodes as a result of the so-called electrorheological effect, namely the coalescing of the nanoparticles into chains along the direction of the electric field that is orthogonal to the flow direction. The DC electric field signals on these two pairs of electrodes are in the form of two square-wave pulses with opposing phases, leading to a time-dependent pressure condition in the ER fluid channel between the two pairs of electrodes. Such temporal pressure variation is then transferred to the control reservoirs at the end of each side channel in layer I through the deformation of a 40-μm-thick PDMS diaphragm located in layer II. The induced pressure variation acting on the control reservoirs through the diaphragm deformation can be adjusted via the applied electric field strength and frequency. **Figure 35(b)**

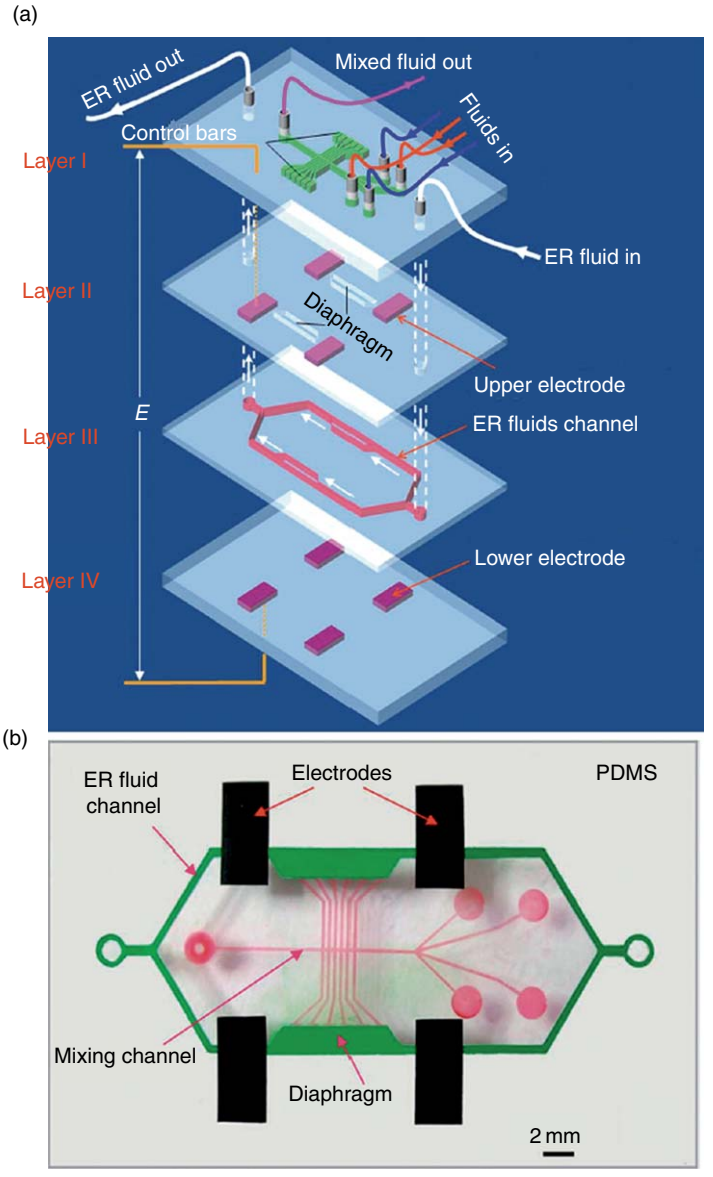


Figure 35 (a) Schematic representation of the mixer design and construction. Layer I is the mixing channel layer, layer II the thin membrane layer, layer III the electrorheological (ER) fluid channel layer, and layer IV the cover layer. Parallel electrodes are located on layers II and IV. (b) A photograph of the micromixer chip. (Source: Niu X, Liu L, Wen W, Sheng P 2006 Active microfluidic mixer chip. *Appl. Phys. Lett.* **88**, 153508.)

shows the image of the micromixer chip. Figure 36 depicts the experimental results of the mixer. In the absence of pressure perturbations in the side channels (Figure 36(a)), two fluid streams are injected into the mixing channel with a velocity of $200\,\mu m\,s^{-1}$, and a clear interface between the two fluid streams is observed, which indicates a poor mixing. In the presence of an electric field applied to the electrodes (Figure 36(b)), the main flow is perturbed by the induced pressure perturbations acting on the fluids in the side channels. The interface line is then stretched and folded intensively, and there is no discernible ink blob existing after a distance of 1.8 mm in the downstream region of the main channel.

2.10.3.2 Thermosiphon-Based Mixer

Usually buoyancy effects in microfluidic devices are very small because of the small length scales and small temperature variations. However, certain

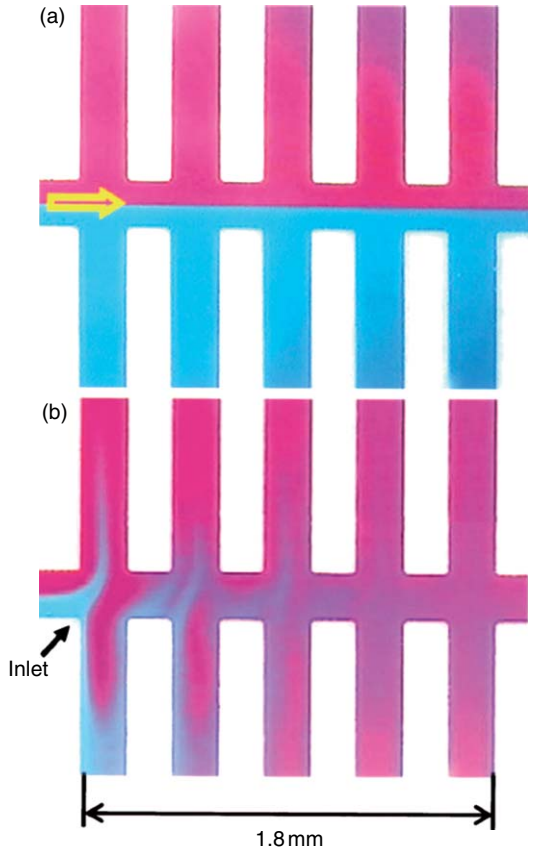


Figure 36 Experimental results of the pressure perturbation micromixer. Optical images of the mixing channel in the (a) absence and (b) presence of pressure perturbations applied to the side channels. (Source: Niu X, Liu L, Wen W, Sheng P 2006 Active microfluidic mixer chip. *Appl. Phys. Lett.* **88**, 153508.)

processes such as polymerase chain reaction (PCR) for DNA amplification require large temperature variations that may be sufficient to induce significant flow velocities even in micrometer-sized conduits. For example, for DNA amplification, Chen *et al.* (2004) and Krishnan *et al.* (2004) successfully developed a thermosiphon-based PCR reactor in which the heating source used for PCR reactions also drives the flow in a closed loop without the need of any mechanical pumps. **Figure 37** depicts the natural convection in a closed cylindrical cavity mounted vertically with the bottom surface set at a constant temperature of 97°C and the top surface is maintained at 61°C. The liquid close to the bottom surface will be less dense than the fluid close to the top surface, thus generating a buoyancy force, which, in turn, induces circulating flows. Natural convection occurs when the dimensionless Rayleigh number

$$Ra = \frac{g\alpha(T_2 - T_1)h^3}{v\kappa}$$

which represents the ratio of buoyant forces driving the instability to diffusive restoring forces acting in opposite direction, exceeds a certain threshold value. Here, g is the gravitational acceleration, α, v, and κ are the thermal expansion coefficient, the kinematic viscosity, and the thermal diffusivity of the fluid, respectively, h is a characteristic length scale, which is the height of the cavity in this case, T_1 and T_2 are the temperatures of the top and bottom surfaces of the cavity, respectively. The experimental and numerical results shown in **Figure 37** indicate that very complicated flow motions can be induced upon variations of the Rayleigh number, the aspect ratio h/d, or both.

When the temperatures T_1 and T_2 are time-dependent, a time-dependent natural convection results and can be used to produce chaotic advection (de la Cruz and Ramos 2006). **Figure 38** schematically depicts an enclosure with the temperatures of the left half of the upper wall periodically reduced and those of the right half of the bottom wall cyclically increased. The temperature changes are outphased. The two vertical walls are thermally insulated. The dimensionless temperatures of the right half of the upper wall and the left half of the bottom wall are maintained at $T = 0$. The temperatures shown in **Figure 38** are dimensionless, and are defined by $T = (\Theta - \Theta_m/\Delta\Theta)$. Here Θ is the dimensional temperature. Θ_H and Θ_C are the maximum and minimum wall temperatures, respectively, and $\Theta_m = (\Theta_H + \Theta_C)/2$ and $\Delta\Theta = \Theta_H - \Theta_C$. The first period of $T_1(t)$ and $T_2(t)$ are defined by the following expressions:

$$T_1(t) = \begin{cases} 0, & \text{for } 0 \le t < \pi \\ -0.5\sin^2(4t - 3\pi), & \text{for } \pi \le t \le \dfrac{9\pi}{8} \\ -0.5, & \text{for } \dfrac{9\pi}{8} \le t \le \dfrac{15\pi}{8} \\ -0.5\sin^2(4t - 6\pi) & \text{for } \dfrac{15\pi}{8} \le t < 2\pi \end{cases} \quad [2]$$

$$T_2(t) = \begin{cases} 0.5\sin^2(4t), & \text{for } 0 \le t \le \dfrac{\pi}{8} \\ 0.5, & \text{for } \dfrac{\pi}{8} \le t \le \dfrac{7\pi}{8} \\ 0.5\sin^2(4t - 3\pi), & \text{for } \dfrac{7\pi}{8} \le t \le \pi \\ 0, & \text{for } \pi \le t < 2\pi \end{cases} \quad [3]$$

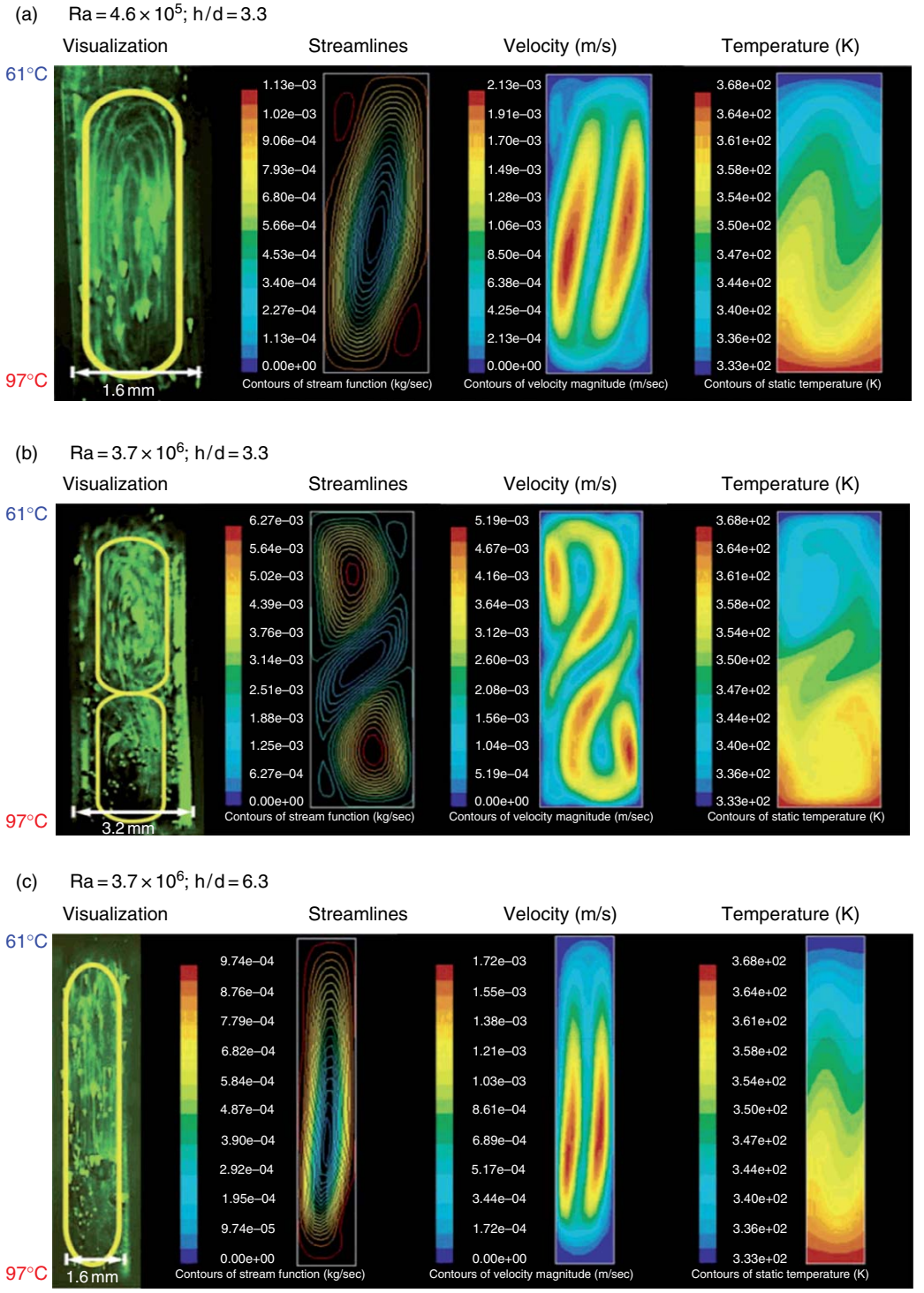


Figure 37 Experimental and numerical results of natural convection in high aspect ratio cavities heated from below and cooled from above. (Source: Krishnan N, Agrawal N, Burns M A, Ugaz V M 2004 Reactions and fluidics in miniaturized natural convection systems. *Anal. Chem.* **76**, 6254–65.)

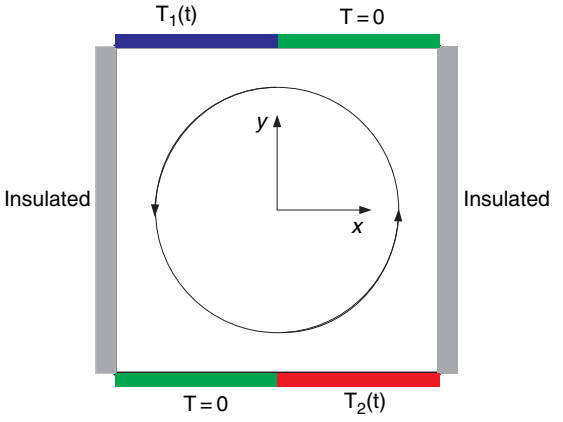


Figure 38 Schematic of a cavity filled with liquid to be mixed and thermal boundary conditions imposed on the walls of the cavity.

In eqns [2] and [3], t is the dimensionless time and the timescale is L^2/α. The temperatures T_1 and T_2 are alternated with a period of 2π. The oscillatory wall temperatures imposed on the left half of the top wall and the right half of the bottom wall yield the formation of alternating ascending and descending thermal plumes in the regions close to the vertical walls. The major feature of the induced transient natural convection is that the center of the vortex defined by the point of zero velocity varies with time. The mixing efficiency of the flow is evaluated by tracing the trajectories of 5×10^4 passive particles initially positioned according to a horizontal line at $y = 0$ where y is defined in **Figure 38**. **Figure 39** depicts the stretching and folding of the passive particles after 1/4, 1/2, 3/4, 1, 5, and 50 periods. The local stretching occurs mostly near the vertical walls where the largest stresses are generated by the convective plumes. The folding takes place near the horizontal walls where the fluid turns. After 5 periods, the material blob is spread across most of the area of the cavity. Witness that this analysis only considered the effect of the induced natural convection on the advection of passive tracers and neglected the molecular diffusion process that is highly dependent on the temperature. The actual mixing performance may be different from the theoretical predictions as obtained by de la Cruz and Ramos (2006).

When the temperature variations in minute systems are not sufficiently large to induce significant natural convection, one can mount the device on a rotating system such as a laboratory on a computer disk (CD) (Chen *et al.* 2004). In this case, the

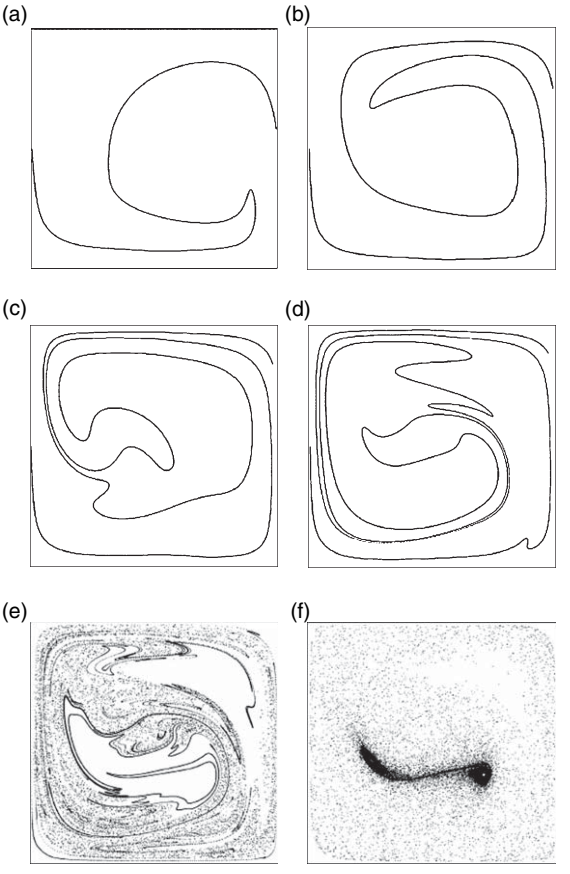


Figure 39 Deformation of a material blob with 5×10^4 passive particles after (a) 1/4; (b) 1/2; (c) 3/4; (d) 1; (e) 5; and (f) 50 cycles. (Source: De la Cruz L M, Ramos E 2006 Mixing with time dependent natural convection. *Int. Commun. Heat Mass Transfer* **33**, 191–8.)

centripetal acceleration, $w^2 r$, will replace the gravitational acceleration (g) in the expression of the Rayleigh number. In the above, w is the angular rotation speed and r is the distance between the center of rotation and a point in the microfluidic channel. Since very high accelerations ($>>9.8\ \mathrm{m\,s^{-2}}$) can be obtained in the rotating system, one can significantly reduce the size of the device and the temperature variations required to induce appreciable natural convection.

2.10.3.3 Magnetic Mixer

A magnetic force-based chaotic micromixer was developed for mixing biofluids with magnetic beads (Suzuki 2003, Suzuki *et al.* 2004). **Figure 40** schematically depicts the concept of the magnetic bead-based microseparation system of biomolecules. Two

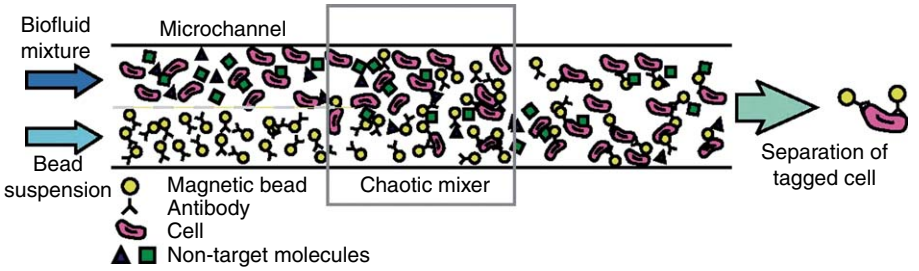


Figure 40 Schematic of magnetic beads-based microcell sorting system. (Source: Suzuki H, Ho C M, Kasagi N 2004 A chaotic mixer for magnetic bead-based micro cell sorter. *J. Microelectromech. Syst.* **13**, 779–90.)

fluid streams, a suspension of biomolecules and a suspension of magnetic beads, meet each other in the microchannel and are mixed in the mixing region. In this region, magnetic beads move around in a complicated way to enhance the biological binding of the magnetic beads to target biomolecules by a local and time-varying magnetic field as generated by microelectrical conductors embedded in the substrate beneath the conduit. Eventually, the tagged biomolecules are collected by using either a permanent magnet or an electromagnet at the downstream.

The developed magnetic micromixer previously described consists of microconductors embedded in the substrate and a serpentine-shaped microchannel formed on the surface (**Figure 41(a)**). The embedded high aspect ratio conductors allow a relatively large current to generate a magnetic field to move magnetic beads. A pair of current-carrying conductors works as a basic unit of a magnetic field source. By shifting the working conductors, it becomes possible to push and pull magnetic beads between high- and low-velocity regions. During the first phase, the

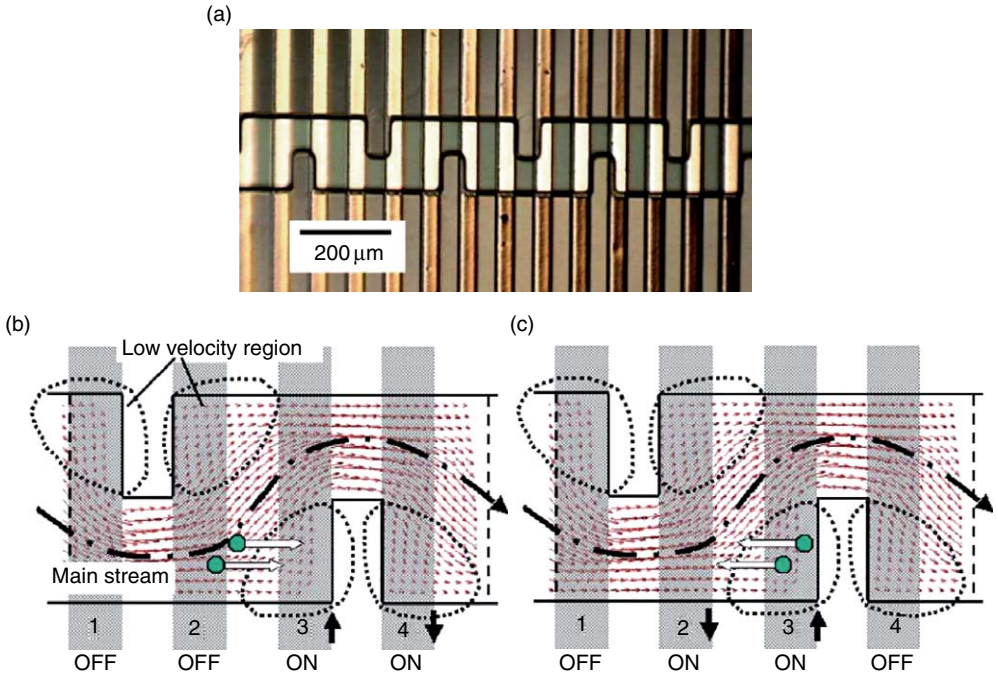


Figure 41 (a) An example of a magnetic mixing device. (b) Flow field and motion of magnetic beads (represented by two dots in the figure) when conductors 3 and 4 are turned on. (c) Flow field and motion of magnetic beads (represented by two dots in the figure) when conductors 2 and 3 are turned on. (Source: Suzuki H 2003 Development of chaotic micro-mixer using magnetic beads. Ph.D. thesis, The University of Tokyo.)

conductors 1 and 2 are turned off, the conductors 3 and 4 are turned on, and some magnetic beads (two dots in **Figure 41(b)** and **41(c)**) are attracted toward the inner edges and pulled into the low-velocity region at the corner of the channel (marked with dotted circles). In the second phase, conductors 2 and 3 are turned on, and the magnetic beads are pulled back to the mainstream. Therefore, by switching a pair of working conductors in a certain manner, it is possible to create stretching and folding by pushing and pulling magnetic beads between high- and low-velocity regions.

Figure 42 depicts a snapshot of the flow when all the conductors are turned off. Deionized water is flowing in the upper side, and a water suspension of 1-μm magnetic beads is flowing in the lower side. The volume ratio of beads is 1.6%, and the flow rate is 200 nl min^{-1}. The clear interface between the two fluid streams demonstrates that there is no mixing effect even though the channel is serpentine-shaped (the Reynolds number is about 10^{-2}). **Figure 43** depicts the sequential images of magnetic beads when the phase shift signal is applied to the conductors (the time period of one phase is 1 s). The time interval between each sequential image is 1 s. The magnetic beads are attracted sequentially by pairs of electrodes, and eventually, magnetic beads spread uniformly in the spanwise direction at the downstream. A similar magnetic mixer has been developed by Biswal and Gast (2004) in which rotations of paramagnetic linked chains of colloidal particles in a rotating magnetic field induce secondary flows to mix fluids.

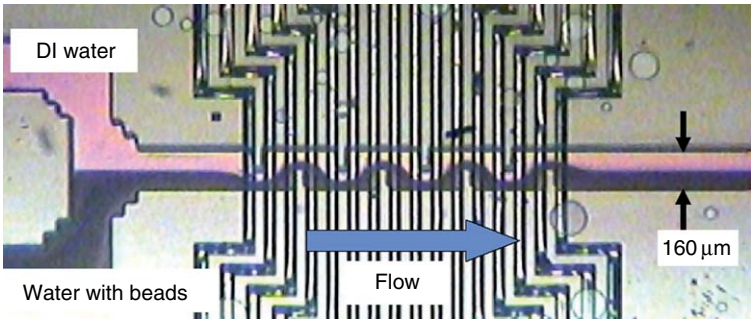


Figure 42 Experimental result of the magnetic mixer when no magnetic force is applied. (Source: Suzuki H, Ho C M, Kasagi N 2004 A chaotic mixer for magnetic bead-based micro cell sorter. *J. Microelectromech. Syst.* **13**, 779–90.)

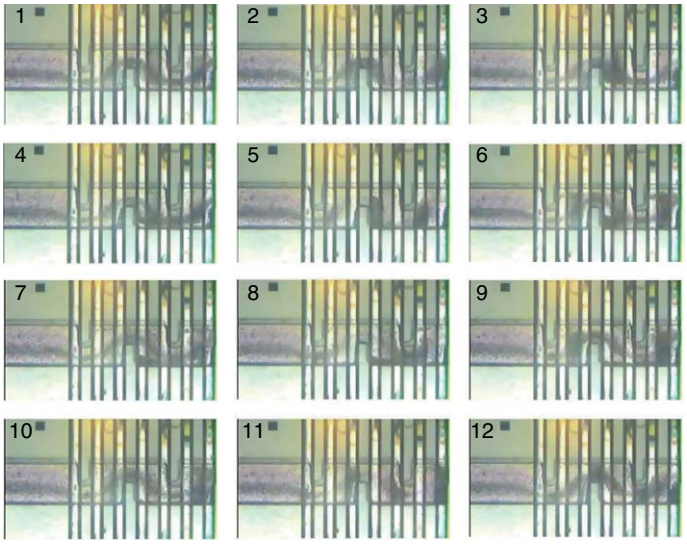


Figure 43 Experimental results of the magnetic mixer when the phase shift signal is applied to the conductors. (Source: Suzuki H 2003 Development of chaotic micro-mixer using magnetic beads. Ph.D. thesis, The University of Tokyo.)

The major disadvantages of the previous magnetic mixers are the high electric currents (i.e., 0.5 – 1 A) needed to generate a strong magnetic field, a high power consumption (about 1 W), and thus a significant Joule heating. Recently, Grumann *et al.* (2005) proposed another novel magnetic mixer in which a temporal magnetic force is generated by the spinning motion through a static magnetic field. **Figure 44** schematically depicts the magnetic mixer, which has an on-disk mixing chamber connected to two symmetric inlet reservoirs and an air vent via microchannels. Hydrophobic valves are embedded in the inlet channels to avoid uncontrolled inflow. The round mixing chamber with a volume of 25 µl connects to an outlet port. To prevent the beads from entering the outlet channel, its depth is smaller than the diameter of the beads. Therefore, the channel acts as a geometrical barrier and retains the beads inside the mixing chamber. A set of conventional permanent magnets resting in the lab frame is specifically aligned in order to periodically divert the magnetic beads. During operation of the mixer, the on-disk mixing chamber is first preloaded with a well-defined number of magnetic beads. Then, two liquids are filled into the inlet reservoirs and the disk is set into a spinning motion with a PC-controlled frequency curve. Instead of applying a time-varying magnetic field, from the viewpoint of the magnetic beads, a temporal magnetic force is generated by the spinning motion through a static magnetic field. Consequently, the beads are periodically deflected inbound and outbound, which induces chaotic advection to the liquid phase due to Stokes drag. **Figure 45** depicts the experimental results, and an ultimate mixing time of <1 s can be achieved with the combination of magnetic and centrifugal forces.

2.10.3.4 Dielectrophoretic Mixer

Dielectrophoresis (DEP) is defined as the lateral motion imparted on uncharged, polarizable particles as a result of polarization induced by nonuniform electric fields (Jones 1995). The DEP force results from the interaction between the nonuniform electric field and the induced dipole moment. Depending on the dielectric properties of the particle and the medium, the particle motion is directed either toward (positive DEP) or away from (negative DEP) regions of high electric field intensity. DEP can be achieved in both DC and AC electric fields, and it is particularly well suited for microfluidic applications and nanoassembly. One can readily control the direction of the electric field by judicious patterning of the electrodes on the substrate, and high-intensity electric fields can be obtained with relatively low potential differences, given the small gaps between the actuating electrodes.

The idea of perturbing the main flow with DEP force to enhance mixing was originally proposed by Deval *et al.* (2002) and Lee *et al.* (2001). **Figure 46** depicts the top view of the DEP micromixer developed by Deval *et al.* (2002). The mixer consists of two side walls located on opposite sides of the main channel 50 µm wide, and each of them is connected to four distinct electrodes at the bottom. Witness that the electrodes are isolated from the liquid by an

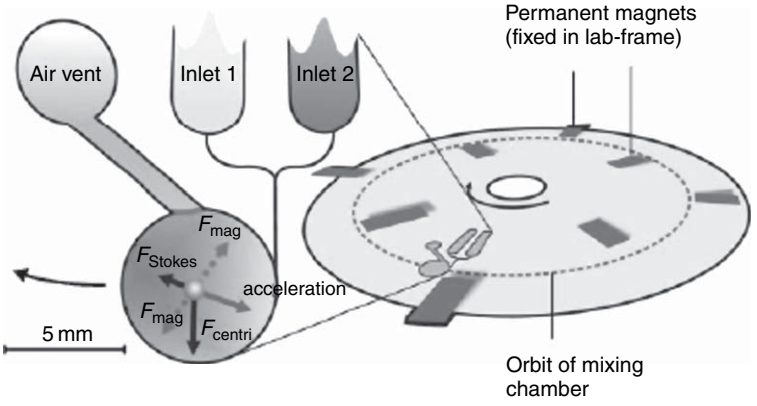


Figure 44 Schematic of a magnetic mixer on a rotating microfluidic disk with a set of permanent magnets resting in the lab frame. (Source: Grumann M, Geipel A, Riegger L, Zengerle R, Ducrée J 2005 Batch-mode mixing on centrifugal microfluidic platforms. *Lab Chip* **5**, 560–5.)

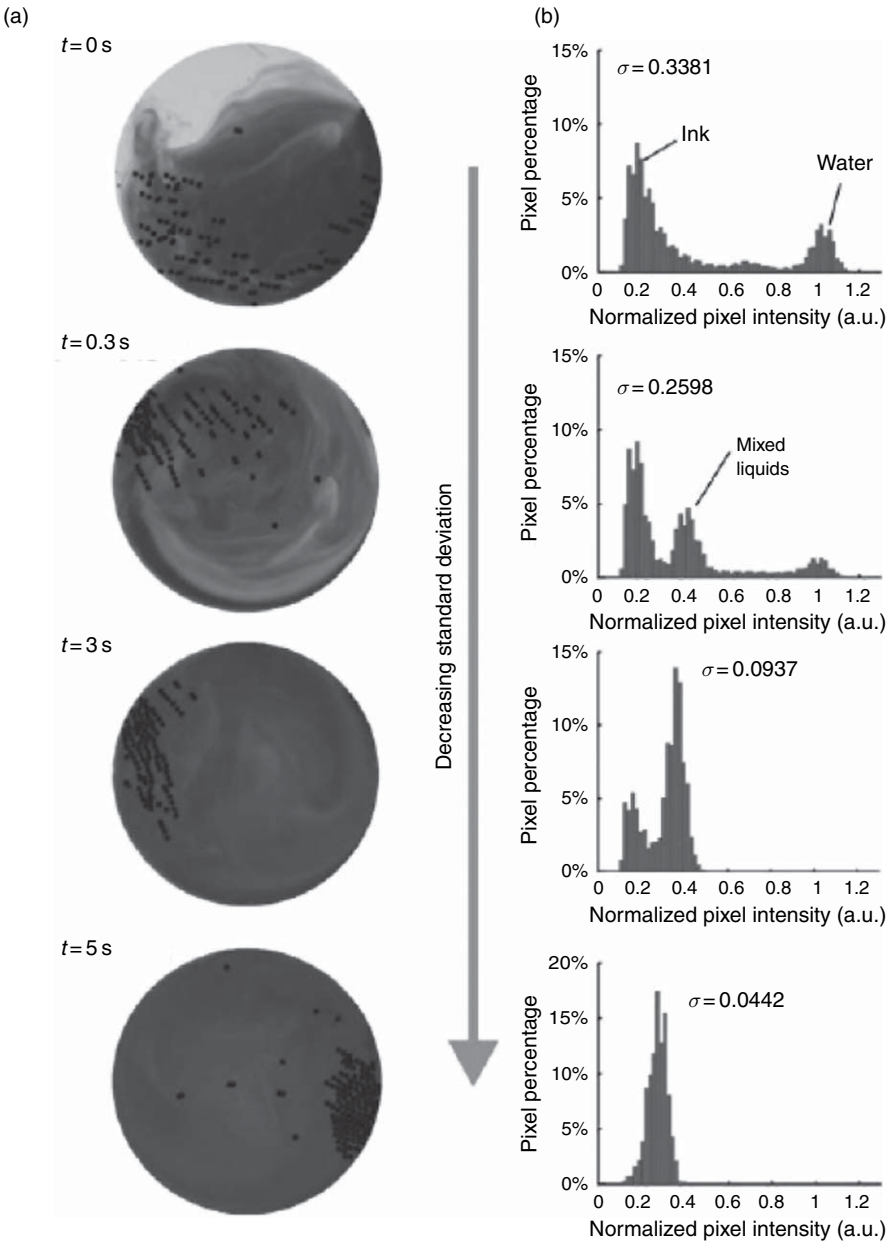


Figure 45 Experimental mixing results. (a) Images and (b) intensity histograms. (Source: Grumann M, Geipel A, Riegger L, Zengerle R, Ducrée J 2005 Batch-mode mixing on centrifugal microfluidic platforms. *Lab Chip* **5**, 560–5.)

Figure 46 Top view of the dielectrophoresis (DEP) mixer. (Source: Deval J, Tabeling P, Ho C-M 2002 A dielectro-phoretic chaotic mixer. *Proc. 15th IEEE Int. Conf. Micro Electro Mechanical Systems 2002*, Las Vegas, NV, USA, pp. 36–9.)

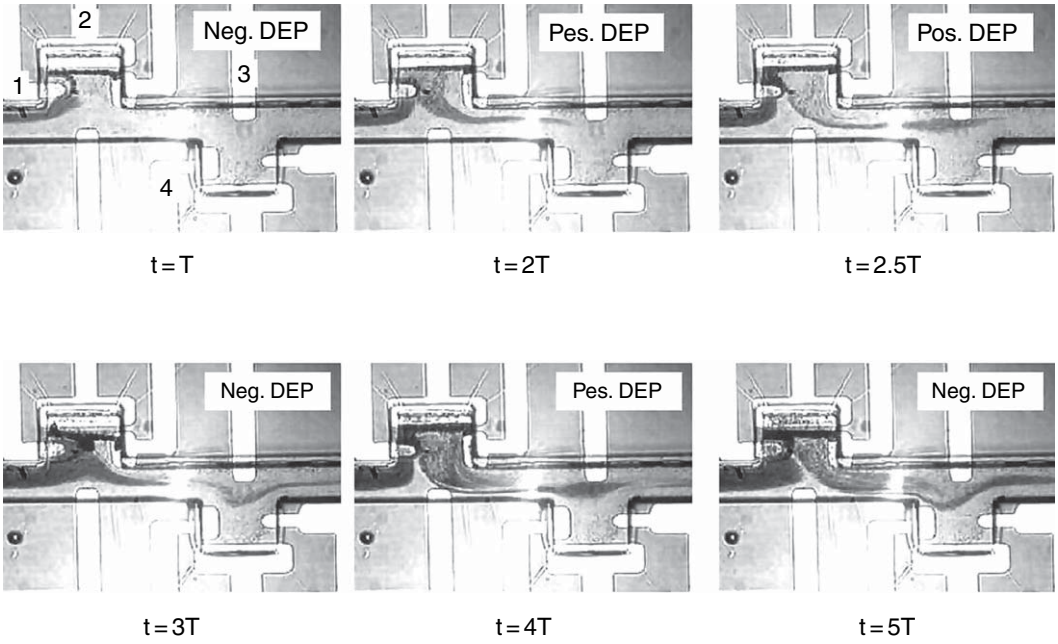


Figure 47 Experimental results of the dielectrophoresis (DEP) mixer. An AC voltage of ± 10 V is applied between electrodes 1 and 2 and between 3 and 4. AC frequency switches between 700 kHz (positive DEP) and 15 MHz (negative DEP) at a rate of $T = 0.5$ s. (Source: Deval J, Tabeling P, Ho C-M 2002 A dielectro-phoretic chaotic mixer. *Proc. 15th IEEE Int. Conf. Micro Electro Mechanical Systems 2002*, Las Vegas, NV, USA, pp. 36–9.)

insulating layer. Figure 47 depicts the successive snapshots of polystyrene particle positions over a total period of 2.5 s when ± 10 V AC is applied between electrodes 1 and 2 and between 3 and 4. The main flow is from left to right. When the particles enter the first cavity, positive DEP pulls them into the low-velocity region, attracted by electrode 1. As the frequency switches, they are repelled from the electrode 1 edges, and particles in the main stream of the flow are dragged downstream by the fluid. As the frequency switches back to the positive DEP, incoming particles are attracted upward, while a folding is created in the cavity as the particles far from the electrode experience a force too small to overcome fluid drag. The combination of the two opposite motions generates folding and stretching and greatly accelerates the homogenization of the particle distribution.

2.10.3.5 Peristaltic Mixer

The peristaltic mixer consists of a fluid-filled cavity bound from below and above by flexible membranes. By printing electrical conductors on these membranes and by passing electric currents through these conductors in the presence of a magnetic field, one can cause the membranes to vibrate with relatively large amplitudes in a prescribed way. Because the conductors can be shaped using microfabrication technology, one may induce fairly complicated motions. For example, by appropriate phasing of the current, one can induce traveling waves in the membranes, which, in turn, will cause peristaltic pumping in the fluid. The vibrating membrane can also be actuated by other means. The end walls serve to restrict the flow and thus force a cellular flow pattern. The induced cellular motion can be used to enhance mixing in a closed cavity (Carlsson *et al.* 2005, Selverov and Stone 2001, Yi *et al.* 2002a, b).

Figure 48 depicts the evolution of a material blob at times $t = 0, 600\,T$, and $1200\,T$ where T is the period of the wave in a peristaltic mixer with wave number $k = 3.8$, wave amplitude $\varepsilon = 0.01$, and Reynolds number, $Re = 25$ (Yi *et al.* 2002a). Initially, a vertical material blob is positioned in the range $1.5 < x < 2$. One can think of two materials, one dark and one light, being stirred. As time goes by, the interface between the two materials stretches and deforms. The thickness of the dark blob reduces dramatically thereby increasing the impact of molecular diffusion. Right on the membranes, the material does not move

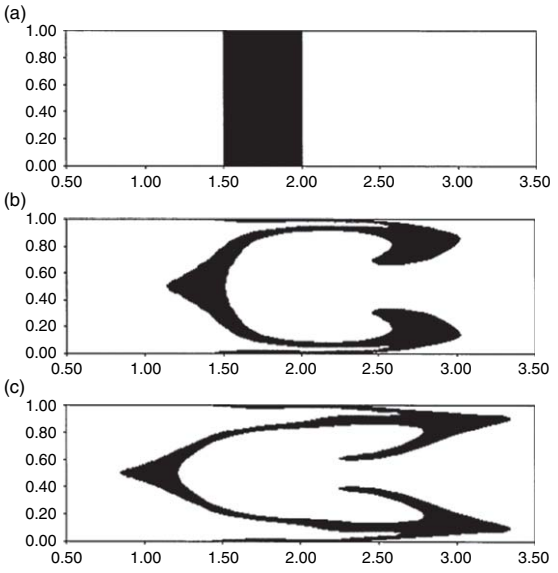


Figure 48 Evolution of a material blob at times (a) $t = 0$; (b) 600T; and (c) 1200T where T is the wave period. (Source: Yi M, Bau H H, Hu H H 2002 Peristaltically induced motion in a closed cavity with two vibrating walls. *Phys. Fluids* **14**, 184–97.)

because of the nonslip condition that holds there. The material in the vicinity of the membranes moves in the direction of the wave while (to conserve mass) the material in the center moves in the opposite direction. The stretching rate is nearly linear with time, which is much better than the rate of diffusion alone that scales with the square root of the time but less effective than the exponential stretching associated with chaotic advection. However, the stretching rate can be enhanced by oscillating both top and bottom walls at different frequencies and with different wave numbers.

2.10.3.6 Acoustic Mixer

When ultrasound is generated in a liquid, the latter would flow in the same direction as the ultrasound waves propagate (Wu and Du 1997). This phenomenon called acoustic streaming is well known for its transport properties: the fluid flows without any external mechanical contact. By appropriate use of the acoustic streaming in an adequate pattern, liquids in a closed chamber can be mixed with high efficiency (Suri *et al.* 2002a, b). **Figure 49** schematically depicts the ultrasonic mixer, which consists of a mixing chamber (closed bottle in **Figure 49**) containing the solution to be mixed and positioned vertically into an acoustic field in a water bath. A generator connected to an amplifier induces ultrasound waves using high-energy piezoceramic transducers. Two transducers are placed in the water bath outside the mixing chamber. One part of

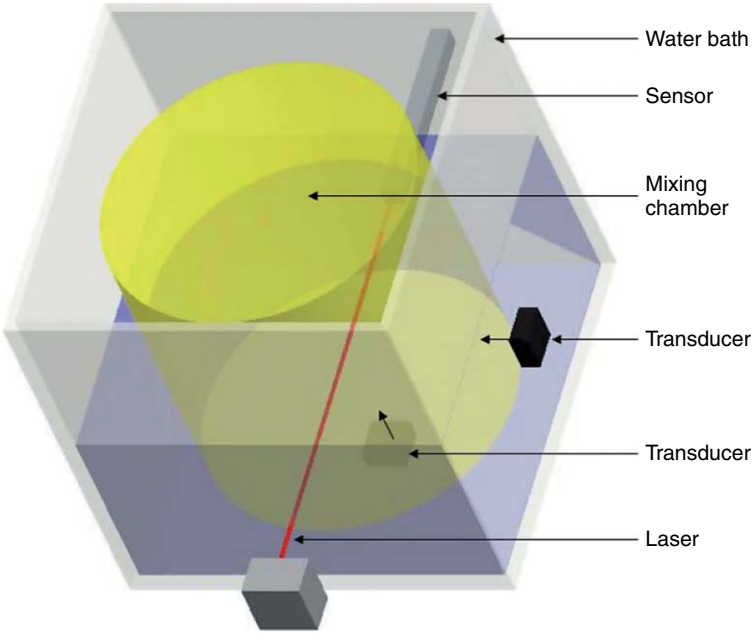


Figure 49 Schematic description of the ultrasonic mixer.

the emitted sound penetrates in the mixing chamber while another part is reflected and dispersed in the bath. The sound waves inside the chamber generate bubbles, the positions of which are changed, and the resulting streaming rotates the solution. To visualize the mixing, a laser beam traverses the glass cylinder and hits a photodiode connected to an oscilloscope. If transducer A is placed on one side of the chamber, for example, in the *x*-direction, the end walls of the mixing chamber serve to restrict the flow and thus a cellular flow pattern is induced. Similarly, when a second transducer B is aligned along a direction perpendicular to that of the transducer A, another cellular flow pattern is generated. By timewise alternations of two or more transducers positioned according to different directions, the resulting periodic acoustic streaming generates effectively chaotic mixing patterns. The main advantage of the mixer is that it allows the spinning of the solution inside a closed chamber without any mechanical contact. However, the temperature rise caused by acoustic energy may limit its applications for biological and chemical analysis.

In contrast, Katou *et al.* (2005a, b) developed a new ultrasonic mixer for mixing samples and reagents in an open-top vessel in which the mixture to be mixed has a free surface. **Figure 50** illustrates the concept of the proposed mixing method, which uses free surface waves in the vessel generated by ultrasound irradiation from outside the vessel. Both the vessel containing the liquid to be mixed and the sound source generating the ultrasound are placed in a water bath. The ultrasound irradiated from the sound source propagates in water, transmits through the vessel wall, and reaches the free surface of the liquid in the vessel. The free surface is tilted by the radiation pressure of ultrasound, which is generated at the boundary between the liquid and the air, and as a result, circulation flow arises and can be used to improve mixing. **Figure 51** depicts the experimental mixing results with 170 µl water and 3 µl dye when the sound source is periodically turned on and off. The mixture becomes homogeneous within about 0.5 s. The mixing performance can be significantly improved by using two ultrasound sources, one of them radiates through the bottom and the other through the sidewall of the vessel. Chaotic advection occurs by periodically actuating the two ultrasound sources (Katou *et al.* 2005b). In this new ultrasonic mixer, there is no significant temperature rise in the solution.

Alternatively, Liu *et al.* (2002, 2003) developed a novel ultrasonic mixer by exciting air bubbles in a liquid medium with a sound field, which is generated by a piezoelectric (lead zirconate titanate, PZT) disk attached to the mixing chamber. The mixer consists of a mixing chamber with a number of air pockets on the inner wall of the polycarbonate mixing chamber for trapping air bubbles, as shown in **Figure 52**. A PZT disk is glued onto the external surface of the chamber. When the PZT is actuated by a sinusoidal voltage, acoustic streaming is induced around the air bubbles. As a result, a churning motion is induced near the individual bubbles in the liquid. **Figure 53** shows the experimental mixing results in a 12 mm $\times$ 15 mm $\times$ 0.125 mm mixing chamber when the PZT is actuated at a frequency of 5 kHz with a

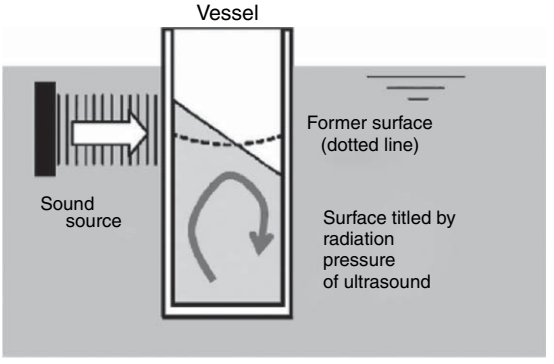


Figure 50 Response of a free surface wave in a vessel to ultrasound. (Source: Katou H, Miyake R, Kambara K, Kawase K, Uchida H 2005 Development of a non-contact micro-liquid mixing method and application to chemical auto-analyzers. *Chem. Eng. Sci.* **60**, 5519–28.)

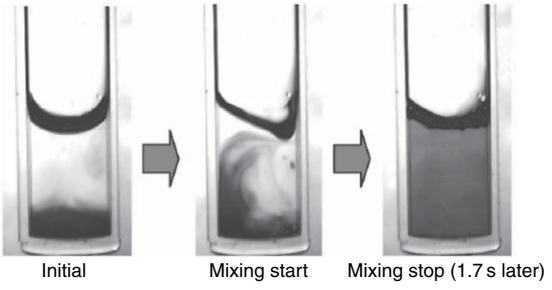


Figure 51 Experimental mixing result of the ultrasonic mixer in which circulation flow is induced by free surface waves in a vessel as generated by ultrasound irradiation from outside the vessel. (Source: Katou H, Miyake R, Kambara K, Kawase K, Uchida H 2005 Development of a non-contact micro-liquid mixing method and application to chemical auto-analyzers. *Chem. Eng. Sci.* **60**, 5519–28.)

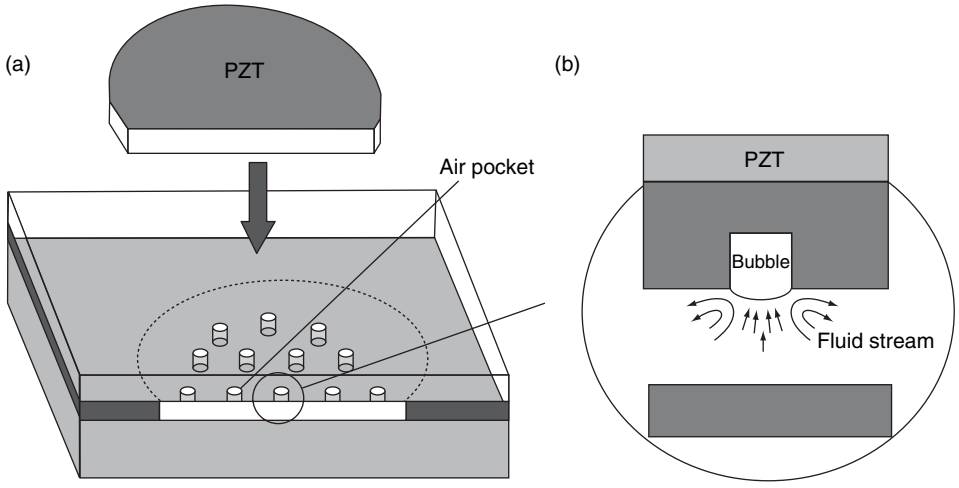


Figure 52 Schematic of the bubble-induced acoustic mixer containing a number of air pockets in the top layer of the mixing chamber. (a) Overview and (b) side view. (Source: Liu R H, Yang J N, Pindera M Z, Athavale M, Grodzinski P 2002 Bubble-induced acoustic micromixing. *Lab Chip* **2**, 151–7.)

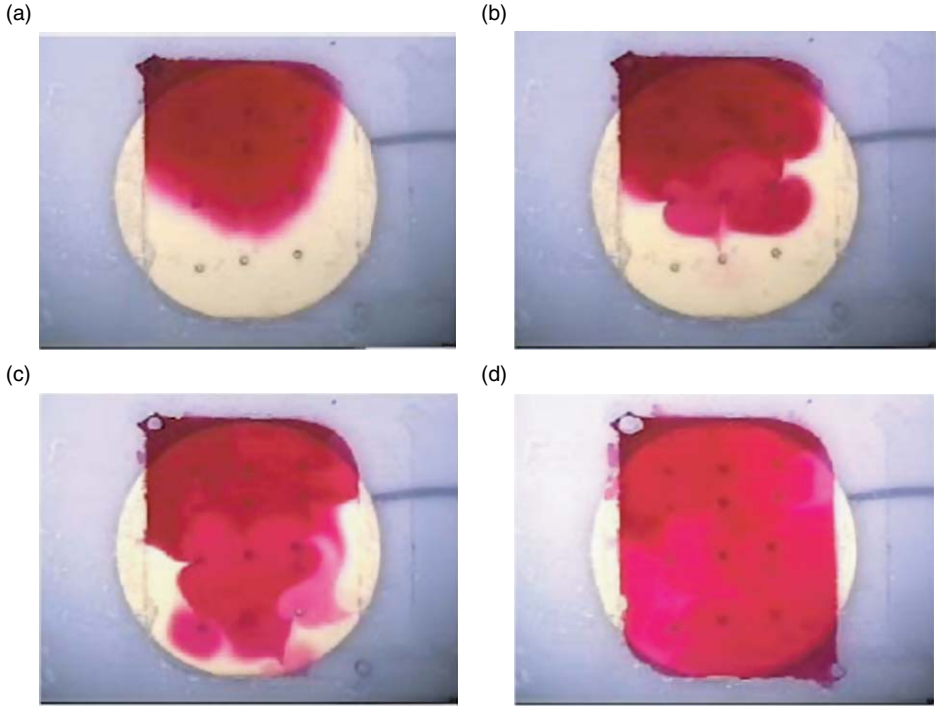


Figure 53 Experimental mixing results of the bubble-induced acoustic mixer at times (a) 0 s; (b) 28 s; (c) 67 s; and (d) 106 s. (Source: Liu R H, Yang J N, Pindera M Z, Athavale M, Grodzinski P 2002 Bubble-induced acoustic micromixing. *Lab Chip* **2**, 151–7.)

peak-to-peak amplitude V_{pp} of 10 V. A complete mixing is achieved across the whole chamber within 105 s, while the mixing based on diffusion alone takes about 6 h for the same chamber. The experimental results also demonstrate that a square sound wave provides faster mixing than does a sinusoidal sound wave, and a higher voltage amplitude results in faster mixing. Due to low frequency (~5 kHz) and low voltage (~10 V), the power consumption of the mixer is low (~2 mW) and the transducer therefore

adds almost no heat to the liquid. Such a mixer has been used in various biomolecular analyses such as immunomagnetic cell capture, bacterial viability assay, and DNA hybridization.

In the previous ultrasonic mixers, no through flow is present, and the only fluid motion in the chamber is due to the agitation induced by the acoustic streaming. However, in certain circumstances, it may be desirable to stir the fluids while it is pumped through the conduit. Such net fluid motion can be driven by an external pressure, i.e., the fluid is pumped continuously with a syringe pump. Yaralioglu *et al.* (2004) utilized acoustic streaming to disturb the pressure-driven main flow in a basic Y-mixer. **Figure 54** schematically depicts the ultrasonic mixer in a Y-shaped microfluidic channel using integrated PZT transducers. Two fluid streams are brought together through the Y-junction and continuously flow through the mixing channel. Ultrasound is introduced into the mixing channel by the embedded PZT transducers made of a thin film of zinc oxide. The thin film of ZnO is sandwiched between two electrodes, and it is deposited on the bottom surface of a quartz substrate. The PDMS channel is aligned along the transducers on the top surface of the substrate. The device is glued on a piece of printed circuit board, which has soldered high-frequency connectors. The electrical connection is achieved by means of a piece of 2-mil gold wire between the connectors and the transducer electrodes using silver ink. The transducers are evenly distributed along the mixing channel on the other side of the quartz wafer. One can fabricate transducers with various electrode patterns as shown in **Figure 54(b)**. **Figure 55** depicts the experimental mixing results obtained by pumping phenolphthalein and NaOH solutions into the mixing channel from the two inlet channels at the same

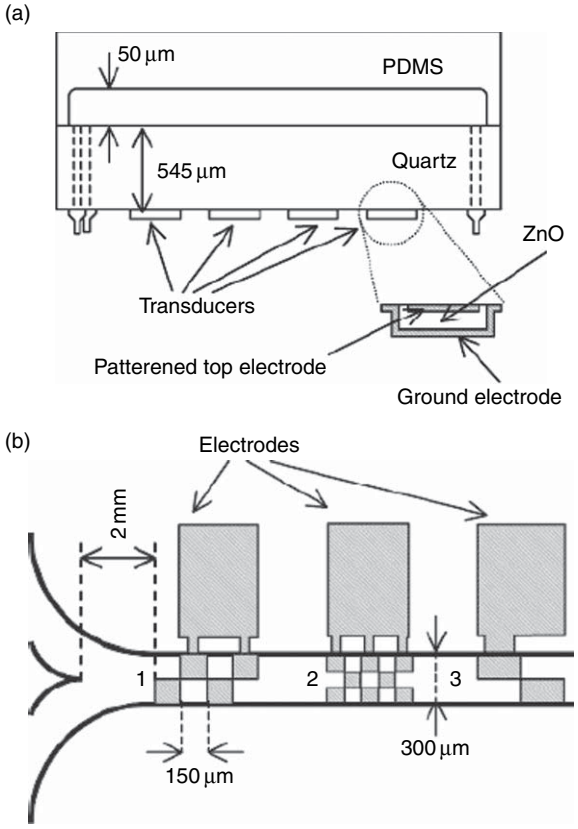


Figure 54 Schematic of the ultrasonic mixer that contains a basic Y-mixer and integrated transducers to induce secondary flows perpendicular to the main flow direction. (a) Cross section of the microfluidic channel and (b) electrode patterns of the first three transducers. (Source: Yaralioglu G G, Wygant I O, Marentis T C, Khuri-Yakub B T 2004 Ultrasonic mixing in microfluidic channels using integrated transducers. *Anal. Chem.* **76**, 3694–8.)

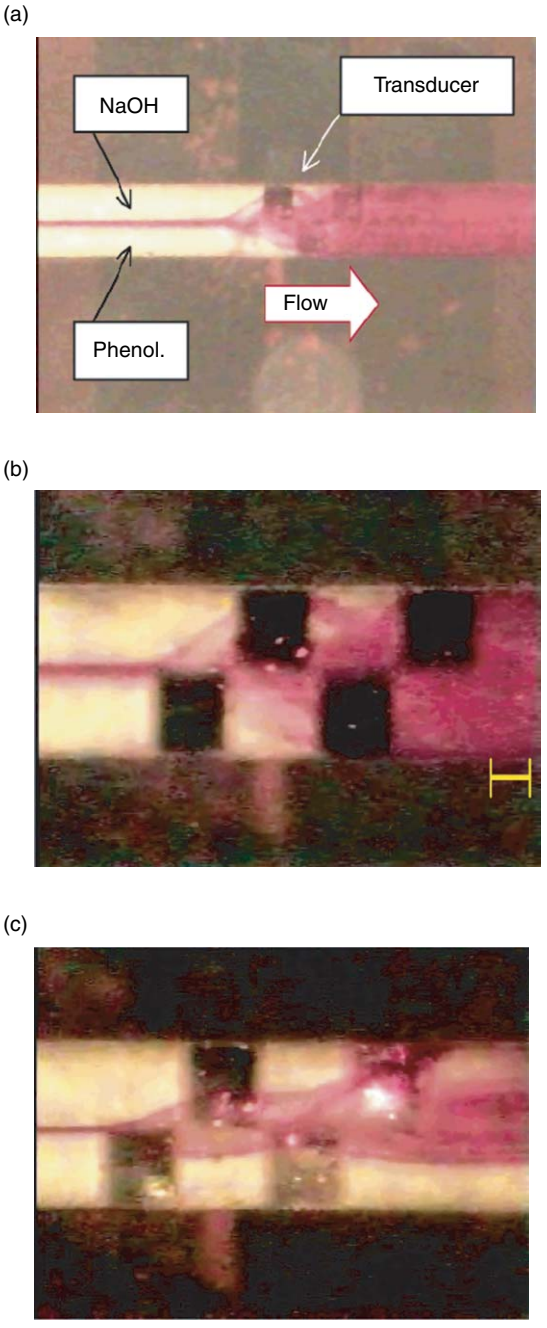


Figure 55 Experimental mixing results of the ultrasonic mixer in which two fluid streams are continuously pumped through the mixer and the main flow is disturbed by a secondary flow induced by the acoustic streaming of the liquid above the transducers. (a) The flow rate in the mixing channel is $10\,\mu l\,min^{-1}$; (b) in the vicinity of the transducers at flow rate $10\,\mu l\,min^{-1}$; and (c) in the vicinity of the transducers at flow rate $60\,\mu l\,min^{-1}$. (Source: Yaralioglu G G, Wygant I O, Marentis T C, Khuri-Yakub B T 2004 Ultrasonic mixing in microfluidic channels using integrated transducers. *Anal. Chem.* **76**, 3694–8.)

flow rates. The colorless solutions generate a pink color when mixed. By observing the intensity of the color, one can estimate the mixed amounts of the two solutions. When the total flow rate in the mixing channel is $10\,\mu l\,min^{-1}$ (the associated Reynolds number is 0.86), the two solutions remain well separated in the mixing channel before passing the transducer, and the diffused chemicals generate only a pink line at the interface of the two fluids. The yellow color of the solution well before the transducer is due to the gold electrode covering the bottom surface of the substrate where the transducers are fabricated. **Figure 55(b)** depicts the zoomed image of **Figure 55(a)** in the vicinity of the transducer, which is driven by a sinusoidal voltage of $V_{pp} = 1.2\,V$ at 450 MHz, and the pink color almost covers the entire channel, which indicates that complete mixing of the chemicals is achieved just following the transducer. **Figure 55(c)** depicts the mixing result when the total flow rate in the mixing channel is $60\,\mu l\,min^{-1}$. As the flow rate increases, color uniformity across the channel degrades. Therefore, for a higher flow rate, one has to increase the power level applied to the transducers to get a uniform mixing across the channel. The main mixing mechanism is the transverse secondary flow generated by the acoustic streaming of the liquid above the transducers. The stream pushes the fluid toward the top of the channel and a circulating flow perpendicular to the main flow is induced.

2.10.3.7 Magnetohydrodynamic Mixer

The application of electromagnetic forces to the pump, confine, and control fluids and is by no means new. MHD is, however, mostly considered in the context of highly conducting fluids such as liquid metals and ionized gases (Davidson 2001, Woodson and Melcher 1969).

Consider a conduit with two electrodes deposited along its opposing walls. The conduit is filled with a mixture of a weakly conductive solution. When a potential difference is applied across the two opposing electrodes, a current density $\mathcal{J}$ transmits through the solution. The entire device is subjected to a magnetic field of intensity B (**Figure 56**). The interaction between the electric current density $\mathbf{J}$ and the magnetic field $\mathbf{B}$ generates a Lorentz force $\mathbf{J} \times \mathbf{B}$, which is directed along the axis of the conduit (in the z-direction that is perpendicular to the cross section) and drives the fluid motion. Witness that the Lorentz force plays a similar role to that of a

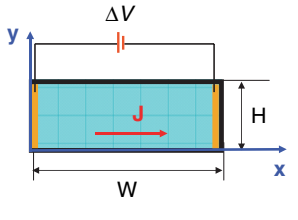


Figure 56 A cross section of a conduit filled with an electrolyte solution and exposed to a magnetic field of intensity *B*. Two electrodes are deposited along the opposing walls and are subjected to a potential difference ΔV, and a current density **J** transmits through the solution. The interaction between the current density **J** and the magnetic field **B** generates a Lorentz force $\mathbf{J} \times \mathbf{B}$ that induces fluid motion.

pressure gradient, and the induced flow motion is similar to a pressure-driven flow.

Since the direction of the current and thus the direction of the force can be controlled by appropriate design of the electrodes, a number of researchers have recently constructed MHD micropumps on silicon and ceramic substrates and demonstrated that these pumps are able to move liquids around in small conduits (Bau 2001, Jang and Lee 2000, Lemoff and Lee 2000, 2003, Zhong *et al.* 2002). Bau *et al.* (2001, 2002, 2003) also showed the feasibility of using MHD forces to control fluid flow in microfluidic networks. By judicious application of various potential differences to different electrode pairs, one can direct the liquid to flow along any desired path without the need of valves and pumps. **Figure 57** depicts a simple example of an MHD microfluidic network fabricated with low-temperature cofired ceramic tapes (Bau *et al.* 2003). The MHD network contains two reservoirs, one stores reagent A and the other reagent B. The solutions are pumped out of the two reservoirs, and flow through the conduit 1 which doubles as a stirrer. Once the reagents are stirred, they are pumped into the torus (conduits 2–3). The electrodes are then reprogrammed to circulate the fluid around the torus (2–3) for a desired number of times. Various regions of the torus can be maintained at different temperatures, allowing thermal cycling and potentially PCR. Once the DNA amplification is completed, the electrodes are reprogrammed and the fluid is pumped back either to conduit 1 or to conduit 4 for further processing. By changing the polarity of the electrodes, one can readily modulate the direction of the flow. Subsequently, the fluid is split into two separate paths (conduits 5 and 6) for parallel analysis. Finally, the waste is recombined at the conduit 7.

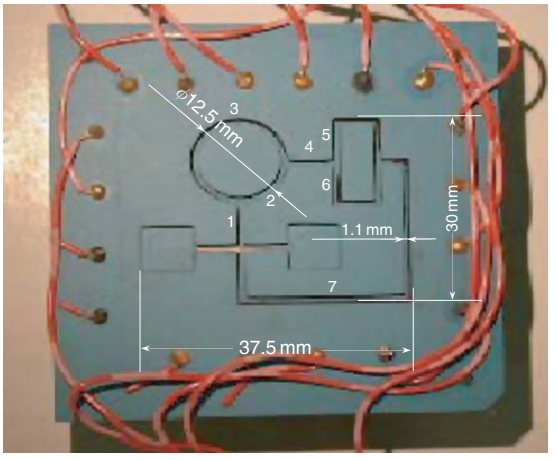


Figure 57 A prototype of a magnetohydrodynamic (MHD) microfluidic network. (Source: Bau H H, Zhu J, Qian S, Xiang Y 2003 A magneto-hydrodynamically controlled fluidic network. *Sens. Actuators B* **88**, 205–16.)

2.10.3.7.1 Annular MHD stirrer

Other applications of microfluidic MHD components and devices are MHD-based stirrers. Gleeson (2005) and Gleeson *et al.* (2004) proposed an annular MHD stirrer. Two electrodes are positioned along the inner and the outer walls of a toroidal conduit to form opposing electrodes, and the device is positioned in an external magnetic field. The direction of the flow is reversed periodically by switching the polarity of the two opposing electrodes. Such a stirrer takes advantage of Taylor dispersion (Taylor 1953) to increase the surface area between the two interacting fluids. **Figure 58** depicts the operation of an idealized MHD annular micromixer, which stretches the fluid interface at each revolution of the fluid around the annulus. This leads to a linear growth of the total interface length with time, which is much better than the rate of diffusion alone that is proportional to the square root of time but less effective than the exponential stretching associated with the MHD chaotic stirrers as developed by Bau and coworkers (Qian and Bau 2005a, b, Qian *et al.* 2002, Xiang and Bau 2003, Yi *et al.* 2002b).

2.10.3.7.2 MHD chaotic stirrer I: MHD stirrer in a chamber

Alternatively, one can pattern electrodes of various shapes, which in turn induce electric fields in different directions. The interaction of such electric fields with the magnetic field induces secondary flows that may

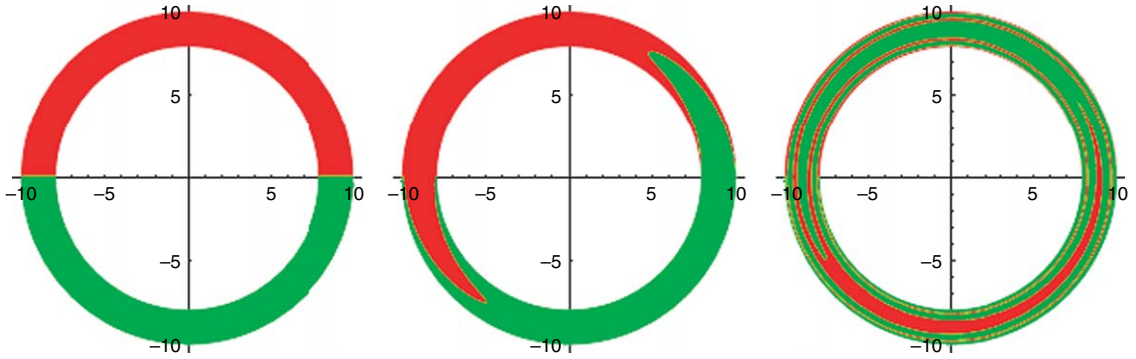


Figure 58 Operation of an idealized magnetohydrodynamic (MHD) annular mixer at three various times. (Source: Gleeson J P, Roche O M, West J, Gelb A 2004 Modelling annular micromixers. *SIAM J. Appl. Math.* **64**, 1294–310.)

benefit stirring and mixing. Although these secondary flows significantly enhance the mixing process, they are well ordered and the mixing remains rather poor. The performance can however be significantly improved. Indeed, by periodically or aperiodically alternating two or more different flow patterns, one can induce (Lagrangian) chaotic advection. Aref (1984) described the general ideas associated with chaotic mixing, and similar ideas were implemented in the context of microfluidic systems and MHD stirrers (Qian and Bau 2002, 2005a, b, Qian *et al.* 2002, Xiang and Bau 2003, Yi *et al.* 2002b).

Yi, Qian, and Bau (hereafter referred to as YQB) described a MHD stirrer consisting of a closed cylindrical cavity with an electrode (denoted C) deposited around its periphery and two additional electrodes A and B placed eccentrically inside the cavity (Yi *et al.* 2002b). **Figure 59** schematically depicts the MHD stirrer in a closed cavity. A magnetic field parallel to the cavity axis is externally applied. When a potential difference is applied across the electrode pair A–C, a circulatory flow pattern results with its center of rotation near the electrode A. **Figure 60(a)** depicts the computed streamlines and **Figure 60(b)** depicts the experimental passive tracer's trajectories when the electrode pair A–C is engaged. When a potential difference is applied across the electrode pair B–C, a similar flow pattern results with its center of rotation next to the electrode B.

When the electric potential differences are applied alternately across the electrode pairs A–C and B–C with a period T, the two different flow patterns are periodically switched on and switched off. At small periods T, the flow is regular and periodic in most of the cavity. At relatively large periods, the passive tracer experiences global chaotic advection. **Figure 61** depicts the Poincaré sections (*left column*) and flow visualizations (*right column*) at various periods. As the period T increases, the complexity of the flow increases.

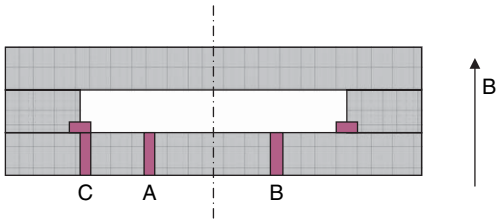


Figure 59 A schematic representation of a magnetohydrodynamic (MHD) stirrer in a cylindrical cavity. The plum segments denote the electrodes. Electrodes A and B are eccentrically located at the bottom of the cavity. Electrode C is shaped like a disk and surrounds the cavity. The cavity is filled with a weakly conductive solution and is positioned in a permanent magnet with magnetic field **B** parallel to the axis of the mixing chamber. (Source: Yi M, Qian S, Bau H H 2002 A magnetohydrodynamic chaotic stirrer. *J. Fluid Mech.* **468**, 153–77.)

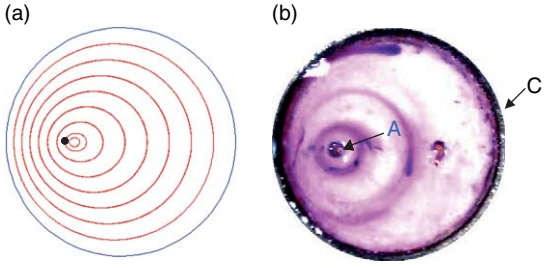


Figure 60 The flow field induced by an electrode pair A–C. (a) Streamlines and (b) a photograph of a flow visualization experiment. (Source: Yi M, Qian S, Bau H H 2002 A magnetohydrodynamic chaotic stirrer. *J. Fluid Mech.* **468**, 153–77.)

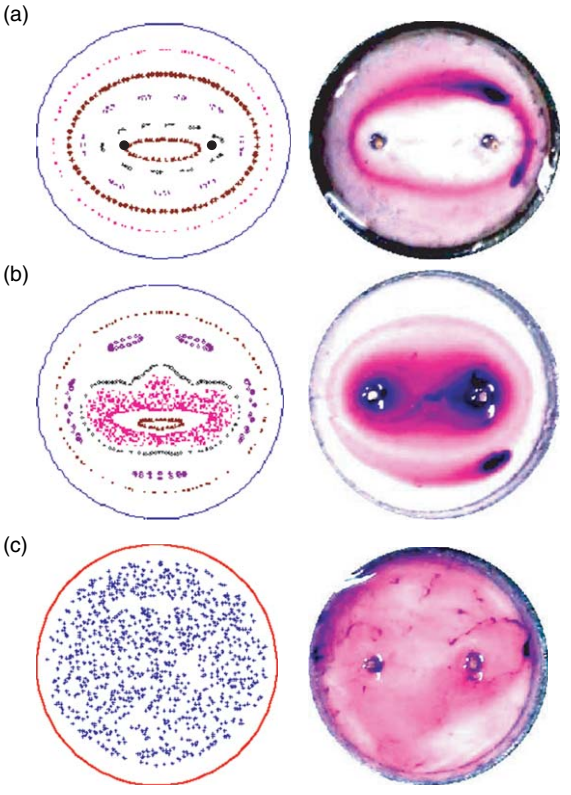


Figure 61 Poincaré sections (stroboscopic images, *left column*) and flow visualization photographs (*right column*) when the electric potential differences are applied alternatively across the electrode pairs A–C and B–C with a period T. (a) $T=4$; (b) $T=10$; and (c) $T=40$. (Source: Yi M, Qian S, Bau H H 2002 A magnetohydrodynamic chaotic stirrer. *J. Fluid Mech.* **468**, 153–77.)

2.10.3.7.3 MHD chaotic stirrer II: MHD stirrer in an MHD-controlled network

Later, Qian *et al.* (2002) developed a variant of YQB's stirrer that is better suited for integration into the MHD network shown in **Figure 57**. Only slight modifications are needed in the network's branches so that they can serve the double function of a pump and a stirrer. The basic building block (branch) of such a network is depicted in **Figure 62**. The branch consists of a conduit with two electrodes (denoted as C^+ and C^-) deposited along its two opposing walls. The conduit is filled with an electrolyte solution. Many conduits of the type depicted in **Figure 62** can be connected to form a network, and one example is shown in **Figure 57**. The entire device is subjected to a uniform magnetic field in the z-direction. When a potential difference is applied across the electrodes C^+ and C^-, the resulting current interacts with the magnetic field to form body

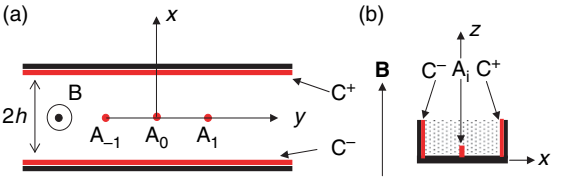


Figure 62 A liquid-filled conduit is confined between two parallel plates ($x=\pm h$). A uniform magnetic field, **B**, is parallel to the z-coordinate. Electrodes C^+ and C^- are deposited along the confining plates ($x=\pm h$) and point electrodes A_{-1}, A_0, and A_1 are positioned along $x=0$ with a separation distance denoted c. When the device operates as a stirrer, the electrodes C^+ and C^- are wired to form a single electrode C. (a) Top view and (b) side view. (Source: Qian S, Zhu J, Bau H H 2002 A stirrer for magnetohydrodynamically controlled minute fluidic networks. *Phys. Fluids* **14**, 3584–92.)

(Lorentz) forces that propel the fluid. By judicious application of different potential differences to different electrode pairs, one can direct the liquid to flow along any desired path. In other words, MHD allows us to control fluid flow in the microfluidic network in very much the same way as one controls electric current flow in an electronic circuit.

Additional point electrodes (denoted as A_i in **Figure 62**) are deposited along the centerline of the conduit's bottom. When the network branch operates as a stirrer, both electrodes C^+ and C^- are connected to the same terminal of a power supply so that they form a single electrode C. When a potential difference is applied across any of the central electrodes A_i and the electrode C, circulatory motion ensues with the liquid rotating around the engaged point electrode A_i.

When two or more electrode pairs are alternatively actuated, complex, chaotic flows are induced. **Figure 63(a)–63(c)** depicts the stroboscopic images of a passive tracer (I) and flow observations (II) when two agitators A_{-1}–C and A_1–C induce corotating motions and the alternation periods are $T=4$, 8, and 20, respectively. Each stroboscopic image is a superposition of snapshots that depict the location of the tracer particle at the end of each period, and it is referred to as Poincaré sections, which are often used as a diagnostic tool to determine the effectiveness of the stirring process. When the points that emerge lie on a smooth curve, the motion is deemed to be nearly regular (poor stirring). A scattered pattern corresponds to chaotic motion. When the period (T) is relatively small (**Figure 63(aI)**, only narrow regions of irregularity are observed. The chaotic region is a result of the homoclinic tangle next to

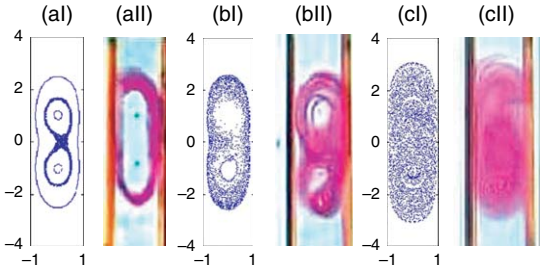


Figure 63 (I) Stroboscopic images and (II) flow visualization photos when two (corotating) electrode pairs are engaged. (a) $T = 4$; (b) $T = 8$; (c) $T = 20$. (Source: Qian S, Zhu J, Bau H H 2002 A stirrer for magnetohydrodynamically controlled minute fluidic networks. *Phys. Fluids* **14**, 3584–92.)

the hyperbolic point. Both the theoretical and the experimental data illustrate that at relatively small periods, the flow has a structure similar to the superposed flow topology observed in the limiting case $T \to 0$. As T increases (**Figure 63(b)**), the size of the chaotic region and the spread of the dye increase. When $T = 20$ (**Figure 63(c)**), the chaotic region has nearly occupied the entire region $-3 < y < 3$. Careful examination, however, reveals the persistence of small regular regions.

Another way to examine the stirring process is to track the history of a material blob inserted into the liquid. **Figure 64** depicts the evolution of a square material blob of edge size 0.1, centered at (0, 0), as a function of time by integrating the trajectories of 10^4 points initially evenly distributed within the square. The figure depicts the blob's evolution at times $t = 2T$ (**Figure 64(a)**), $3T$ (**Figure 64(b)**), $4T$ (**Figure 64(c)**), $6T$ (**Figure 64(d)**), and $20T$ (**Figure 64(e)**), and compares the computational results with experimental observations. The computational results qualitatively agree with the experimental flow visualizations. After about 6 periods, the material blob has covered most of the conduit.

This arrangement is similar to Aref's (1984) blinking vortex. However, in contrast to Aref's highly idealized, inviscid system, the stirrer developed here can be readily constructed, which is yet another example of Stokes flow that exhibits chaotic advection.

2.10.3.7.4 MHD chaotic stirrer III: MHD stirrer without internal electrodes

All the MHD stirrers described above require some of the electrodes to be patterned inside the conduit or chamber and away from the walls of the conduit/

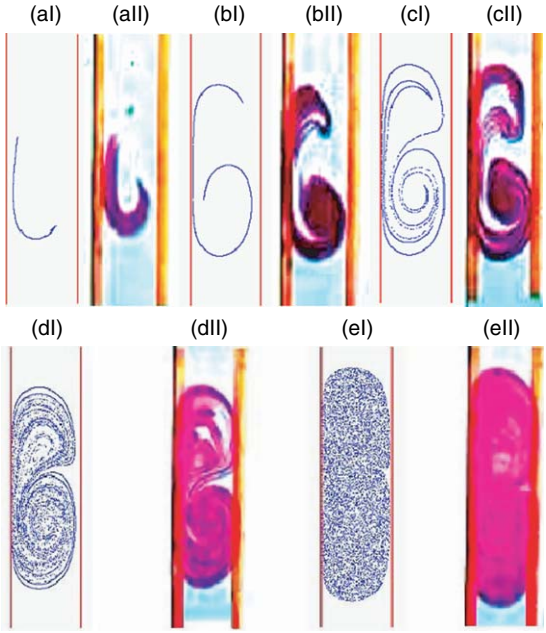


Figure 64 Deformation of a material blob alternatively excited by two electrode pairs. Counterrotating, $c = 1$ and $T = 20$. (I) and (II) represent the computational results and the visualization photos of the experiment, respectively. (a) $t = 2T$; (b) $t = 3T$; (c) $t = 4T$; (d) $t = 6T$; (e) $t = 20T$. (Source: Qian S, Zhu J, Bau H H 2002 A stirrer for magnetohydrodynamically controlled minute fluidic networks. *Phys. Fluids* **14**, 3584–92.)

chamber. In some cases, such internal electrodes may be intrusive. To alleviate this potential shortcoming, a new stirrer design that does not require any interior electrodes was developed (Qian and Bau 2005a). The same electrodes used for pumping are used for stirring. This arrangement requires fewer fabrication steps than were required in the previous designs, and it also minimizes the intrusion that arises when introducing internal electrodes. The newly designed MHD stirrer can operate as a stand-alone component or it can be incorporated into a MHD-controlled network.

The stirrer consists of a conduit equipped with individually controlled electrodes that are positioned along the opposing walls of the conduit, and there is a small dielectric gap between the adjacent electrodes. **Figure 65** depicts a photograph of the device. The conduit is filled with a weakly conductive electrolyte solution, and the device is placed in a static magnetic field of flux density **B**.

When the device serves as a pump, the electrodes along one side of the conduit's walls are wired to form a single electrode C^+, which is connected

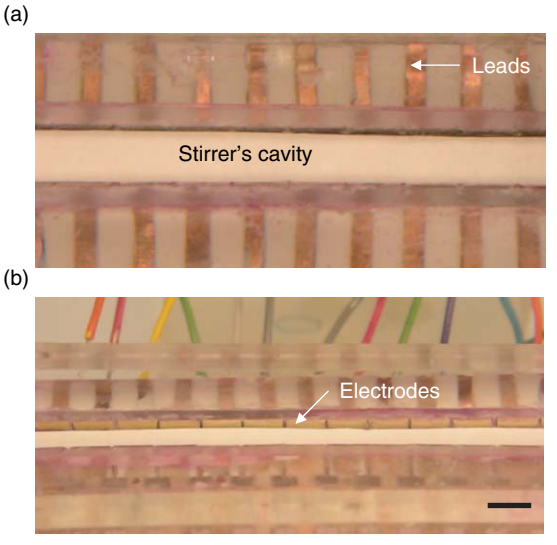


Figure 65 A photograph of the magnetohydrodynamic (MHD) stirrer without internal electrodes inside the microfluidic channel. (a) Top view showing the leads of the electrodes. (b) Oblique view showing the individual electrodes positioned along the wall of the stirrer's conduit. Scale bar, 2 mm.

to one terminal of the power supply. The electrodes along the other side of the conduit's walls are similarly wired to form a single electrode C^-. When a potential difference ΔV is applied between the two groups of electrodes C^+ and C^-, the current direction is nearly normal to the surface of the electrodes, the Lorentz force is directed along the conduit's axis, and the device operates as a pump.

When only two of the electrodes that are diagonally positioned along the opposing walls are activated (i.e., electrode pair $C_0^- - C_1^+$ as depicted in **Figure 66**), the direction of the current flow is oblique to the electrodes' surfaces and the resulting Lorentz force has a component transverse to the conduit's axis. As a result, one observes cellular flow to enhance mixing. When a potential difference is applied across another diagonal electrode pair such as $C_0^- - C_{-1}^+$, a nearly mirror image of the flow pattern forms. By periodically or aperiodically alternating the two flow patterns, one can obtain more complicated trajectories of the passive tracer particles and can achieve chaotic advection that elongates the length of the interface at an exponential rate. The stirring process is described more vividly in **Figure 66**. Red and green dye blobs were introduced into the conduit, and their evolution was tracked as a function of time. At $t = 0$, the red and the green dyes are well separated (**Figure 66**, $t = 0$). After one period, some of the red dye is surrounded by the green dye (**Figure 66**, $t = T$). As the time increases, through continuous stretching and folding, which are

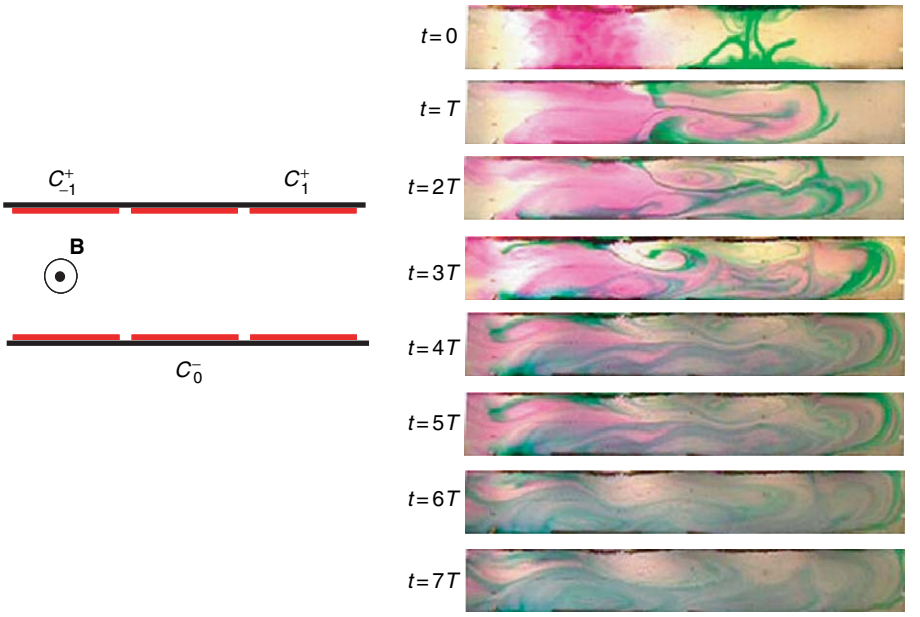


Figure 66 Images of the blending process when initially well-separated red and green dye blobs are introduced into the stirrer. The various images correspond to time $t = 0$, T, $2T$, $3T$, $4T$, $5T$, $6T$, and $7T$. Two diagonally positioned electrode pairs $C_0^- - C_1^+$ and $C_0^- - C_{-1}^+$ are alternatively active with a potential difference $\Delta V = 2.5$ V and a period $T = 4$ s. (Source: Qian S, Bau H H 2005 Magnetohydrodynamic stirrer for stationary and moving fluids. *Sens. Actuators B* **106**, 859–70.)

characteristics of chaotic advection, the two dyes blend in the mixing region. Here only three electrodes, C_0^-, C_{-1}^+, and C_1^+, are engaged, and one can increase the volume of the mixing region by increasing the number of engaged electrodes.

So far, we have described the operation of the stirrer when the only fluid motion in the conduit was due to the agitation induced by the MHD stirrer. In other words, no through flow was present. In certain circumstances, it may be desirable to stir the fluids while it is pumped through the conduit. Such net fluid motion can be driven either by an external pressure or by an MHD drive located some distance away from the stirring region. **Figure 67(a)** schematically depicts the new design in which both fluid propulsion and mixing are achieved by MHD. The device consists of a microfluidic channel with pumping electrodes C^+ and C^- located at both ends and deposited along the opposing walls of the channel. When a potential difference is applied across the electrodes C^+ and C^-, the fluid is pumped from one end of the device to the other. Between the pumping electrodes, there is a mixing region where several individually controlled mixing electrodes denoted as TO, TE, BO, and BE are positioned along the opposing walls with a small dielectric gap between the adjacent electrodes. In the mixing region, a transverse velocity component is created when a potential difference is applied across electrodes that are diagonally positioned. **Figure 67(b)** depicts the flow field and trajectories of passive particles in the mixing region when a potential difference is applied across electrode pair TO–BE. In the upstream of the mixing region, the velocity profile is parabolic. In the mixing region, the main flow is perturbed by the induced Lorentz forces and zigzag trajectories of passive particles result. As before, a potential difference is alternatively applied across electrode pairs TO–BE and TE–BO with a period T and a chaotic advection is induced in the mixing region. The fluids are mixed after they are pumped through the mixing region.

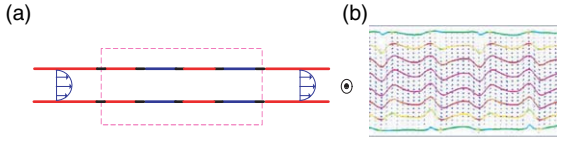


Figure 67 (a) Schematic representation of magnetohydrodynamic (MHD) pumping and mixing in a microfluidic channel. Pumping electrodes C^+ and C^- are located at both ends of the channel and are deposited along the opposing walls, and several individually controlled mixing electrodes denoted as TO, TE, BO, and BE are positioned along the opposing walls with a small dielectric gap between adjacent electrodes. The device is positioned in a magnetic field **B**. (b) Flow field and trajectories of passive particles in the mixing region when a potential difference is applied to the diagonally positioned electrode pair TO–BE.

2.10.3.8 Electroosmotic Mixer

Electroosmosis is very attractive for manipulating fluids in microdevices. It can be readily seen from eqn [1] that the zeta potential is an important factor that affects the EOFs. Zeta potential is related to, among other things, the surface chemistry or surface charge density of the channel, eventually the pH value of the solution, and the electrolyte concentration. A plug flow is generated if the surfaces are homogeneously charged, but heterogeneous charge surfaces are often unavoidable because of surface defects. In some situations, an inhomogeneous surface charge distribution is designed for generating desirable flow structures. For example, 3D recirculating EOFs are generated by patterning materials of different zeta potentials inside a channel, as discussed in Section 2.10.2.5.

To control the EOF, the zeta potential of a channel wall can be dynamically modulated as a result of changes in the potential applied to gate electrodes embedded beneath the solid–liquid interface (Lee *et al.* 1990, Lin *et al.* 2005, Schasfoort *et al.* 1999, Wu and Liu 2005). **Figure 68** schematically depicts the control of EOF in a channel with one embedded electrode called the gate electrode. Witness that the gate electrode does not contact the liquid. The zeta potential can be estimated from the following expression (Wu and Liu 2005):

$$\zeta = \sinh^{-1}\left(\frac{\kappa(\sigma_v - \sigma_s)}{4n_0 ze}\right)\exp(-\kappa d) \qquad [4]$$

In the above, $\kappa = (2n_0 z^2 e^2 / k_B T \varepsilon_0 \varepsilon_r)^{1/2}$ is the reciprocal screening Debye length, n_0 is the bulk ionic concentration, z is the valance of the symmetric ions, e is the elementary charge, k_B is the Boltzmann constant, T is the absolute temperature of the liquid solution, d is the distance from the outer Helmholtz plane to the shear plane, $\varepsilon_0\varepsilon_r$ is the dielectric permittivity of the solution, $\sigma_v = \varepsilon_0$

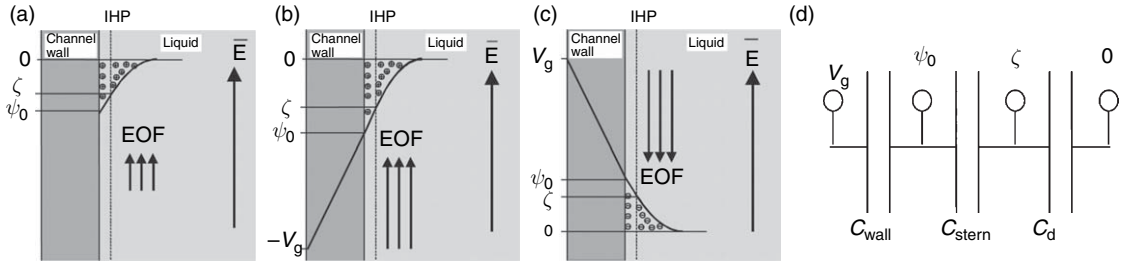


Figure 68 Schematic representation of controlling electroosmotic flow (EOF) through the modulation of the gate potential applied to the gate electrode embedded beneath the solid–liquid interface. (a) Illustration of the development of the zeta potential and EOF at the solid–liquid interface; (b) enhance EOF; (c) reverse EOF; (d) the three capacitors in series that are used to describe the effect of V_g on zeta potential, ζ. ψ_0 represents the surface potential of the insulator surface. (Source: Schasfoort R B M, Schlautmann S, Hendrikse J, Van den Berg A 1999 Field-effect flow control for microfabricated fluidic networks. *Science* **286**, 942–5.)

$\varepsilon_w(V/W)$ is the induced surface charge density due to the potential applied to the gate electrode, $\sigma_s = -(\gamma e/(1 + [H^+]/K_a))$ is the surface charge density at the solid–liquid interface, $\varepsilon_0\varepsilon_w$ is the dielectric permittivity of the channel wall, W is the thickness of the channel wall, V is the applied gate potential, γ is the sum of both the ionized and the protonated surface silanol group concentrations when the channel wall is made of SiO_2-based material, $[H^+]$ is bulk buffer hydrogen ionic concentration, and finally K_a is the acid dissociation constant of silanol groups.

From eqn [4], the modulation of the zeta potential through the control of the gate potential, V, strongly depends on the pH value and the concentration of the solution. The numerical simulations demonstrated that the lower the pH value and the lower the electrolyte concentration, the more sensitive the modulation of the zeta potential via the applied gate potential (Wu and Liu 2005). When there is no gate potential applied to the gate electrodes and the channel wall is negatively charged, the positive ions are attracted toward the charged surface, and the liquid moves in the same direction as that of the driving electric field (**Figure 68(a)**). One can increase the magnitude of the zeta potential, thereby enhancing the flow, by applying a negative gate potential to the embedded gate electrode (**Figure 68(b)**). One can also reverse the polarity of the zeta potential by applying a positive gate potential to the gate electrode, as shown in **Figure 68(c)**, which leads to a reversed flow.

One can pattern several individually controlled gate electrodes embedded beneath the solid–liquid interface, and complex flow field for enhancing the mixing can be generated by spatial and temporal modulation of the zeta potentials along the channel walls (Qian and Bau 2002, 2005b). **Figure 69** depicts two examples of flow patterns in a closed cavity with four electrodes (red segments) embedded along the top and bottom channel walls. The driving electrodes, i.e., the anode and the cathode, are located at the left and the right end, respectively, of the cavity. By controlling the gate potentials applied to the gate electrodes, the top and bottom channel walls are controlled to be negatively and positively charged, respectively. The liquid in the EDL next to the top channel wall moves in the same direction as that of the electric field, and the liquid in the EDL next to the bottom channel wall moves in the direction opposite to that of the applied electric field. The arrows in **Figure 69** depict the directions of the EOFs at the edge of the EDL. The left and right walls serve to restrict the flow and thus a clockwise

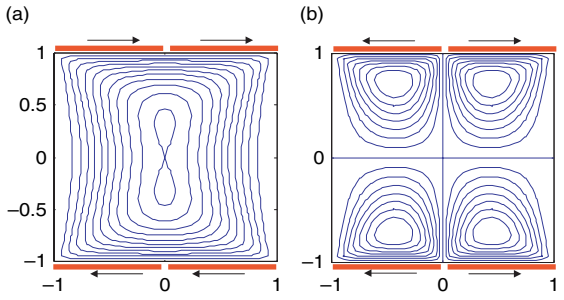


Figure 69 Streamline patterns for electroosmotic flows (EOFs) within a cavity with nonuniform zeta potentials at the top and bottom surfaces. The red lines represent the embedded gate electrodes, and the arrows denote the directions of the electroosmotic velocities at the edge of the electric double layer.

cellular flow pattern occurs (**Figure 69(a)**). By controlling the gate potentials applied to the gate electrodes located at the left half of the upper wall and the right half of the bottom wall, the polarities of the zeta potentials at the left half of the upper wall and the right half of the bottom wall are reversed and four counterrotating vortices occur within the cavity (**Figure 69(b)**).

We choose two basic flow patterns, say, A (**Figure 69(a)**) and B (**Figure 69(b)**), and maintain the flow field type A for a time interval $0 < t < T/2$, switch to the flow field type B for the time interval $T/2 < t < T$, and then switch back to the flow A. The process is repeated with a period T. The switching from one flow pattern to the other is accomplished by controlling the zeta potential distribution with the aid of the embedded electrodes. It is easy to imagine the trajectories of a passive (nondiffusing) tracer particle. In the time interval $0 < t < T/2$, the particle will follow the streamlines that correspond to the flow pattern A. In the time interval $T/2 < t < T$, the particle will follow the path dictated by the streamlines that correspond to the pattern B. As the flow field switches from A to B, the particle will trace a zigzag path that eventually will cause it to sample most of the cavity's area. **Figure 70** depicts the evolution of a material blob (i.e., dye) with 10^4 passive tracer particles when $T = 8$. These figures illustrate the stretching and folding process of the blob. Ultimately, the fluid particles spread to cover almost the entire area of the cavity. Numerical simulations conclude that chaotic mixing occurs by alternating two or more flow patterns in a cavity and in an open channel through the spatial and temporal control of

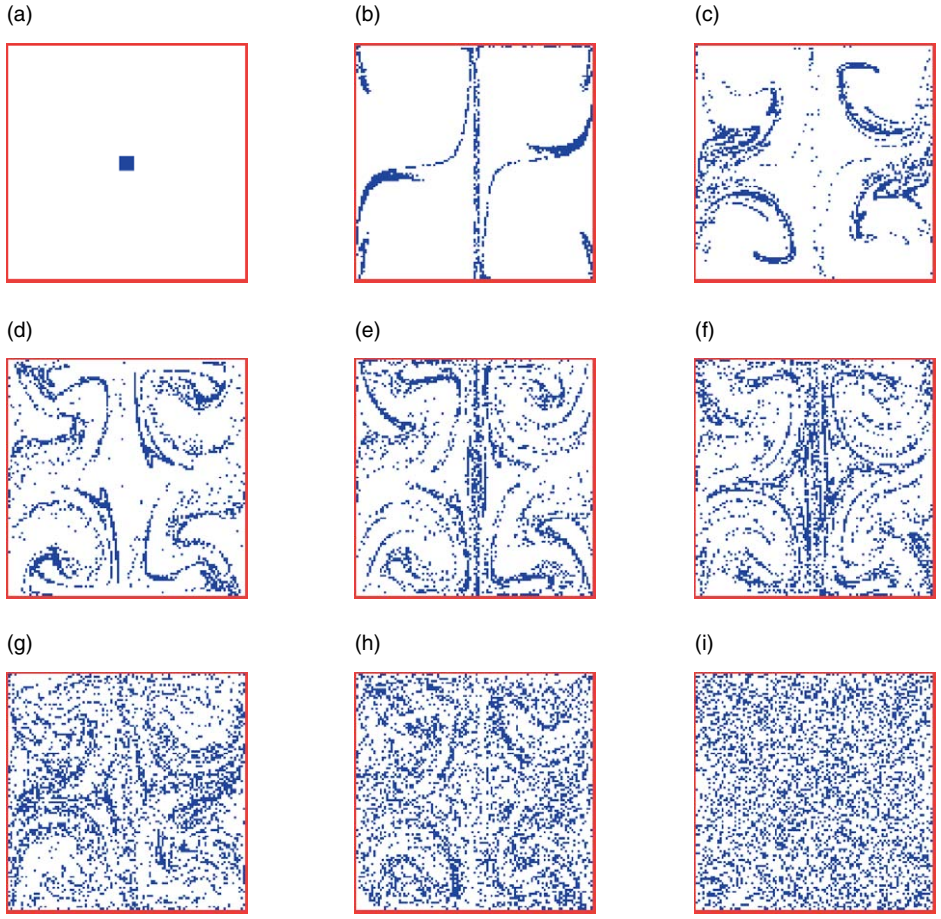


Figure 70 Deformation of a material blob of edge size 0.1 initially ($t = 0$) centered at (0, 0) at $T = 8$. (a) $t = 0$; (b) $t = 2T$; (c) $t = 4T$; (d) $t = 6T$; (e) $t = 7T$; (f) $t = 8T$; (g) $t = 10T$; (h) $t = 12T$; and (i) $t = 20T$. (Source: Qian S, Bau H H 2005 Theoretical investigation of electro-osmotic flows and chaotic stirring in rectangular cavities. *Appl. Math. Modell.* **29**, 726–53.)

the gate potentials (Qian and Bau 2002, 2005b). Later, the proposed electroosmotic mixer has been successfully tested in experiments (Lin *et al.* 2005, Wu and Liu 2005).

Figure 71 shows photograph of the electroosmotic mixer developed by Lin *et al.* (2005). Two fluid streams are brought together through the Y-junction and flow downstream in the straight microchannel equipped with inclined shielding electrodes (black segments). One can dynamically control the local zeta potential variations along the surface of silica-based microchannel through the control of the electric potential applied to the shielding electrodes. When a high voltage is applied to the shielding electrodes, negative charges will be induced on the channel surface by the capacitance effect. **Figure 72** depicts the images of fluorescence dye in the mixing region where the five-pair shielding electrodes are located. During the mixing, the shielding electrodes 1, 4, 5, 8, and 9 are activated with an applied potential of 900 V during the first half period, and the shielding electrodes 2, 3, 6, 7, and 10 are activated with an applied potential of 900 V during the second half period, and this procedure is repeated at a frequency of 0.5 Hz. Initially, the flow is laminar and stable with the two fluids clearly separated. When the control shielding voltage is alternately applied on the inclined shielding electrodes, significant mixing is observed due to the induced wave-shape flow pattern.

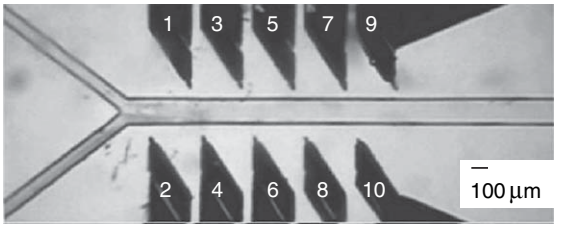


Figure 71 Photograph of the electroosmotic mixer with inclined shielding electrodes. Spatial and temporal variations of the zeta potential along the microchannel are controlled through the control of the potentials applied to the individually shielding electrodes. (Source: Lin J L, Lee K H, Lee G B 2005 Active mixing inside microchannels utilizing dynamic variation of gradient zeta potentials. *Electrophoresis* **26**, 4605–15.)

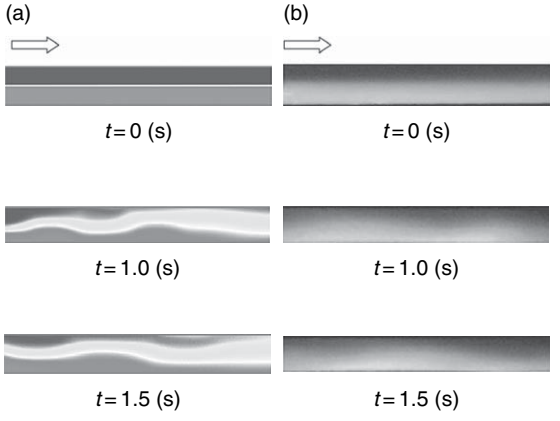


Figure 72 Images of fluorescence dye in the mixing region with five-pair inclined shielding electrodes. (a) Simulation results and (b) experimental results. (Source: Lin J L, Lee K H, Lee G B 2005 Active mixing inside microchannels utilizing dynamic variation of gradient zeta potentials. *Electrophoresis* **26**, 4605–15.)

A similar electroosmotic mixer was developed by Wu and Liu (2005) with Al electrodes embedded at the middle of the straight channel. By periodic control of the zeta potentials via the gate potentials applied to the embedded Al electrodes, excellent mixing efficiency of over 90% has been successfully achieved after the liquid flows through a 5-mm-long microchannel.

Since the electroosmotic velocity is proportional to the local zeta potential and the local electric field, instead of dynamically controlling the variations of the local zeta potentials, one can also manipulate the electric field with fixed zeta potentials along the channel walls. For example, one can enhance mixing using alternatively switching EOF by applying a switching DC driving electric field (Lin *et al.* 2004a,b). This mixer is basically a T-mixer that works with a switching DC field to induce EOFs for sample driving and flow instability generation. A ground electrode is located at the outlet of the T-mixer. In a convectional switching mode, the driving voltage is applied alternatively first to one inlet reservoir and then to the other. In other words, a driving voltage is applied to one inlet, denoted as inlet A, and the other inlet, denoted as inlet B, remains open during the first half period. During the second half period, the driving voltage is applied to the inlet B, and the inlet A is open. This procedure is repeated with a period T. The switching of the driving voltage from one sample reservoir to the other is achieved with a high-voltage relay. A floating voltage that is about 80% of the applied driving voltage is established at the inlet B when the driving voltage is applied to the inlet A, and the presence of the floating voltage prevents the fluid exiting from the inlet A from entering the inlet channel B. Therefore, the fluids from the two inlet channels flow alternatively into the mixing channel, which

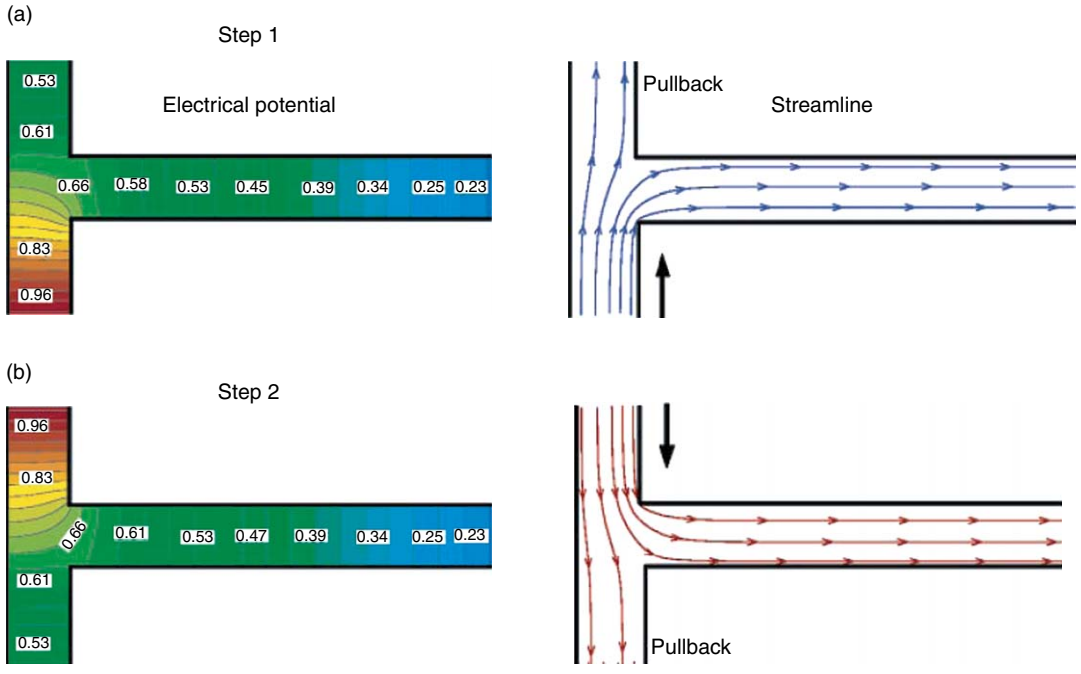


Figure 73 Electric potential distributions and streamlines in pinched switching model. (Source: Lin C H, Fu L M, Chien Y S 2004 Microfluidic T-form mixer utilizing switching electroosmotic flow. *Anal. Chem.* **76**, 5265–72.)

increases the flow instability and the contact area of the two fluid samples. However, the maximum mixing efficiency is below 60% at different switching frequencies. To improve the mixing efficiency, a pinched switching model is proposed in which a lower voltage is established at the open inlet channel by connecting a series of high-ohmic resistors between the reservoir and the ground end. Therefore, the fluid from the driven inlet flows not only into the mixing channel but also partially into the open inlet as shown in **Figure 73**. When the applied voltage switches to the other inlet, the fluid that previously flowed into the open inlet is expelled. As compared to the convectional switching model, continuous flow instability with larger amplitude is generated in the pinched switching model. The pinched switching model significantly improves the mixing performance and a 97% mixing performance has been experimentally achieved within a mixing channel 1 mm downstream from the T-junction when a 60 V cm^{-1} driving electric field is applied at a switching frequency of 2 Hz. **Figure 74** depicts the numerical and experimental results of the T-mixer with pinched switching EOF. The maximum mixing efficiency of 90% is obtained within 2 s for the driving electric field of 180 V cm^{-1} at a switching frequency of 8 Hz. Later, Fu *et al.* (2005)

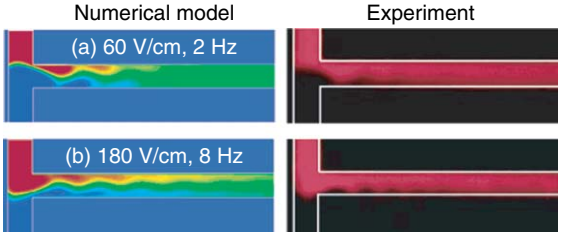


Figure 74 Numerical and experimental results of the T-mixer with alternately switching electroosmotic flow (EOF). (Source: Lin C H, Fu L M, Chien Y S 2004 Microfluidic T-form mixer utilizing switching electroosmotic flow. *Anal. Chem.* **76**, 5265–72.)

utilized the same strategy of switching driving voltage in a double T-mixer. Relative to the previous electroosmotic mixer in which the local zeta potentials are dynamically controlled, the fabrication and the operation of this mixer are simpler since it does not require the fabrication and delicate control of the embedded gate electrodes.

When a driving potential is applied to a pair of electrodes that are arranged in such a manner that they remain parallel to the channel, it can be expected that the EOF crosses the channel, and such flow can be used to perturb the main pressure-driven flow in the mixing channel (Sasaki *et al.* 2006, Sundaram and Tafti 2004). **Figure 75** schematically

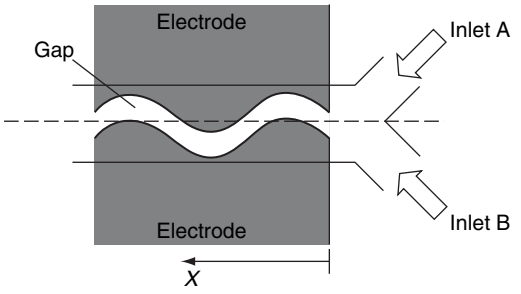


Figure 75 Schematic of a Y-shaped microfluidic channel with a pair of meandering electrodes located at the bottom of the channel to induce AC-EOF. (Source: Sasaki N, Kitamori T, Kim H B 2006 AC electroosmotic micromixer for chemical processing in a microchannel. *Lab Chip* **6**, 550–4.)

depicts a Y-shaped microfluidic channel with a pair of meandering electrodes patterned at the bottom of the channel. Two solutions are introduced into the Y-shaped channel with a two-port syringe pump, and they flow downstream in the straight microchannel of a 120 μm width and a 40 μm depth. The gap between the electrodes is 40 μm, and the periodicity of the meandering structure is 250 μm. When an AC voltage is applied to the meandering electrodes, an asymmetric circulating EOF perpendicular to the main flow is induced. The purpose of the meandering electrodes is to make periodical changes in both direction and velocity of the induced AC-EOF along the fluid stream. **Figure 76** depicts the typical experimental results when two solutions are introduced into the channel at a rate of 1 μl min^{-1}, and the total flow rate of the confluence is 2 μl min^{-1}, which corresponds to a mean velocity of 8.1 mm s^{-1} and $Re \approx 0.5$. In the images, the fluorescence intensity from part of the electrodes is enhanced due to reflection effects on the electrode surface. Before applying the AC voltage (**Figure 76(a)**), the clear interface between the two fluid streams indicates poor mixing within a length of several hundred micrometers from the junction. When a sinusoidal voltage ($V_{pp} = 20$ V at 1.0 kHz) is applied to the meandering electrodes, two fluid streams are distorted immediately and the fluorescence from the upper half side of the channel is observed at about $x > 250$ μm (**Figure 76(b)**). The experimental results from Sasaki *et al.* (2006) demonstrated that the mixing length increased almost linearly with the flow velocity of solution and the mixing time was independent of the flow velocity. For a flow velocity up to 12 mm s^{-1}, 90% mixing has been experimentally achieved within 0.2 s, while 4.4 s is required for diffusion mixing under the same condition. The mixing length also increases with the viscosity and/or the ionic strength of the solution due to the decrease in the magnitude of the AC-EOF.

When two fluid streams with different conductivities meet in a microchannel, electrokinetic instability (EKI) originating from polarization due to a conductivity gradient under high electric field conditions occurs (Chen *et al.* 2005, Lin *et al.* 2004a, b, Park *et al.* 2005, Shin *et al.* 2005). **Figure 77** depicts the EKI in a microchannel of 1 mm width, 100 μm depth, and 40 mm length as observed by Lin *et al.* (2004a, b). Two buffer solutions with different

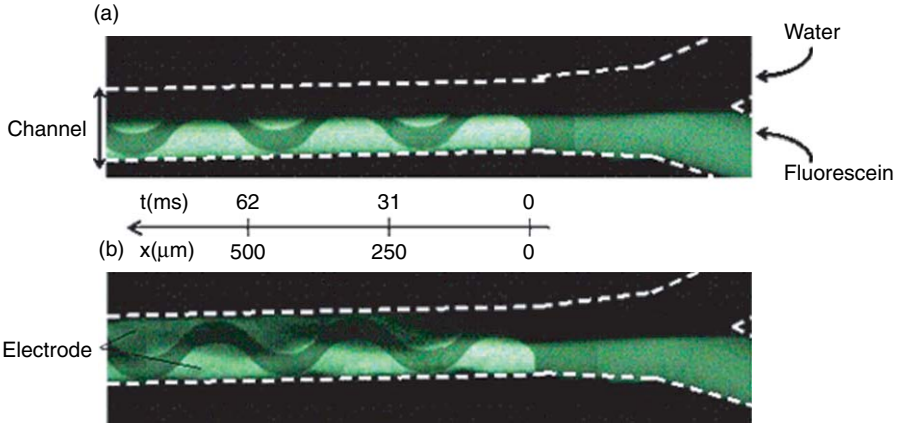


Figure 76 Experimental mixing results in (a) the absence and (b) the presence of an AC voltage ($V_{pp} = 20$ V at 1 kHz) applied to the meandering electrodes. (Source: Sasaki N, Kitamori T, Kim H B 2006 AC electroosmotic micromixer for chemical processing in a microchannel. *Lab Chip* **6**, 550–4.)

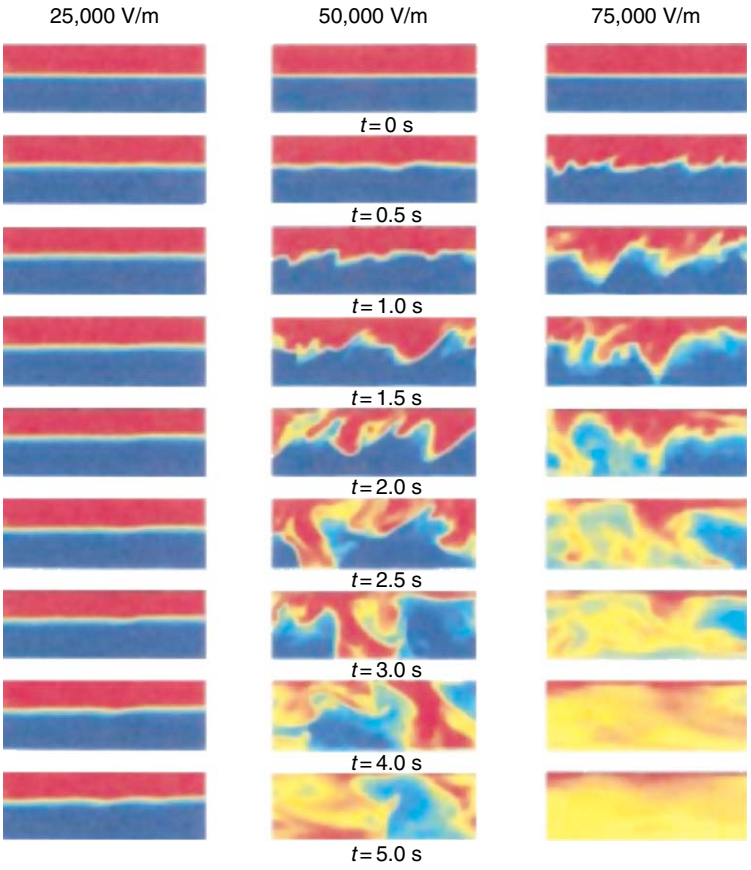


Figure 77 Instability of an electroosmotic flow (EOF) with solutions of different electrical conductivities under various electric fields as a function of the time after introduction in the microchannel. (Source: Lin H, Storey B D, Oddy M H, Chen C, Santiago J G 2004 Instability of electrokinetic microchannel flows with conductivity gradients. *Phys. Fluids* **16**, 1922–35.)

conductivities, one dyed buffer solution with a conductivity of 50 $\mu s\,cm^{-1}$and the other undyed buffer solution with a conductivity of 5 $\mu s\,cm^{-1}$, are introduced into the microchannel using a syringe pump. After the fluids are loaded into the channel, the syringe pump is deactivated and a potential difference is applied to the driving electrodes located at the inlet and outlet reservoirs that connect the channel. The driving electric field initiates an EOF in the channel. In this figure, the blue color pertains to the undyed, low-conductivity stream, and the red color represents the dyed stream with high conductivity. For a low electric field of 25 $kV\,m^{-1}$, the interface between the two streams is only slightly perturbed at 4.0 and 5.0 s. When the strength of the applied electric field exceeds a certain threshold value, an unstable fluid motion is generated, which provokes stretching and folding of the interface and thereby a rapid mixing of the two streams. At an electric field of 75 $kV\,m^{-1}$, the fluids are well mixed within 5 s. The conclusion of this is that EKI phenomenon can be successfully used to enhance mixing in microfluidic channels.

The EKI phenomenon has been adopted for micromixing by applying DC or AC electric fields in the lateral direction on liquid layers with different ionic concentrations. Strong downstream mixing was observed by the induced EKI for EOF (Oddy *et al.* 2001, Park *et al.* 2005). **Figure 78** schematically depicts a micromixer that incorporates the EKI as the stirring mechanism for rapid stirring of two fluid streams. The micromixer consists of two fluid channels of 300 μm width and 100 μm depth and a square mixing chamber (1 mm × 1 mm × 100 μm). The micromixer is fabricated using two wet-etched Borofloat glass substrates. The two fluid streams are

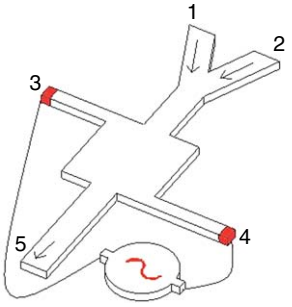


Figure 78 Schematic of an electrokinetic instability (EKI) mixer.

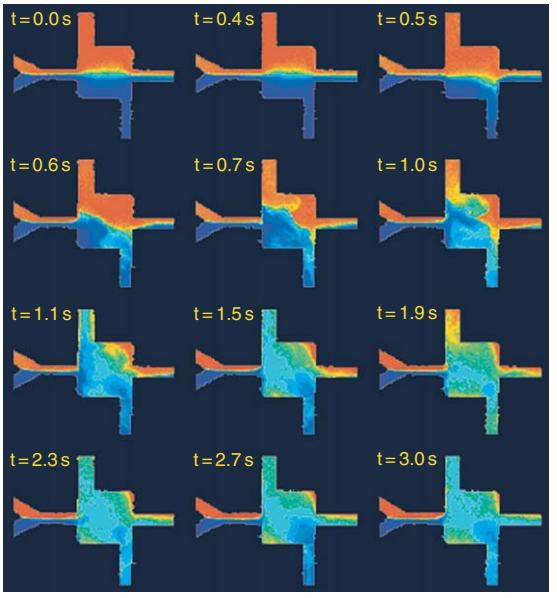


Figure 79 Experimental results of the electrokinetic instability (EKI) mixer. Fluids A and B are continuously advected from left to right, and the fluid interface is immediately disturbed upon application of the AC electric field. (Source: Oddy M H, Santiago J G, Mikkelsen J C 2001 Electrokinetic instability micromixing. *Anal. Chem.* **73**, 5822–32.)

introduced into the inlet ports 1 and 2 and are advected either electroosmotically or with pressure toward the square mixing chamber. Side channel ports 3 and 4, connected to either side of the mixing chamber, have two electrodes connected to them between which an AC electric field is applied. Molecular diffusion continues the mixing process while the stirred fluid is advected downstream from the mixing chamber toward port 5. Under certain oscillating frequencies and strengths of the applied AC electric field in the mixing chamber, EKI is induced, which in turn leads to stretching and folding of the fluids in the mixing chamber, and thus enhances mixing. In contrast to the previous electroosmotic mixer, depicted in **Figure 75**, where there is no conductivity gradient in the introduced fluids, the mixing mechanism in the EKI mixer is the EKI in the presence of conductivity gradients in the bulk fluids. **Figure 79** depicts the experimental results, and the fluid in the mixing chamber is qualitatively well mixed approximately 2.5 s after the application of the AC electric field. The disadvantage of this mixer design is that low-frequency AC voltages on the order of kilovolts are required for efficient mixing.

2.10.4 Conclusions

Laminar mixing in microfluidic devices is difficult to achieve due to the inherently low Reynolds numbers of the flows and the lack of turbulence. Various passive and active mixers have been developed to enhance laminar mixing in microfluidic LOC devices. In passive mixers, the mixing is obtained by the natural motion of the fluid as it flows through the mixing elements where mixing relies on molecular diffusion and/or chaotic advection. The microchannel geometries of passive mixers are often complicated and require intricate fabrication to create complex 3D geometries. The active mixers are principally driven by chaotic advection and are therefore usually more effective in mixing than their passive counterparts. However, the appropriate functioning of active mixers requires the use of accessories to supply adequate time-dependent external energy sources, which makes them difficult and expensive to be integrated into microfluidic systems. In contrast, passive mixers can be readily integrated in microfluidic systems because they do not need external power sources except for those engineered in the field of fluid delivery.

Although many different types of micromixers have been proposed, each type of mixer has its own advantages, drawbacks, and limitations. There is no single mixing device that is suitable for all applications and some micromixers are more appropriate for a certain application than others. The comparisons of the various mixers used in microsystems are documented in **Table 1**.

Table 1 Various micromixers for microsystems

	Description	*Advantages*	*Disadvantages*
Passive mixers			
Parallel lamination mixers	Split the inlet fluid streams into several substreams and subsequently combine these into one single stream	Simple structure; no moving parts	Slow mixing speed; requires long channel length
Split-and-recombine mixers	One stream is splitted into n laminae with width equal to $1/n$ of the original channel's width. Subsequently, two alternating thin laminae are recombined until all partial flows are united into one broad outlet channel	No moving parts; fast mixing	Large pressure drop; substantial dead volumes; delicate fabrication
Injection mixers	The fluids are injected through micronozzles and subsequently converted into microplumes	No moving parts; fast mixing	Large pressure drop; delicate fabrication
Droplet mixers	Mix the liquids within a droplet moving through winding microchannels	No moving parts; fast mixing	Poor kinetic performance of concentrated systems to be mixed
Chaotic mixers	Spatial-dependent transverse velocity components are created with complex 3D structures or channel walls patterned with various surface charges	No moving parts; chaotic advection	Complex 3D structure; delicate fabrication; relatively large Reynolds numbers are required; long channel length
Active mixers			
Pressure perturbation mixer	Time-variant pressure perturbations are applied transversely to the main stream	Chaotic advection; no moving parts	External pressure source and switching valves are required
Thermosiphon-based mixer	Natural convection is induced in the presence of large, time-dependent temperature variations	No moving parts	Requires large temperature variations and thus possibly alters the chemical/structural properties of thermosensitive systems to be mixed
Magnetic mixer	By time-wise and local control of the magnetic field, chaotic trajectories of the magnetic beads are generated, which results in mixing	No moving parts; chaotic advection	High power consumption; Joule heating
Dielectrophoretic mixer	Polymer particles are driven by time-wise DEP forces to generate chaotic trajectories	Chaotic advection; no moving parts	Requires high electric field; Joule heating
Peristaltic mixer	The fluid motion is induced by the traveling wave along the boundary wall	Chaotic advection	Moving mechanical parts (walls)
Acoustic mixer	Fluid flows in the same direction as the direction of propagation of the ultrasonic wave. The 3D flow patterns can be generated by time-wise activation of two or more transducers along different directions	Chaotic advection; no moving parts; nonintrusive	High power consumption; Joule heating
Magnetohydrodynamic mixer	Secondary flow is induced by the Lorentz force through the interaction between the electric and magnetic fields	No moving parts; low voltage; convenient circular flow; easy fabrication	The solution must be weakly conductive; electrode corrosion; bubble generation; body force
Electroosmotic mixer	Complex electroosmotic flows are induced by temporally varying the wall's zeta potentials or the applied electric field	No moving part	Requires high electric field

References

Aref H 1984 Stirring by chaotic advection. *J. Fluid Mech.* **143**, 1–21

Aref H 1990 Chaotic advection of fluid particles. *Proc. R. Soc. Lond. A* **333**, 273–88

Aref H 2002 The development of chaotic advection. *Phys. Fluids* **14**, 1315–25

Bau H H 2001 A case for magneto-hydrodynamics (MHD). *IMECE 2001*, *MEMS 23884 Symposium Proceedings*, New York, NY, November 2001

Bau H H, Zhong J, Yi M 2001 A minute magneto-hydrodynamic (MHD) mixer. *Sens. Actuators B* **79**, 205–13

Bau H H, Zhu J, Qian S, Xiang Y 2002 A magneto-hydrodynamic micro fluidic network. *IMECE 2002-33559*, *Proceedings of IMECE'02*, *2002 ASME International Mechanical Engineering Congress and Exposition*, New Orleans, LA, USA, November 17–22, 2002

Bau H H, Zhu J, Qian S, Xiang Y 2003 A magneto-hydrodynamically controlled fluidic network. *Sens. Actuators B* **88**, 205–16

Bessoth F G, deMello A J, Manz A 1999 Microstructure for efficient continuous flow mixing. *Anal. Commun.* **36**, 213–15

Biddiss E, Erickson D, Li D 2004 Heterogeneous surface charge enhanced micromixing for electrokinetic flows. *Anal. Chem.* **76**, 3208–13

Biswal S L, Gast A P 2004 Micromixing with linked chains of paramagnetic particles. *Anal. Chem.* **76**, 6448–55

Bringer M R, Gerdts C J, Song H, Tice J D, Ismagilov R F 2004 Microfluidic systems for chemical kinetics that rely on chaotic mixing in droplets. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng.* **362**, 1087–104

Carlsson F, Sen M, Lofdahl L 2005 Fluid mixing induced by vibrating walls. *Eur. J. Mech. B Fluids* **24**, 366–78

Chen H, Meiners J-C 2004 Topologic mixing on a microfluidic chip. *Appl. Phys. Lett.* **84**, 2193–5

Chen Z, Qian S, Abrams W, Malamud D, Bau H H 2004 Thermosiphon-based PCR reactor: Experiment and modeling. *Anal. Chem.* **76**, 3707–15

Chen C H, Lin H, Lele S K, Santiago J G 2005 Convective and absolute electrokinetic instability with conductivity gradients. *J. Fluid Mech.* **524**, 263–303

Chow A W 2002 Lab-on-a-chip: Opportunities for chemical engineering. *AICHE J.* **48**, 1590–5

Davidson P A 2001 *An Introduction to Magnetohydrodynamics*. Cambridge Press, Cambridge

De la Cruz L M, Ramos E 2006 Mixing with time dependent natural convection. *Int. Commun. Heat Mass Transfer* **33**, 191–8

Deval J, Tabeling P, Ho C-M 2002 A dielectro-phoretic chaotic mixer. *Proc. 15th IEEE Int. Conf. Micro Electro Mechanical Systems 2002*, Las Vegas, NV, USA, pp. 36–9

Dittrich P S, Manz A 2006 Lab-on-a-chip: Microfluidics in drug discovery. *Nat. Rev. Drug Discov.* **5**, 210–18

Dittrich P S, Tachikawa K, Manz A 2006 Micro total analysis systems. Latest advancements and trends. *Anal. Chem.* **78**, 3887–907

Erbacher C, Bessoth F G, Busch M, Verpoorte E, Manz A 1999 Towards integrated continuous-flow chemical reactors. *Mikrochim. Acta* **131**, 19–24

Fu L M, Yang R J, Lin C H, Chien Y S 2005 A novel microfluidic mixer utilizing electrokinetic driving force under low switching frequency. *Electrophoresis* **5**, 1814–24

Gardeniers H, Van den Berg A 2004 Micro- and nanofluidic devices for environmental and biomedical applications. *Int. J. Environ. Anal. Chem.* **84**, 809–19

Gleeson J P 2005 Transient micromixing: Examples of laminar and chaotic stirring. *Phys. Fluids* **17**, 100614–9

Gleeson J P, Roche O M, West J, Gelb A 2004 Modelling annular micromixers. *SIAM J. Appl. Math.* **64**, 1294–310

Gobby D, Angeli P, Gavriilidis A 2001 Mixing characteristics of T-type microfluidic mixers. *J. Micromech. Microeng.* **11**, 126–32

Grumann M, Geipel A, Riegger L, Zengerle R, Ducrée J 2005 Batch-mode mixing on centrifugal microfluidic platforms. *Lab Chip* **5**, 560–565

Günther A, Jhunjhunwala M, Thalmann M, Schmidt M A, Jensen K F 2005 Micromixing of miscible liquids in segmented gas–liquid flow. *Langmuir* **21**, 1547–55

Hadd A G, Raymond D E, Halliwell J W, Jacobson S C, Ramsey J M 1997 Microchip device for performing enzyme assays. *Anal. Chem.* **69**, 3407–12

He B, Burke B J, Zhang X, Zhang R, Regnier F E 2001 A picoliter-volume mixer for microfluidic analytical systems. *Anal. Chem.* **73**, 1942–7

Hertzog D E, Michalet X, Jäger M, Kong X, Santiago J G, Weiss S, Bakajin O 2004 Femtomole mixer for microsecond kinetic studies of protein folding. *Anal. Chem*. **76**, 7169–78

Hinsmann P, Frank J, Svasek P, Harasek M, Lendl B 2001 Design, simulation and application of a new micromixing device for time resolved infrared spectroscopy of chemical reactions in solutions. *Lab Chip* **1**, 16–21

Hoffmann M, Schluter M, Rabiger N 2006 Experimental investigation of liquid–liquid mixing in T-shaped micro-mixers using μ-LIF and μ-PIV. *Chem. Eng. Sci.* **61**, 2968–76

Howell P B, Mott D R, Fertig S, Kaplan C R, Golden J P, Oran E S, Ligler F S 2005 A microfluidic mixer with grooves placed on the top and bottom of the channel. *Lab Chip* **5**, 524–30

Huikko K, Kostiainen R, Kotiaho T 2003 Introduction to micro-analytical systems: Bioanalytical and pharmaceutical applications. *Eur. J. Pharm. Sci.* **20**, 149–71

Jain K K 2000 Biochips and microarrays: Technology and commercial potential. Pharmaceutical Industry Report, published by Informa Pharmaceuticals.

Jang J, Lee S S 2000 Theoretical and experimental study of MHD (Magnetohydrodynamic) micropump. *Sens. Actuators A* **80**, 84–9

Jen C-P, Wu C-Y, Lin Y-C, Wu C-Y 2003 Design and simulation of the micromixer with chaotic advection in twisted microchannels. *Lab Chip* **3**, 77–81

Jensen K 1998 Chemical kinetics – smaller, faster chemistry. *Nature* **393**, 735–7

Jensen K F 1999 Microchemical systems: Status, challenges, and opportunities. *AIChE J.* **45**, 2051–4

Jiang F, Drese K S, Hardt S, Kupper M, Schonfeld F 2004 Helical flows and chaotic mixing in curved micro channels. *AICHE J.* **50**, 2297–305

Johnson T J, Ross D, Locascio L E 2002 Rapid microfluidic mixing. *Anal. Chem.* **74**, 45–51

Jones T 1995 *Electromechanics of Particles*. Cambridge University Press, Cambridge

Kamholz A E, Weigl B H, Finlayson B A, Yanger P 1999 Quantitative analysis of molecular interactive in microfluidic channel: The T-sensor. *Anal. Chem.* **71**, 5340–7

Katou H, Miyake R, Kambara K, Kawase K, Uchida H 2005a Development of a non-contact micro-liquid mixing method and application to chemical auto-analyzers. *Chem. Eng. Sci.* **60**, 5519–28

Katou H, Miyake R, Terayama T 2005b Non-contact micro-liquid mixing method using ultrasound. *JSME Int. J. Series B Fluids Therm. Eng.* **48**, 350–5

Knight J B, Vishwanath A, Brody J P, Austin R H 1998 Hydrodynamic focusing on a silicon chip: Mixing nanoliters in microseconds. *Phys. Rev. Lett.* **80**, 3863–6

Krishnan N, Agrawal N, Burns M A, Ugaz V M 2004 Reactions and fluidics in miniaturized natural convection systems. *Anal. Chem.* **76**, 6254–65

Langer R 2000 Biomaterials: Status, challenges, and perspectives. *AIChE J.* **46**, 1286–9
Lee Y K 2002 Lyapunov exponents of a micro chaotic mixer. *Int. J. Nonlinear Sci. Numer. Simul.* **3**, 561–4
Lee C S, Blanchard W C, Wu C T 1990 Direct control of the electroosmosis in capillary zone electrophoresis by using an external electric field. *Anal. Chem.* **62**, 1550–2
Lee Y K, Tabeling P, Shih C, Ho C M 2000 Characterization of a MEMS-fabricated mixing device. *Proceedings of MEMS, ASME International Mechanical Engineering Congress and Exposition*, Orlando, FL, USA, November 2000, pp. 505–11
Lee Y K, Deval J, Tabeling P, Ho C M 2001 Chaotic mixing in electrokinetically and pressure driven micro flows. *IEEE 14th Int. Conf. Micro Electro Mechanical System (MEMS 2001)*, Interlaken, Switzerland, pp. 483–6
Lee J S H, Ren C R, Li D 2005 Effects of surface heterogeneity on flow circulation in electroosmotic flow in microchannels. *Anal. Chim. Acta* **530**, 273–82
Lemoff A V, Lee A P 2000 An AC magnetohydrodynamic micropump. *Sens. Actuators B* **63**, 178–85
Lemoff A V, Lee A P 2003 An AC magnetohydrodynamic microfluidic switch for micro total analysis systems. *Biomed. Microdevices* **5**, 55–60
Liau A, Karnik R, Majumdar A, Cate J H D 2005 Mixing crowded biological solutions in milliseconds. *Anal. Chem.* **77**, 7618–25
Lin C H, Fu L M, Chien Y S 2004a Microfluidic T-form mixer utilizing switching electroosmotic flow. *Anal. Chem.* **76**, 5265–72
Lin H, Storey B D, Oddy M H, Chen C, Santiago J G 2004b Instability of electrokinetic microchannel flows with conductivity gradients. *Phys. Fluids* **16**, 1922–35
Lin J L, Lee K H, Lee G B 2005 Active mixing inside microchannels utilizing dynamic variation of gradient zeta potentials. *Electrophoresis* **26**, 4605–15
Liu R H, Stremler M A, Sharp K V, Olsen M G, Santiago J G, Adrian R J, Aref H, Beebe D J 2000 Passive mixing in a three-dimensional serpentine microchannel. *J. Microelectromech. Syst.* **9**, 190–7
Liu R H, Yang J N, Pindera M Z, Athavale M, Grodzinski P 2002 Bubble-induced acoustic micromixing. *Lab Chip* **2**, 151–7
Liu R H, Lenigk R, Druyor-Sanchez R L, Yang J N, Grodzinski P 2003 Hybridization enhancement using cavitation microstreaming. *Anal. Chem.* **75**, 1911–17
Mengeaud V, Josserand J, Girault H H 2002 Mixing processes in a zigzag microchannel: Finite element simulations and optical study. *Anal. Chem.* **74**, 4279–86
Miyake R, Lammerinkz T S J, Elwenspoek M, Fluitman J H J 1993 Micro mixer with fast diffusion. *Proc. MEMS'93, 6th IEEE Int. Workshop Micro Electromechanical System*, San Diego, CA, USA, pp. 248–53
Munson M S, Yager P 2004 Simple quantitative optical method for monitoring the extent of mixing applied to a novel microfluidic mixer. *Anal. Chim. Acta* **507**, 63–71
Muradoglu M, Stone H A 2005 Mixing in a drop moving through a serpentine channel: A computational study. *Phys. Fluids* **17**, 073305
Niu X, Lee Y K 2003 Efficient spatial–temporal chaotic mixing in microchannels. *J. Micromech. Microeng.* **13**, 454–62
Niu X, Liu L, Wen W, Sheng P 2006 Active microfluidic mixer chip. *Appl. Phys. Lett.* **88**, 153508
Ng A S W, Hau W L W, Lee Y K, Zohar Y 2004 Electrokinetic generation of microvortex patterns in a microchannel liquid flow. *J. Micromech. Microeng.* **14**, 247–55
Nguyen N T, Wu Z G 2005 Micromixers – A review. *J. Micromech. Microeng*. **15**, R1–16
Oddy M H, Santiago J G, Mikkelsen J C 2001 Electrokinetic instability micromixing. *Anal. Chem.* **73**, 5822–32
Ottino J M 1989 *The Kinematics of Mixing: Stretching, Chaos, and Transport*. Cambridge University Press, Cambridge
Park J, Shin S M, Huh K Y, Kang I S 2005 Application of electrokinetic instability for enhanced mixing in various micro-T-channel geometries. *Phys. Fluids* **17**, 118101
Park H Y, Qiu X, Rhoades E, Korlach J, Kwok L W, Zipfel W R, Webb W W, Pollack L 2006 Achieving uniform mixing in a microfluidic device: Hydrodynamic focusing prior to mixing. *Anal. Chem.* **78**, 4465–73
Qian S, Bau H H 2002 A chaotic electroosmotic stirrer. *Anal. Chem.* **74**, 3616–25
Qian S, Bau H H 2005a Magnetohydrodynamic stirrer for stationary and moving fluids. *Sens. Actuators B* **106**, 859–70
Qian S, Bau H H 2005b Theoretical investigation of electro-osmotic flows and chaotic stirring in rectangular cavities. *Appl. Math. Modell.* **29**, 726–53
Qian S, Zhu J, Bau H H 2002 A stirrer for magnetohydrodynamically controlled minute fluidic networks. *Phys. Fluids* **14**, 3584–92
Rivallin M, Benmami M, Kanaev A, Gaunand A 2005 Sol–gel reactor with rapid micromixing – Modelling and measurements of titanium oxide nano-particle growth. *Chem. Eng. Res. Design* **83**, 67–74
Sasaki N, Kitamori T, Kim H B 2006 AC electroosmotic micromixer for chemical processing in a microchannel. *Lab Chip* **6**, 550–4
Schasfoort R B M, Schlautmann S, Hendrikse J, Van den Berg A 1999 Field-effect flow control for microfabricated fluidic networks. *Science* **286**, 942–5
Schonfeld F, Hardt S 2004 Simulation of helical flows in microchannels. *AICHE J.* **50**, 771–8
Schwarzer H C, Peukert W 2004 Combined experimental/numerical study on the precipitation of nanoparticles. *AICHE J.* **50**, 3234–47
Selverov K P, Stone H A 2001 Peristaltically driven channel flows with applications toward micromixing. *Phys. Fluids* **13**, 1837–59
Shin S M, Kang I S, Cho Y-K 2005 Mixing enhancement by using electrokinetic instability under time-periodic electric field. *J. Micromech. Microeng.* **15**, 455–62
Song H, Bringer M R, Tice J D, Gerdts C J, Ismagilov R F 2003 Experimental test of scaling of mixing by chaotic advection in droplets moving through microfluidic channels. *Appl. Phys. Lett.* **83**, 4664–6
Stahl M, Aslund B, Rasmuson A C 2004 Aging of reaction-crystallized benzoic acid. *Ind. Eng. Chem. Res.* **43**, 6694–702
Stone H A, Kim S 2001 Microfluidics: Basic issues, applications, and challenges. *AICHE J.* **47**, 1250–4
Stremler M A, Olsen M G, Adrian R J, Aref H, Beebe D J 2000 Chaotic mixing in microfluidic systems. *Solid-state Sensor and Actuator Workshop*, Hilton Head, SC, USA, June 4–8, 2000
Stroock A D, Weck M, Chiu D T, Huck W T S, Kenis P J A, Ismagilov R F, Whitesides G M 2000 Patterning electroosmotic flow with patterned surface charge. *Phys. Rev. Lett.* **84**, 3314–17
Stroock A D, Dertinger S K W, Ajdari A, Mezic I, Stone H A, Whitesides G M 2002 Chaotic mixer for microchannels. *Science* **295**, 647–51
Sudarsan A P, Ugaz V M 2006 Multivortex micromixing. *Proc. Natl. Acad. Sci. U S A* **103**, 7228–33
Sundaram N, Tafti D K 2004 Evaluation of microchamber geometries and surface conditions for electrokinetic driven mixing. *Anal. Chem.* **76**, 3785–93
Suri C, Takenaka K, Kojima Y, Koyama K 2002a Experimental study of a new liquid mixing method using acoustic streaming. *J. Chem. Eng. Japan* **35**, 497–502
Suri C, Takenaka K, Yanagida H, Kojima Y, Koyama K 2002b Chaotic mixing generated by acoustic streaming. *Ultrasonics* **40**, 393–6
Suzuki H 2003 Development of chaotic micro-mixer using magnetic beads. Ph.D. thesis, The University of Tokyo

Suzuki H, Ho C M, Kasagi N 2004 A chaotic mixer for magnetic bead-based micro cell sorter. *J. Microelectromech. Syst.* **13**, 779–90

Taylor G 1953 Dispersion of soluble matter in solvent flowing slowly through a tube. *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.* **219**, 186–203

Therriault D, White S R, Lewis J A 2003 Chaotic mixing in three-dimensional microvascular networks fabricated by direct-write assembly. *Nat. Mater.* **2**, 265–71

Thilmany J 2005 Think small – Lab-on-a-chip technology shrinks the biological laboratory to the micro scale and expands the potential for future applications. *EMBO Rep.* **6**, 913–16

Tice J D, Song H, Lyon A D, Ismagilov R F 2003 Formation of droplets and mixing in multiphase microfluidics at low values of the Reynolds and the capillary numbers. *Langmuir* **19**, 9127–33

Veenstra T T 1999 Characterization method for a new diffusion mixer applicable in micro flow injection analysis systems. *J. Micromech. Microeng.* **9**, 199–202

Verpoorte E 2002 Microfluidic chips for clinical and forensic analysis. *Electrophoresis* **23**, 677–712

Woodson H H, Melcher J R 1969 *Electromechanical Dynamics*, John Wiley, New York, Vol. III

Wu J R, Du G H 1997 Streaming generated by a bubble in an ultrasound field. *J. Acoust. Soc. Am.* **101**, 1899–907

Wu H Y, Liu C H 2005 A noval electrokinetic micromixer. *Sens. Actuators A* **118**, 107–15

Xia H M, Wan S Y M, Shu C, Chew Y T 2005 Chaotic micromixers using two-layer crossing channels to exhibit fast mixing at low Reynolds numbers. *Lab Chip* **5**, 748–55

Xia H M, Shu C, Wan S Y M, Chew Y T 2006 Influence of the Reynolds number on chaotic mixing in a spatially periodic micromixer and its characterization using dynamical system techniques. *J. Micromech. Microeng.* **16**, 53–61

Xiang Y, Bau H H 2003 Complex magnetohydrodynamic low-Reynolds-number flows. *Phys. Rev. E* **68**, 016312

Yamaguchi Y, Ogino K, Yamashita K, Maeda H 2004 Rapid micromixing based on multilayer laminar flows. *J. Chem. Eng. Jpn* **37**, 1265–70

Yang R, Williams J D, Wang W 2004 A rapid micro-mixer/reactor based on arrays of spatially impinging micro-jets. *J. Micromech. Microeng.* **14**, 1345–51

Yaralioglu G G, Wygant I O, Marentis T C, Khuri-Yakub B T 2004 Ultrasonic mixing in microfluidic channels using integrated transducers. *Anal. Chem.* **76**, 3694–8

Yi M, Bau H H 2003 The kinematics of bend-induced mixing in micro-conduits. *Int. J. Heat Fluid Flow* **24**, 645–56

Yi M, Bau H H, Hu H H 2002a Peristaltically induced motion in a closed cavity with two vibrating walls. *Phys. Fluids* **14**, 184–97

Yi M, Qian S, Bau H H 2002b A magnetohydrodynamic chaotic stirrer. *J. Fluid Mech.* **468**, 153–77

Zhong J, Yi M, Bau H H 2002 Magneto hydrodynamic (MHD) pump fabricated with ceramic tapes. *Sens. Actuators A* **96**, 59–66

Biographies

Shizhi Qian received his first Ph.D. for his work on multiphase flow in gas–solid fluidized bed boilers from Huazhong University of Science and Technology, P.R. China, in 1998. Since 2000, he has been pursuing research in the area of lab-on-a-chip technology at the University of Pennsylvania (Upenn) and obtained his second Ph.D. on modeling complex interactions in microfluidic systems from Upenn in 2004. He is currently an assistant professor in the Department of Mechanical Engineering, University of Nevada, Las Vegas. His current research interests focus on lab-on-a-chip technology for biomedical, clinical, and environmental applications, electrokinetics and electrodics, and magnetohydrodynamics.

Jérôme F.L. Duval graduated as a Physico-Chemist engineer in 1999 from the National Graduate School of Chemistry and Physics of Bordeaux in France. From 1999 to 2003, he was with the Department of Physical Chemistry and Colloid Science of Wageningen University and Research Center, Wageningen, the Netherlands, as a Ph.D. student under the supervision of Profs. H.P. van Leeuwen and J. Lyklema. He obtained his Ph.D. thesis (Cum Laude) with a manuscript entitled Faradaic and Adsorption-Mediated Depolarization of Electric Double Layers in Colloids. After 2 years of a joined postdoctoral research program in Wageningen and in the Department of Analytical and Biophysical Environmental Chemistry, Geneva, Switzerland, he entered the Centre National de la Recherche Scientifique (CNRS) in France as a Chargé de Recherche (Assistant Professor). He is currently working in the Laboratory Environment and Mineralurgy in Nancy (France) and his current research topics cover the electrokinetics of complex interfaces (bio- and environmental particles), microfluidics, the electrokinetic phenomena of the second kind, particle interactions, and the coupling between electrode kinetics and electrokinetics.

2.11 Gas Sensors

Ulrich Bonne, Hopkins, MN, USA

Nomenclature

Variables

A	Heat transfer surface area (cm^2)
Bn	Bonne number, defined as $Bn = D_{O_2}/(M_{gas}/M_{CH_4})^{0.5}$
C	Peak or channel capacity of a GC system; or Coulomb, unit of electrical charge
c_p	Specific heat in J (mol K)$^{-1}$ or cal (mole K)$^{-1}$; 1 cal = 4.184 J
C_V	Volumetric composition correction factor (dimensionless)
d	Diameter (cm) or (in.)
D	Mass diffusivity ($cm^2\ s^{-1}$)
D_{O_2}	Molar demand for oxygen of a fuel to achieve stoichiometric combustion, e.g., $D_{O_2}(CH_4) = 2$
D_T	Thermal diffusivity ($cm^2\ s^{-1}$)
f	Frequency (Hz)
G	Sensor output; ΔG = output resulting from the difference between the downstream–upstream sensing element resistances of the sensor in the inset in Figure 1. Outputs in V or mV
h	Heat transfer per unit surface area and temperature difference (W $(m^2\ K)^{-1}$)
H	Height of a theoretical GC separation plate
k	Thermal conductivity (W (m K)$^{-1}$)
k'	Retention or capacity factor. $k' = K/\beta$
K	Equilibrium constant
L	Length (cm)
m	Mass flow (g s^{-1})
m_x	Mass flux (g $(cm^2\ s)^{-1}$)
M	Molecular weight (g mol^{-1})
Nu	Nusselt number, total heat transfer/conductive heat transfer. $Nu = Q/(k\Delta TA/d) = hd/k$
p	Pressure (dyne cm^{-2}); 1 bar = 10^6 dyne cm^{-2} = 10^5 Pa
Pr	Prandtl number, momentum diffusivity/thermal diffusivity, $Pr = v/D_t = \eta c_p/(kM)$
Re	Reynolds number, inertial force/viscous force, $Re = vd/v = m_x d/\eta = md/(A\eta)$
R	Universal gas constant, R = 1.9872 cal (mole K)$^{-1}$ = 8.3145 J (mol K)$^{-1}$
R^*	Resolution, $R^* = t_R/W_{1/2}$
t_R	Retention time (s)
t_o	Flow-trough time (s)

v	Velocity (cm s^{-1})
V	Volume (cm^3) or abbreviation for volts
$W_{1/2}$	Half-width and $W^*_{1/2}$ average half-width
α_1	First TCR, with $R = R_o(1 + \alpha_1\Delta T_o + \alpha_2\Delta T_o)$, $\alpha_1 \sim 3500$ ppm ($°C^{-1}$) for Pt
α_2	Second TCR
β	Volumetric ratio of mobile phase/stationary phase = gas/film in a GC column
γ	Ratio of specific heats, $\gamma = c_p/c_v$
Δ	Difference, as in Δp
η	(Dynamic) viscosity in Poise or g $(cm\ s)^{-1}$; 0.000178 P for N_2 at 20°C
ν	Kinematic viscosity, $\nu = \eta/\rho(\cdots$"
π	Constant, 3.1416
ρ	Density (g cm^{-3}) and specific gravity, ρ/ρ_{air}
τ	Response time, to achieve 63.2% of final signal value (s or ms)
Subscripts	
c	Capillary
m	Mass
M	Main flow channel, or mole-related
o	Reference
p	Pressure
v, V	Volume

Glossary

AC Alternating Current
ACS American Chemical Society
AED Atomic Emission Detector for GC
AEC Amperometric Electrochemical
amu Atomic mass unit
ARI Atmosphere Recovery, Inc.
CID Chemical Impedance Detector for GC or stand-alone
CPAC Center for Process Analytical Chemistry, at the University of Washington in Seattle, WA
CRDS Cavity Ring-Down Spectroscopy
CSTL Chemical Science and Technology Laboratory
DC Direct Current
DARPA US Defense Advanced Research Projects Agency
DMS Differential Mobility Spectroscopy. See also IMS
DOE US Department of Energy
DTA Differential Thermal Analysis
EC Electro-Chemical
EM Electromagnetic
EMI Electromagnetic Interference
FAR False Alarm Rate
FID (Hydrogen) Flame Ionization Detectors for GC
GC Gas Chromatograph(y)
HW Half-Width of a peak or pulse
IMS Ion Mobility Spectroscopy for GC or stand-alone. See also DMS
IS Intrinsic Safety
ITMS Ion Trap Mass Spectrometry for GC or stand-alone
MB Micro Bridge
MK Microbrick™
MDD Microdischarge Detector for GC or stand-alone
MDL Minimum Detection Level, typically defined as the analyte concentration level at which $S/N = 3$
MGA Micro Gas Analyzers
MM Microminiaturized or Man–Machine (Interface)
MS Mass Spectrometer
MW Molecular Weight, e.g., in g $mole^{-1}$
ng Nanogram
nl Nanoliter, 1 000 000 nl = 1 ml = 1 cm^3
NEL National Engineering Lab, UK
NDIR Non-Dispersive Infra Red
NIST National Institute of Standards and Technology
ORML Oak Ridge National Library
PC Preconcentrator or Preconcentration
PE Piezoelectric
pg Picogram, 1 pg = 0.001 ng
PHASED Phased Heater Array Structure for Enhanced Detection
PI Principal Investigator
ppb Parts per billion or $1/10^9$
ppm Parts per million or $1/10^6$
ppt Parts per trillion or $1/10^{12}$
PDMS Polydimethylsiloxane
PPDS Physical Property Data Service Co., USA
PTFE Polytetrafluoroethylene or Teflon
PZT Lead Zirconate Titanate, a well known and much used polycrystalline piezoelectric ceramic
RC Resonator on time constant based on resistive and capacitive elements in a circuit
R&D Research and Development
RF Radio Frequency
RMS Root Mean Square
RLGA Raman Laser Gas Analyzer

SAE Society for Automative Engineers
SAW Surface Acoustic Wave (detector), indicating adsorbed mass detector via frequency change
SEM Scanning Electron Microscope
SERS Surface-Enhanced Raman Spectrometry for GC or stand-alone
S/N Signal-to-Noise ratio
TCD Thermal Conductivity Detector for GC or stand-alone
TCR Temperature Coefficient of Resistance

2.11.1 Introduction and Fundamentals

Gas sensors are generally understood as providing a measurement of the concentration of some analyte of interest, such as CO, CO_2, NO_x, SO_2, without at this point dwelling on the plethora of underlying approaches such as optical absorption, electrical conductivity, electrochemical (EC) and catalytic bead (see Section 2.11.3). However, and as discussed in Section 2.11.2, many other gas sensors measure a physical property of the environment around them, such as simple temperature, pressure, flow, thermal conductivity, and specific heat, or more complex properties such as heating value, supercompressibility, and octane number for gaseous fuels. The latter may require capital-intensive (engines) or destructive testing, e.g., via combustion, or involve the measurement of a number of parameters to serve as inputs to a correlation with the complex property of interest.

When the sensor provides a multiplicity of outputs, as with optical or mass spectrometers (MSs), we refer to it as a gas analyzer. Gas chromatography (GC), differential thermal analysis (DTA), ion mobility, and nuclear magnetic resonance (NMR) are additional examples, some of which will be detailed in Section 2.11.4. Such analyzers, preferred by the author, should not be confused with sensor arrays, in which different sensing materials (typically polymers and metal oxides) are used on each element of the array, which then needs to conform to difficult-to-achieve stability requirements.

The performance of all of the above-mentioned sensors and analyzers may be characterized by their signal-to-noise (S/N) ratio, minimum detectable limit (MDL), selectivity, and response time. Increasingly, power consumption, size, and weight are becoming more important as interest and demand increases for handheld, battery-powered sensors, with or without wireless capability. These specifications may be viewed as simple performance parameters, because they are relatively simple to quantify.

Self-calibration, drift, S/N, and false alarm rate (FAR) (mainly for composition sensors or analyzers) require more sophisticated approaches, but are of increasing importance in all applications such as for medical, industrial, environmental, security, and first-responder use. Section 2.11.5 goes into the details of this subject.

Another classification of gas sensors and analyzers could be based on their sampling method: by diffusion, pumped transport, or via remote optical sampling to induce fluorescence, absorption, or scattering.

Researchers, designers, and planners continually face the need to make development or fabrication decisions before all the facts are available. Therefore, there is a perennial need to generate estimates about the performance and sensitivity of devices, structure, and also sensor systems. This is where mathematical modeling, be it simple or complex or *ad hoc* (to simulate a specific sensor) or multipurpose (such as ANSYS, FLUENT to simulate heat transfer or flow of a sensor within a given programmatic framework, which is adapted to individual geometries and conditions) can be of tremendous help. Rather than dedicating a section specifically to this subject, several examples will be discussed throughout this chapter.

The intent of this chapter is not to provide an exhaustive review of the world of gas sensors, nor a history of their development, but to highlight and share selected gas-sensing approaches that impressed the author in meeting modern expectations for performance, features, and cost.

The issue of cost merits additional comments. Contrary to initial negative reactions one might have about it, because of its potential commercialism, the author subscribes to the view that cost just adds a tough professional challenge (conservation, sustainable development, affordability) to all the others related to achieving generic and useful sensor performance attributes mentioned above. In fact, many elegant sensing approaches are withering on the shelf because few potential users could afford to implement them.

2.11.2 Physical Property Gas Sensors

2.11.2.1 Introduction

Among gas sensors to determine physical properties such as pressure, flow, specific heat, viscosity, and density, this section focuses on the fabrication and performance of microsensors primarily measure thermal effects. Some of the sensor outputs are direct thermal properties themselves, such as temperature, thermal conductivity, and specific heat, while others are indirect, inferred, or correlated such as flow, density, viscosity, heating value, oxygen demand, and octane and cetane numbers (Bonne 1992, 1996, Higashi *et al.* 1986). Let us start with the measurement of flow and pressure, which may be viewed as special cases in between the above-mentioned direct and indirect groups.

Traditional flow sensing with hot-wire anemometers had evolved to hot-film microanemometry in the 1980s (Higashi *et al.* 1986), for which the hot film was a 50- to 100-nm film of permalloy (FeNi) or platinum, passivated by and sandwiched between unsupported films of silicon nitride of ~1 μm in total thickness. This MB structure is created by etching out an underlying area to a depth of about 150 μm of the supporting silicon chip, which only measures 1.7 mm × 1.7 mm. **Figure 1** illustrates the makeup of this MB in cross section and in top views showing the heating and sensing elements (as well as other resistors to be detailed later on); the scanning electron microscope (SEM) photograph, shows the contact pads for gold-wire bonding. (The heating and sensing elements are also visible but are shown more clearly in the line drawing.) All resistive elements consist of the above-mentioned thin-metal film located either on the MB (R_h, R_u, R_d) or on the solid substrate (**Figures 2** and **3**). The substrate resistors are thereby held at or close to ambient temperature.

The basic measurement consists of accurately sensing changes in the absolute (for temperature) or relative (for flow) resistance of the elements. The heater (R_h, see central element in **Figure 1(a)–1(c)**) is controlled to maintain an approximately constant temperature differential of 70–200°C above ambient (~160°C for most commercial AWM-Series flow sensors (Catalog #15 1998)), as achieved by the circuit of **Figure 2(a)**. The zero temperature coefficient of

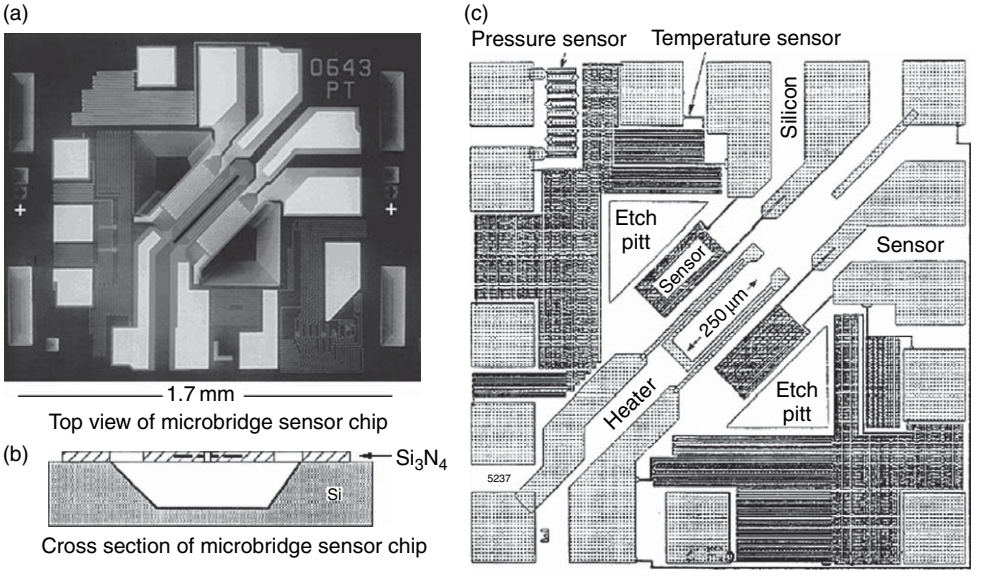


Figure 1 Various views of microbridge flow sensor chip, showing flow sensing, center heater, pressure-, and temperature-sensing elements, as well as two large reference resistor serpentines. (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission. Bonne U 1994 Thermal microsensors for environmental and industrial controls. *NIST Workshop on Gas Sensors: Strategy for Future Growth*, Gaithersburg, MD, USA, September 8–9, 1993, NIST Special Publication No. 865, 1994, pp. 31–8, ISSN: 1048-776X; reproduced with permission. **Figure 1(c)**: Bonne U, Kubisiak D 1994 Overpressure-proof, thermal pressure sensor for gases. *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, June 13–16, 1994, Tech. Digest, p. 76. Bonne U 1998 Microsensor fabrication and application. *11th IGT Symp. Gas Quality and Energy Measurement*, St. Petersburg, FL, USA, February 2–4, 1998, Proceedings; © 1998 reproduced with permission from Gas Technology Institute.)

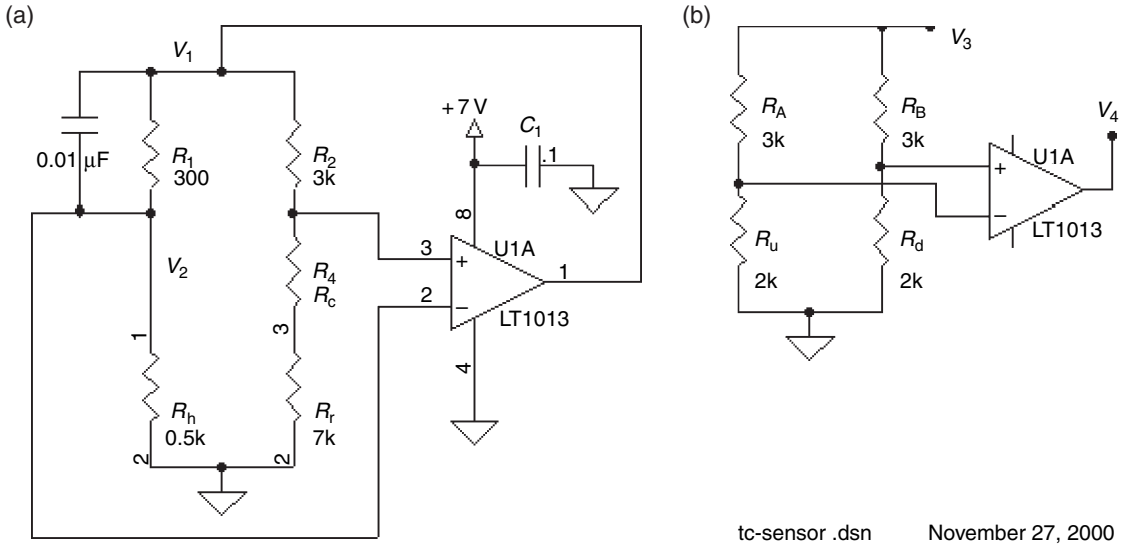


Figure 2 Basic microbridge flow sensor circuits. Elements R_h, R_u, and R_d reside on the microbridge, while R_1, R_2, R_A, R_B, and R_C reside on the silicon chip substrate, but are not shown in **Figures 1(b)** and **3(a)**. R_A and R_B may even be of zero-thermal conductivity detector (TCR) and positioned off the chip. R_C is preferably of zero or low TCR. (a) Heater control to maintain the temperature rise of R_h. (Source: Bonne U, Kubisiak D, Creasy K, Francisco T 2001a Rugged microsensor for process stream composition monitoring. *15th Int. Forum on Process Analytical Chemistry (IFPAC)*, Amelia Island, FL, USA, January 21–24, 2001. *J. Process Anal. Chem.* **7**, 1 (2001–2002); reproduced with permission.) (b) Flow-sensing bridge for upstream and downstream sensor elements R_u and R_d.

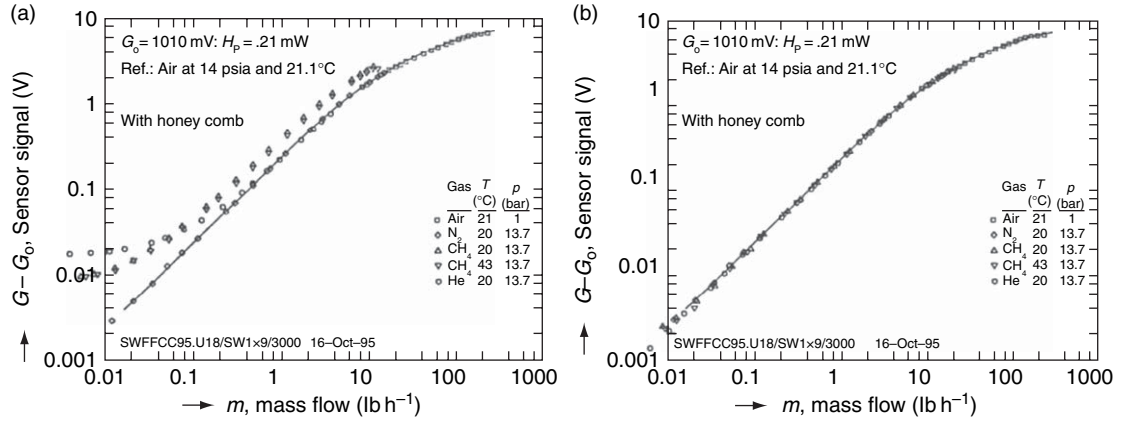


Figure 3 Microbridge flow sensor responses and composition correction to four different gases: (a) before and (b) after composition correction of mass flow. (Source: Bonne U 1996a Sensing fuel properties with thermal microsensors. Proceedings: *SPIE Smart Electronics and MEMS Conference*, San Diego, CA, USA, February 25–29, 1996, Paper No. 2722-24, p. 165; reproduced with permission.)

resistance (TCR) resistor R_c compensates for the temperature influence on the value of Pt coefficient of resistance (TCR), i.e., the nonzero value of α_2, so that the flow or thermal conductivity outputs are largely temperature-compensated. Such flow sensor chips offer a large dynamic flow range (~3000:1), a short response time (~1 ms), low power (0.1–10 mW), chip-to-chip interchangeability within 5%, wide temperature range (−60°C to 160°C), good stability (drift <50 ppm after dozens of −20°C to 120°C thermal cycles) (Bonne 1992, 1996), and low chip manufacturing cost, but have limitations imposed by packaging materials and electronics.

It is this basic MB structure (Higashi and Johnson 1987) (see **Figures 1** and **5**) that spawned a number of sensor ideas, research efforts and productized sensors: of physical gas properties (Bonne 1992, 1994, Bonne and Kubisiak 2001a), uncooled optical radiation

(especially IR) bolometers (Wilson *et al.* 1991), flow or wind direction (Bonne *et al.* 2003a, Djorup 1993, Huising *et al.* 1993), gas composition (several) (Bonne 1994b, Bonne *et al.* 2000a, Cole *et al.* 1986, Schierbaum *et al.* 1991, Zanini *et al.* 1995), tilt and acceleration (MEMSIC Inc., Zhao and Cai 2005), and vibratory sensors or actuation-based sensors (Bonne and Kubisiak 2001a). The following are the core features that enabled the fabrication of the above sensors:

(1) Measuring thermal conductivity, specific heat, and temperature and correlating these with pressure (see point (3)), gas density, viscosity, heating value, supercompressibility, and octane number (Bonne 1991–1993). Others leveraged the microthermal gradients to make tilt, level, acceleration (MEMSIC Inc., Zhao and Cai 2005), or rotation sensors.
(2) Depositing special coatings on the heatable MB (or microhotplate (Cavicchi *et al.* 2000), such as metal oxides (SnO_2, WO_3, Pt- or In-doped SnO_2, etc.) to sense NO_x, CO, and combustible gases nonspecifically (Boger *et al.* 2005, Cavicchi *et al.* 2000, Dable *et al.*, 2004, Kunt *et al.* 2000, Meier *et al.* 2004, Tiffany *et al.* 2000, Visser *et al.* 1994)), hydrated LiCl to sense dew point (Bonne *et al.* 1996), and yttria-stabilized zirconia (YSZ) to sense O_2 (Cole *et al.* 1986).
(3) Making the MB structures either so small that they become commensurate with the mean free path of ambient air molecules, so that the measurable heater power becomes pressure dependent (Bonne and Kubisiak 1994, James *et al.* 1988, Priti *et al.* 1993), or sensitive enough to respond to the small, pressure-induced changes in thermal conductivity (Bonne and Kubisiak 1994, Priti *et al.* 1993).
(4) Exploiting various effects obtained by exposing such thermal microflow sensors to oscillatory flow induced by off-the-shelf miniactuators such as microphone speakers (Bonne *et al.* 2001a, Bonne *et al.* 2001), or to electronic resonance effects associated with their circuits (Bonne and Kubisiak 2001b).

2.11.2.2 Sensing Gas Flow, Thermal Conductivity, *k*, Specific Heat, c_p, and Temperature

In describing the above except for point (2), to which we will come back to in Section 2.11.3, we will now focus on various gas sensor concepts based on MB and Microbrick™ structures (Padmanabhan *et al.* 2006), whereby the heater and sensing thin-film film resistors of the latter are supported on bulk solid insulators. Although consuming a little more power, MK flow sensitivity (S/N ratio) is higher and the response time is shorter.

But it was the shift in thermal flow sensor output signal, as shown in **Figure 3(a)** (and more dramatically also in **Figure 15**), that induced the effort to find ways to correct for it, and in the process led to further ideas on how to leverage the gained insights into additional uses of such sensors. **Figure 3** shows how the flow signal curves shift (without a change in shape; at low flows it is important to properly determine the offset, G_o) with gas composition, which provided an early clue about the existence and mathematical structure of flow-independent composition correction factors. Solving this problem was of fundamental interest to thermal anemometry and especially if we were to use thermal microanemometers for metering natural gas, the composition of which changes enough with time to cause uncorrected metering errors of up to 10%. This effort coincided in time (1985–2000) with a growing interest by the natural gas distribution industry to modernize traditional bellows-based gas meters with electronic and wireless ones (Bonne 1991, 1992, Bonne *et al.* 1990, Visser *et al.* 1994). Early researchers had tried to correlate physical gas properties to heating value, to avoid destructive and bulky calorimetry (e.g., correlations with dielectric constant and density, thermal conductivity, speed of sound, and viscosity (Altemark *et al.* 1979, Durst *et al.* 1999, Sommers *et al.* 1979)), but did not make use of specific heat or try to use the correlations to correct flow sensor outputs, because electronic gas meters were still in their infancy.

The composition correction factor was defined as C_V = (uncorrected flow/corrected flow), where for practical reasons, flow here means standard volume flow. C_V had to meet a number of expectations:

- Be pressure independent to avoid the need for pressure sensors. The pressure independence of thermal flow sensors, was experimentally validated and holds for constant gas composition (Bonne *et al.* 1995a), but does not hold when composition is varied.
- Revert to a value of unity for flow sensing with the reference or calibration gas.
- Show little or no temperature sensitivity, in view of the inherent temperature compensation built into available flow sensors (Catalog #15 1998).

- One correction constant had to be valid over the full dynamic flow range in view of the uncovered simple shift of the flow curves for different gases, when plotted on a log–log scale.

The form of the empirically correlated composition correction factor, C_V, meets the above expectations, with even only the first two terms in:

$$C_V = \left(\frac{k}{k_o}\right)^{-0.8827} \left(\frac{C_p}{C_{po}}\right)^{1.110} \left(\frac{\mathrm{Pr}}{\mathrm{Pr}_o}\right)^{-0.333} \quad [1]$$

The subscript o indicates the values for the reference gas. A theoretical foundation and derivation for eqn 1] was provided later (Bonne 1990–1992), which rationalized the need for the third but weak term involving the Prandtl number. The physical properties in all three terms are fundamentally independent of pressure, but not of temperature. By way of example, if one defines $C_V = 1$ for N_2, then $C_V = 2.12$ for ethane (C_2H_6), despite having close to the same molecular weight (MW) and thus density. **Figure 3(b)** shows the result of applying composition correction, which was valid for more than four orders of magnitude, for gases with vastly different properties as in the case of helium (He), methane (CH_4), nitrogen (N_2), and C_2H_6.

After uncovering and demonstrating thermal flow sensor composition correction based on the measurement of k and c_p, preferably at two different temperatures (Bonne 1990–1992, 1994a, 1996, Bonne *et al.* 1990), the question about accuracy came to the fore (Frampton *et al.* 1996) and how large a S/N could be achieved, which prompted the development of a standardized approach to visualize S/N and short-term drift and temperature correction, which is discussed in detail in Section 2.11.4. But it also prompted the question about correlation to other gas properties such as natural gas density (see **Figure 4**, demonstrating the feasibility for a good correlation of the simple type $\rho = \rho(k(T), c_p(T))$), viscosity, heating value (Bonne 1992, Bonne and Kubisiak 1996, Bonne *et al.* 1994), supercompressibility (Bonne 1993), octane number (Bonne 1993), and oxygen demand (Bonne 1996b), as well as the determination of equivalent liquid properties (Bonne 1996a). The above correlations consisted of multiterm polynomials and were in large part enabled by the high degree of orthogonality between k and c_p. Being able to predict oxygen demand of a fuel, made feedforward control of fuel/air ratio possible, so that no sensors would have to be exposed to the harsh exhaust gas environment. See also **Table 2** for such control.

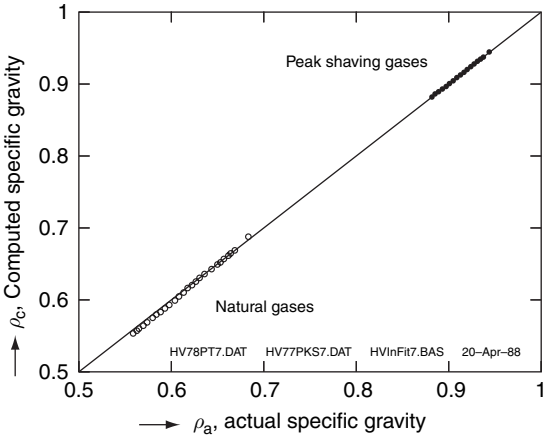


Figure 4 Microbridge chip-based k and c_p measurement and correlation with gas density, expressed in terms of specific gravity, of over 140 fuel gases (natural gases and peak shaving gases (with a 50% addition of propane and air). The standard deviation for this correlation was ±0.33%. (Source: Bonne U 1996a Sensing fuel properties with thermal microsensors. Proceedings: *SPIE Smart Electronics and MEMS Conference*, San Diego, CA, USA, February 25–29, 1996, Paper No. 2722-24, p. 165; reproduced with permission. Bonne U 1994 Thermal microsensors for environmental and industrial controls. *NIST Workshop on Gas Sensors: Strategy for Future Growth*, Gaithersburg, MD, USA, September 8–9, 1993, NIST Special Publication No. 865, 1994, pp. 31–8, ISSN: 1048-776X; reproduced with permission.)

2.11.2.3 Sensing Pressure with Thermal Microsensors

Extending the state-of-the art applications of micro-Pirani pressure sensors (James *et al.* 1988, Mastrangelo and Muller 1991, Priti *et al.* 1993, Bonne and Kubisiak 1994) Bonne and Kubisiak:

(1) Demonstrated an operating range of thermal conductivity-based pressure sensors (Pirani gauges) to above 200 bar.
(2) Corrected their output for the influence of changes in gas composition and temperature.
(3) Demonstrated repeatabilities and short-term S/N comparable to the state-of-the-art precision (±0/1%) industrial pressure sensors (Bonne and Kubisiak 1994), although their output sensitivity to drift in the used k and c_p measurements was higher.

The used sensor chips (see **Figure 1(c)**) were special in the sense of being endowed with a micro-Pirani pressure-sensing structure used earlier (James *et al.* 1988; see **Figure 7**), and not used in Honeywell's commercial MB sensor chips (Catalog #15 1998).

Figure 9 shows the measured output of the small Pirani microsensor of Figure 7 over a pressure range spanning more than six orders of magnitude, in N_2. The drop in pressure sensitivity to be less than a millibar is associated with the larger than 100-μm mean free path, being much larger than the dimensions of the thermal conductivity sensing structure. Larger structures are known to measure down to lower pressures. The still changing output above 1 bar is what caught our attention as being a novel feature, which we attribute to the nonideal gas changes in thermal conductivity, which may be used for making burst-proof pressure sensors in the future.

The online composition corrections of these thermal pressure sensors were made via thermal measurements with two MB sensors of different dimensions (see Figure 1(c), *upper left corner*, and its magnified view in Figure 7) so that the effect of absolute thermal conductivity changes are sensed and effectively eliminated. The required measurements include temperature, heater conditions (in gas or vaccum), and steady state and transient heat transfer conditions, and were performed with single 1.7 mm × 1.7 mm, Si chip sensors machined from single Si wafers, as depicted in Figure 1(c). Figure 8 shows the measurement data before and after applying the composition correction, as well as the used expression for the correction. The values of the involved constants were $Q_{vac} = 0.33$ mW, i.e., the measured residual heat loss in vacuum, $k_{air} = 68.255$ μcal (cm K s)$^{-1}$, $A_1 = 0.64998$, $A_2 = 1.291638 \times 10^{-6}$ psi cm, $A_3 = 0.65996$, and $A_4 = 2.741976 \times 10^{-7}$ psi cm. The remaining or unknown gas parameters are k_i = thermal conductivity of gas i in μcal (cm K s)$^{-1}$ and L_i = its mean free path in centimeter. By sensing Q we could determine the pressure, p (in psia). In the expression in Figure 8, k_i thus defines the heat loss at high pressure and L_i the transition region from the Q_{vac}. The second exponential term was needed to improve the quality of the curve fits, which was not deemed surprising in view of the odd geometry of the narrow MB structure.

To visualize the achieved S/N of the MB sensor graphically and compare it to pressure sensors available commercially, Figure 10 shows measured data of both the former and the latter (for which we chose an available Honeywell ST3000 piezoresistive pressure sensor), labeled P_{MB} and P_{ST3000}, respectively. Note that the log(y_i)-axis data were plotted after increasing their magnitude 20-fold, i.e., $y_i = C_i(P_i/P_1)^{20}$, so that a 1% noise level would correspond to the shown scale bar. The C_i are simply spacing constants to separate the individual parameter traces on the chart. As shown, the peak–peak noise levels differ by only a factor of about 2. Section 2.11.4 provides further discussion about such visualization plots, which we found very revealing and educational.

2.11.2.4 Actuation-Based and Self-Oscillating Sensors

The original motivation for exploring the behavior of small and constant amplitude AC-modulated flow was to create repeatable, oscillating, volumetric flow signals at the prevailing temperature and fluid pressure, so that these signals would reflect the characteristic of the fluid, without being influenced by the main flow, and enable one to use such signals to correct flow sensor output errors resulting from changes in gas composition.

The setup depicted in Figure 11 to generate AC-modulated or reciprocating flow at the second or normalizing sensor at first used a sawtooth drive generated by a Hewlett-Packard HP8116A signal generator. The flow sensor amplifier with filter was a Model 113 by Princeton Applied Research. A key requirement is that both flow sensors be made as equal as possible.

Exposing thermal flow microsensors to oscillatory or reciprocating flow with the above setup enabled us to demonstrate a number of exciting new sensing approaches and characterizations, as follows (Bonne and Kubisiak 2001a):

- Submillisecond flow sensor response time versus flow velocity and microsensor structure.
- Compact and affordable composition correction, C_V, for volumetric fluid flow sensors, without the need to sense k or c_p.
- Concentration of binary mixtures, based on measurement of C_V.
- Fluid properties based on actuator-induced flow or compression, such as viscosity or $\gamma = c_p/c_v$, respectively. The determination of $\gamma = c_p/c_v$ from the induced temperature changes due to adiabatic compression of the gas in a closed cavity, as is discussed in a later publication.

The used micromachined thermal flow sensor chips consisted of either:

(1) Off-the-shelf, front-etched MB sensor chips of ~1.7 mm × 1.7 mm, with bridges of ~0.2 mm × 0.25 mm as depicted in **Figure 1(a)**. (Figure 1(c) has additional features, such as the pressure sensor, not included in the photo of the commercial version). (Higashi and Johnson 1987, Higashi *et al.* 1986, Honami *et al.* 1999).

(2) Developmental micromembrane sensor chips, similar to MB sensors except for etching away the silicon under the sensing membrane from the backside of the chip, rather than through the openings at the front side. The gases and vapors to be sensed then had access only to the front side, which would therefore prevent the problem we had seen with occasional gasoline vapor condensates and with breakage of MBs resulting from excessively high mass flux, e.g., over 30 g $(s\ cm)^{-1}$. In addition, such membrane-based sensors would burst at differential pressures between 50 and 100 psid, and exhibit undesirable stress-effect errors caused by changes in differential fluid pressure. Such effects were ~60× larger than the small pressure dependence of the thermal conductivity of air, which is only 0.18% bar^{-1}. We therefore decided to proceed only with sensors discussed in (1) and (3).

(3) Developmental, very rugged, Microbrick sensor chips of equal size but without etched cavities (see top of **Figure 5** and **Figure 11(b)**), so that the thin-film resistors are supported by a polymer (e.g., carefully back-filled epoxy) or glass (Pyrex or quartz). The purpose of using such sensor chips was to make high-performance sensors also available for flow and property measurements in applications involving harsh environments such as condensing vapors or liquids, high mass flux process streams and those carrying dust particles. Condensates collected under the MB microspaces are extremely hard to remove even by raising the microheater temperature by an extra 80°C. Moderate mass flux conditions in either gases or liquids (>5 g $cm^{-2}\ s^{-1}$), especially when associated with entrained dust (sand or rust particles of up to 0.3 mm diameter), have broken such sensors, so that reliable operation in industrial applications with gas or water mass fluxes of up to 80 or 500 g $cm^{-2}\ s^{-1}$, respectively, would have been short-lived. The concern that such Microbrick sensors would trade ruggedness for sensitivity and speed was found to be unsubstantiated. **Table 1** lists further design and performance data of Microbrick sensors. They are clearly burst-proof and designed for operation in harsh environments, and they operate with gas or liquid mass fluxes up to 500 g $cm^{-2}\ s^{-1}$, with condensable vapors and suspended sand or dust (they survived MIL Spec ice storm and sandstorm tests).

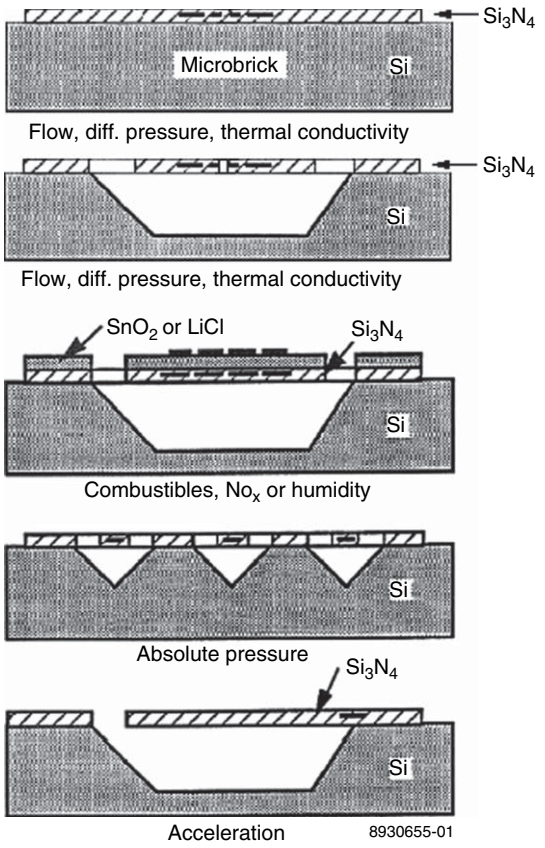


Figure 5 On the versatility and flexibility of microbridge (MB)/Microbrick structure sensors. (Source: Bonne U 1994b Thermal microsensors for environmental and industrial controls. *NIST Workshop on Gas Sensors: Strategy for Future Growth,* Gaithersburg, MD, USA, September 8–9, 1993, NIST Special Publication No. 865, 1994, pp. 31–8, ISSN: 1048-776X; reproduced with permission. Bonne U, Kubisiak D 1994 Overpressure-proof, thermal pressure sensor for gases. *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, June 13–16, 1994, Tech. Digest, p. 76; © 1994 reproduced with permission from IEEE. Bonne U 1998 Microsensor fabrication and application. *11th IGT Symp. Gas Quality and Energy Measurement*, St. Petersburg, FL, USA, February 2–4, 1998, Proceedings; © 1998 reproduced with permission from Gas Technology Institute.)

Figure 6(b) shows a comparison of flow responses between a MB and an early Microbrick sensor, via measurements obtained before and after backfilling the etch pit of the same MB sensor with epoxy. In this way, small chip–chip differences were eliminated, and the observed shift was solely due to the change in heat transfer characteristics. For this comparison, the heater temperature rise above ambient was kept at ~30°C by suitable adjustment of the heater control circuit. **Figure 6** shows that the signal noise, as

Table 1 Microbrick™ fluid flow and property sensor performance data

Mechanical/packaging	
Sensor chip	FWB Microbrick-II, mask #0857A
Primary package	Chip on Au-plated TO18 header with glass cover seal; 0.168″ OD platform
Electronics	
Input heater power	$\sim 2^{\circ}C\,mW^{-1}$, also known as thermal efficiency
Environmental	
Approvals	Microbridge chip approved for operation in flammable gas mixtures[a], no UL, CSA, or PTB approvals are in place for Microbrick sensors
Temperature range	$-40^{\circ}C$ to $150^{\circ}C$ for chip and substrate
Heater temperature rise	$\sim 6^{\circ}C$ to $140^{\circ}C$, depending on application and fluid phase (liquid or gas)
Sample fluid pressure range	No limitation
Chip protection from	
Flow and particulates	None. Passed Mil-Std-810e Sand Blast Test
Performance	
Sensitivity	
Flow	0.8 V from 0 to 100 m s^{-1}, for $\Delta T_{htr} = 35^{\circ}C$
TC	*~1.5 V for change from methanol to water, ~1.230 V for air to n-butane*
Noise (short term)	$<\pm 1.6\,mV_{RMS}$
Flow range	0.2–200 m s^{-1} in air
Drift	≤400 ppm in 10 years or after >1 million thermal cycles from $-40^{\circ}C$ to $120^{\circ}C$
Chip response time (63%)	<1–0.2 ms in air; $\tau_{electronic}$<4.8 ms
Signal stability (total)	
Versus ambient temperature	With compensation: $\leq 0.03\,mV^{\circ}C^{-1}$ from $0^{\circ}C$ to $60^{\circ}C$ ambient temperature
Versus sample liquid pressure	*$-0.25\,mV\,bar^{-1}$ (0.017 mV psi^{-1}) in 100% methanol (natural TC versus pressure), for 8–56 bar (100–800 psig), 20°C*
	$-5.3\,mV\,bar^{-1}$, a decrease in line with natural TC increase of air with pressure
Versus input voltage	*≤5 mV V^{-1} in air, from 15 to 22 V*
Versus vibration	*≤5 mV for 1–100 g at 600–6000 RPM. ≤22 mV g^{-1} for ±1 g in methanol*
	≤1.3 mV g^{-1} for ±1 g in air

[a]Certificate No. BVS 94.C. 7001 U, for Honeywell Flow Sensor Type AWM 2***V, DMT-Gesellschaft für Forschung und Prüfung mbH, Bergbau-Versuchsstrecke, Dortmund, Germany, July 15, 1994. The approval is for operation in continuously (=Zone 0) explosive environments (even in H_2–air), while powered with <26.9 mW at 25°C or <22.2 mW at 65°C. The approval type EEx Ia I/II C means: (E)uropean (Ex)plosion protection (I)ntrinsic safety; a, category that allows two faults in the circuit; I, mining/II, industrial application; C, applicable for gases with quenching distances of <0.5 mm (CH_4–air being about 2 mm and ignition current of <0.45× smaller than that of CH_4.
Source: Bonne and Kubisiak (2001a); © 2001 reproduced with permission from IOP
Data in italics are for performance as property (thermal conductivity) sensor only.

indicated by the 1-sigma error bars, is ~5–20% lower for the Microbrick. In view of its faster response and thus its expected greater sensitivity to microturbulence, we are still searching for an explanation of this beneficial effect.

For actuators, we used commercially available, 10- to 12-mm OD, membrane-based, low-cost, solenoid earphone speakers, with resonances in the 2-kHz region (International Components Co.). The displacement of such membranes was measured under a microscope and provided a surprisingly linear response under DC conditions, with center displacements on the order of 0.01 mm V^{-1} (or ~0.001 mm V^{-1} for the lead zirconate titanate (PZT)-based versions), as reported earlier (see data in **Figures 7 and 8** of Bonne and Kubisiak 2001a).

Two sizes of compensated flow sensor packages were designed and fabricated, with maximum main flows of 1 and 200 L min^{-1}, featuring flow channel cross sections of 0.015 and 1.77 cm^2, respectively. They hold the two flow sensors and the actuator membrane, as indicated in the conceptual sketch of **Figure 11**. Actually, all four sensors were off-the-shelf AWM4300 flow sensor packages rated to operate at a maximum nominal flow of 1 L min^{-1} (Catalog #15 1998), except that in the nominal 200 L min^{-1} version, the main flow sensor was mounted in a bypass across the honeycomb flow straightener, all housed in a Delrin polymer package, as shown in **Figure 12**, in comparison with the much bulkier "k and c_p" version. For the final, built-in electronics for the determination of C_V, we switched the solenoid drive from a sawtooth to a sine wave (and monitored the root mean square (RMS) sensor output rather than the peak–peak output) because of the ~10× reduction in the noise level.

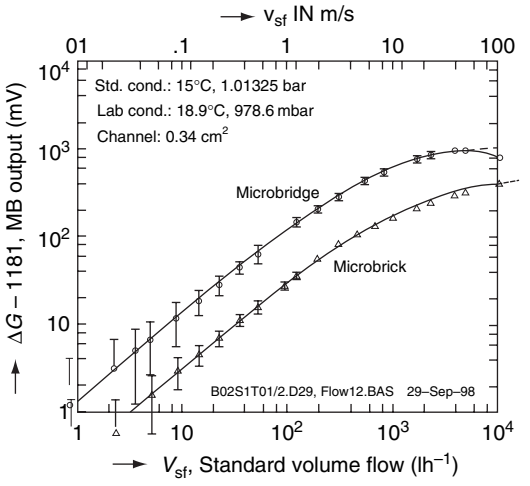


Figure 6 Comparison of microstructure flow sensor outputs before and after backfilling the MB with epoxy, to make an early version of a Microbrick. (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185.)

Speed of response: The useful operating frequency range of both sensors and actuators, with due consideration to resonance effects, was in the 40- to 100-Hz range and the one most free of disturbances for the used actuators (Bonne and Kubisiak 2001a). The flow sensors themselves showed the capability to operate beyond 500 Hz, especially the faster and more rugged Microbrick version, which showed response times down to ~0.2 ms.

Figures 13 and **14** show the results of measuring the speed of response of the above-mentioned MB and Microbrick sensors. The first experiment was designed to determine the response time of the sensor, in order to define a frequency range of operation. **Figure 13(a)** and **13(b)** shows two examples: with triangle wave inputs of 6 V_{p-p} to generate square wave AC flows of the indicated frequencies we obtained the signals shown, from which we then derived the response times, i.e., the times needed to achieve 63.2% of the final signal value.

Repeating such measurements at different frequencies resulted in the plotted response time and sensor output signal data in **Figure 14**, for both MB and Microbrick sensors. These sensors were mounted in standard AWM43000 packages (see **Figure 11(d)**), which feature flow channels of 0.5 mm × 3 mm in cross section at the sensor chip, which resulted in the linear gas velocities shown on the top *x*-axis. (The bottom *x*-axis shows the flow-generating actuator frequency, from which the top-axis velocity past the normalizing sensor was calculated, based on the known actuator displacement.) At 1 m s^{-1} on that scale in **Figure 14**, for example, the 63.2% response times are $\tau \sim 0.8$ and 0.3 ms, and at 3 m s^{-1} about 0.4 and 0.2 ms, for the MB and Microbrick chips, respectively. The decreases in response times with increasing velocities, are in line with increased convective heat transfer and stronger thermal coupling. We interpreted the reason for the shorter response times of the Microbrick versus the MB to the stronger thermal coupling between the heater and the upstream and downstream sensing elements (see the top two cross-sectional sketches in **Figure 5**). The strength of such coupling also influences the shape of the sensor output curves at very high flows. As **Figure 6** indicates, the MB signal peaks near 30 m s^{-1} (4000 l h^{-1} in the 0.34 cm^2 flow channel) and decreases after that due to the insufficient heat flux in the thin membrane to maintain the needed temperature of the sensing elements. Such limiting heat flux may not occur in the structure of

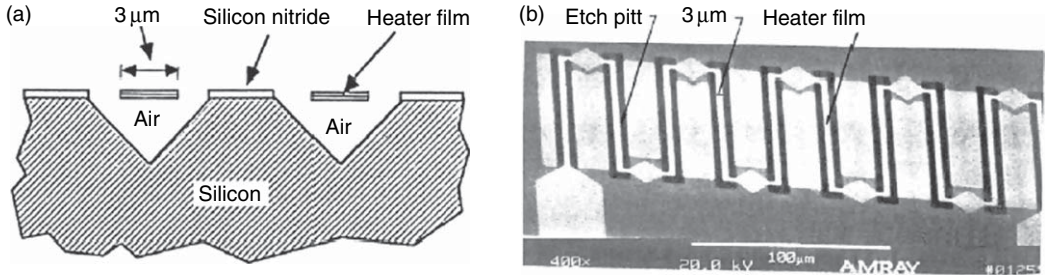


Figure 7 Micropressure sensor based on thermal conductivity effects across structure sizes commensurate with mean free path: (a) Cross section and (b) scanning electron micrograph (SEM) showing gaps of 3 µm. (Source: James S D, Johnson R G, Higashi R E 1988 A broad-range, absolute pressure microsensor. *IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head, SC, USA, June 6–9, 1988, Book of Abstracts, p. 107; Bonne U, Kubisiak D 1994 Overpressure-proof, thermal pressure sensor for gases. *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, June 13–16. 1994, Tech. Digest, p. 76; © 1994 reproduced with permission from IEEE.)

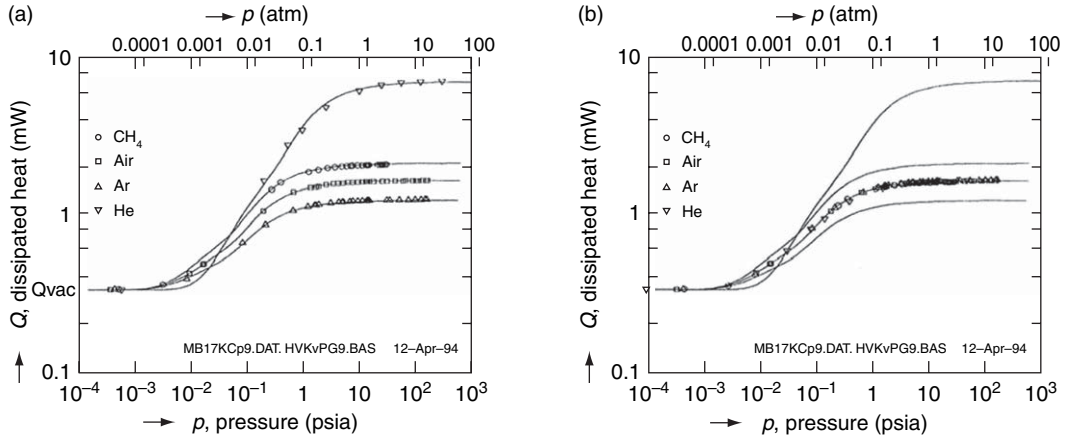


Figure 8 Heat dissipated by a wide MB pressure sensor heater for four gases, without measurable flow (Source: Bonne U, Kubisiak D 1994 Overpressure-proof, thermal pressure sensor for gases. *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, June 13-16, 1994, Tech. Digest, p. 76; © 1994 reproduced with permission from IEEE.) (a) Before and (b) after composition correction, with fitted functions of the form:

$$Q = \frac{Q_{\text{vac}} + k_j}{k_{\text{air}}\left\{A_1 \exp\left(-\frac{A_2}{p/L_i}\right) + A_3 \exp\left(-\left[\frac{A_4}{p/L_i}\right]^{0.5}\right)\right\}}$$

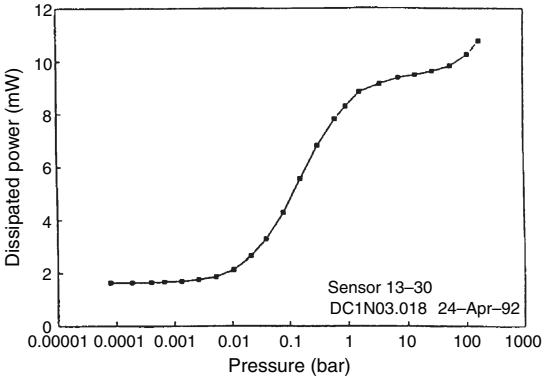


Figure 9 Heat dissipated by a Pirani-type, narrow MB pressure sensor, up to high pressure, before linearization, in N_2. (Source: Bonne U, Kubisiak D 1994 Overpressure-proof, thermal pressure sensor for gases. *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, June 13-16, 1994, Tech. Digest, p. 76; © 1994 reproduced with permission from IEEE.)

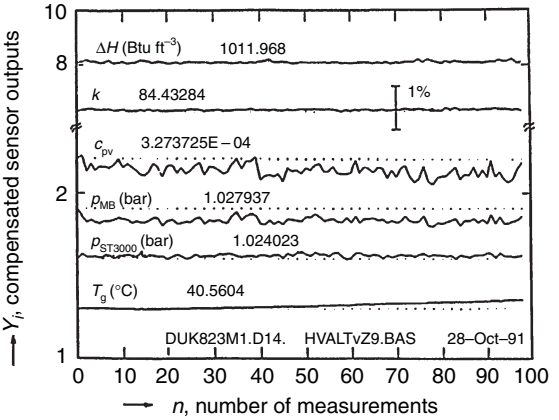


Figure 10 Short-term repeatability of thermal pressure MB sensor measurements, including its compensated output, P_{MB}, compared to an industrial-grade Si piezoresistive sensor, PST_{3000}, and to additional MB outputs, for CH_4; *y*-scale: 20×. (Source: Bonne U, Kubisiak D 1994 Overpressure-proof, thermal pressure sensor for gases. *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, June 13–16, 1994, Tech. Digest, p. 76; © 1994 reproduced with permission from IEEE.)

the Microbrick sensor, which in this case was the identical MB chip, after it had been ruggedized by filling epoxy under the bridge, the output of which is also shown in **Figure 6** and seen to be shifted to higher flows by a factor of about 4×, after subtracting out the 1181 mV zero-flow offset. This offset is purposely set by the circuit, so that small negative flows can be observed as well. For additional performance parameters of the Microbrick sensor chip, see **Table 1**.

In summary, the above response time measurements have taught us that we can safely use normalizing actuator frequencies, f, of up to $f = 1/(2\pi\tau)$, which, with conservative values of $\tau = 0.8$ and 0.3 ms amount to 200 and 530 Hz for the MB and Microbrick, respectively, provided no other effects such as resonances limit such measurements.

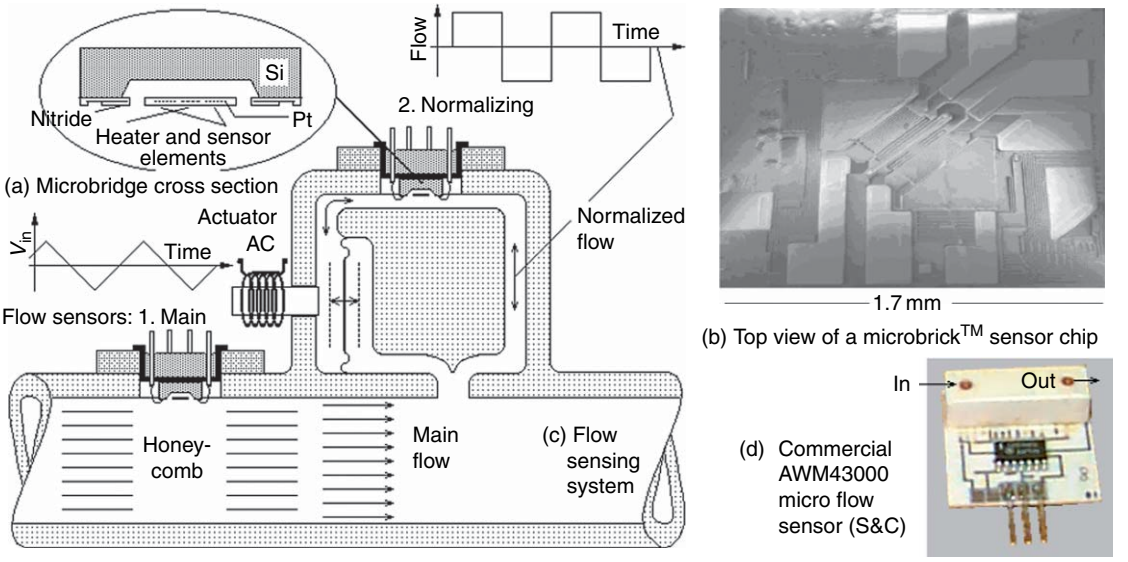


Figure 11 Functional sketch of composition-compensated flow-sensing system. The use of the shown sensor mounting (on TO5 header or on alumina circuit board with flow channel (as in (d)) and thermal flow microsensors is only exemplary. Other types of flow sensors may be used as well. (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission. Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *Proceedings GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002, © 2002 reproduced with permission from Gas Technology Institute.)

Composition correction factor, C_V: Having demonstrated the ability to measure the small reciprocating flows generated by small membrane actuators, and assuming that the actuator displacement is:

- stable over time, as one is led to believe, given its rugged iron membrane set in motion by the electromagnetic (EM) field (see further details in the discussion on viscosity sensing).
- independent of changes in gas density, pressure, or viscosity.
- generating sensor output signals that are proportional to actuator frequency, to indicate that the actuator is not near conditions or subject to signals distorted by resonance frequency effects as proven in Bonne and Kubisiak (2001a).
- generating sensor output signals that are proportional to absolute pressure and temperature-dependent gas density (because thermal flow sensors are accurate mass flow sensors), and
- generating sensor outputs proportional to the individual gas properties that come into play to cause flow errors due to gas composition changes.

The reciprocating flow sensor signal is then proportional to the composition correction factor, C_V, described in Section 2.11.2.2. This means that C_V can now be directly measured, without the need for the more cumbersome measurement of k and c_p described in Section 2.11.2.2. Furthermore, this general approach is not limited to thermal mass flow sensors. However, for each new sensor one needs to determine whether the used flow sensor is a volumetric or mass flow sensor, so that the pressure and temperature corrections can be applied as appropriate.

The measured data shown in **Figure 15** was also obtained with the setup of **Figure 11**, with N_2 as reference gas and C_2H_6 and He as test gases. The shifts in output for these gases would be much smaller if plotted versus mass flow, but still too large (~100% for C_2H_6 versus N_2, see **Figure 15**) to be negligible. Because of the mentioned 10-fold decrease in signal noise for sine wave versus sawtooth operation, we continued work on flow normalization and composition correction (but not on response time or viscosity measurements) with the former. Additional influences on the quality of the obtained results are as follows:

(1) Proximity of the chosen frequency to the 50- or 60-Hz line frequency, which influences the amount of undesirable pickup.
(2) Drive amplitude, which influences distortion. **Figure 15** was obtained at 60-Hz sine wave and $3V_{p-p}$.
(3) Proximity of the chosen frequency to any resonance frequencies of the speaker membranes.

Figure 12 Comparison of composition-compensated thermal flow sensors, based on sensing: *left*: C_V via actuation-based flow, for maximum flows of 200 (*top*) and 1 l min^{-1} (*bottom*). Right: k and c_p in a dead-ended cavity, also for 200 l min^{-1} package, with circuit-board area of 260 cm^2, compared to 12.6 cm^2 for the C_V-based measurement.

The data points by the sloped lines in **Figure 15** represent the uncorrected main flow experimental data for C_2H_6, N_2, and He, obtained with the 200 l min^{-1} package. The sloped lines represent uncorrected flow data predicted theoretically, relative to reference gas N_2, for which we chose to set the volumetric composition correction factor, $C_V = 1$. As shown, they agree to within ±3%, even before a large effort is made to reduce this number further.

The arrays of approximately constant $\Delta G'$ versus flow data points represent the normalizing sensor outputs, which should not be affected by the main flow, as demonstrated, until turbulence at high flows does seem to exert some influence. C_V now results from a simple ratio:)

$$C_V = \Delta G'_{\text{norm,RMS}} / \Delta G'_{\text{norm,RMS,N}_2} \qquad [2]$$

where $\Delta G'_{\text{norm}}$, RMS, $N_{2=\ 48.5\ \text{mV}}$ as plotted, and represents the nulled signal, i.e., the measured signal, ΔG, without its zero-flow offset, ΔG_o. $\Delta G' = \Delta G - \Delta G_o$. The obtained values of C_V are in good agreement with those "obtained via the "k and c_p" approach" and predicted theoretically: numerically, the actuator-based values for C_V are 1.0 (by definition), 2.1, and 0.112, respectively, while the "theoretical, "k- and c_p"-based" approach according to eqn [3] yielded 1.0, 2.2, and 0.105, respectively. For gases of known k and c_p, the computation of C_V via eqn [3], derived in Section 2.11.2.2, is easy:

$$C_V = C_V = \left(\frac{k}{k_o}\right)^{-0.8827} \left(\frac{C_p}{C_{po}}\right)^{1.110} \left(\frac{\text{Pr}}{\text{Pr}_o}\right)^{-0.333} \left(\frac{T}{T_o}\right)^{m} \qquad [3]$$

but more cumbersome if the k and c_p have to be measured. Measurement of the Prandtl number, Pr, could generally be omitted because of its small

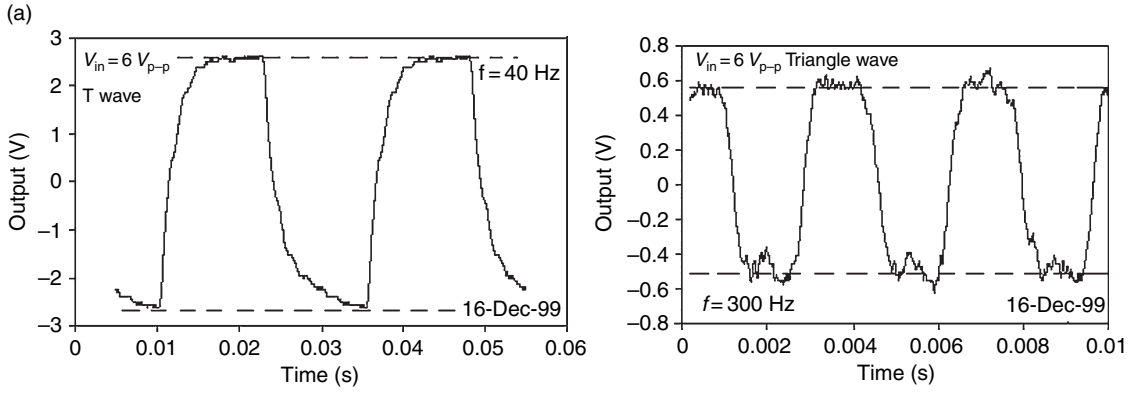


Figure 13 Responses to reciprocating flow generated by triangle wave drive to the earphone speaker, which in turn generates square wave AC flow: (a) microbridge at 40 Hz and (b) Microbrick at 300 Hz (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission. Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *Proceedings GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002, © 2002 reproduced with permission from Gas Technology Institute.)

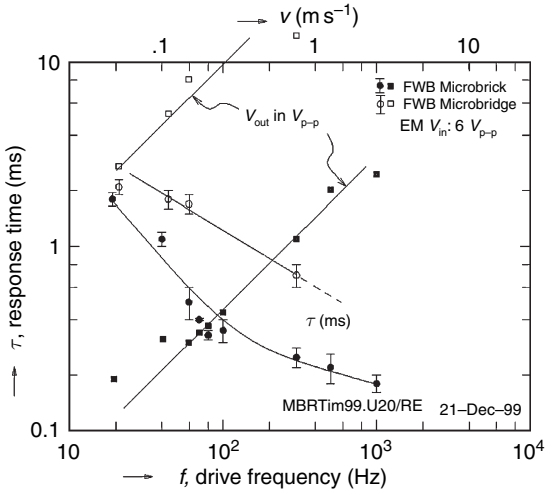


Figure 14 Flow sensor response times measured with the setup of **Figure 11**, using various actuator (earphone) frequencies. For the constant displacements, increasing frequency entails higher flow velocities as indicated on the top scale. (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission. Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002, Proceedings; © 2002 reproduced with permission from Gas Technology Institute.)

deviation from $Pr = 0.71$ for a range of gases and because of the smallness of the exponent. The temperature term serves to compensate for a combination of sensor material (Pt in our case) and electronic temperature effects, and may serve to improve accuracies beyond about 1% of reading. Such temperature compensation may need to be added to eqn [1] as well.

To summarize, this new volumetric flow sensor normalization approach:

(1) No longer requires the much more costly measurement of thermal conductivity and specific heat to determine C_V, discussed above, but requires the measurement of absolute temperature and pressure if mass flow normalization is called for; ideally a membrane able to displace and move the same and constant mass flow would be preferred for mass flow normalizations.

(2) The above flow normalization approach is universal and is not limited to composition correction of thermal flow sensors. Although the self-normalizing flow sensor packages in **Figures 11** and **12**, based on the above AC flow modulation, were built with thermal flow sensors (**Figure 11(d)**), other types of flow sensors can be used, such as Δp-based, whereby the two sensors used for the main and normalizing flows and their flow channels would be as similar as possible.

To obtain compensated standard volume flow sensor signals, additional gas pressure compensation is needed, if the actuator is of the constant volume flow displacement type. For compensated mass flow sensing, a constant mass (rather than the above constant volume) flow displacement actuator is needed, which may be a bit more difficult.

Composition sensing via C_V: Easy measurability of C_V begs the question of whether this parameter has any additional advantageous applications. As the parametric comparison in **Figure 16** shows for a binary mixture of CH_4 and C_2H_6, C_V stands out among the other plotted properties (k, c_p, D_T, v, Pr, and M, from software by PPDS/NEL) as showing the strongest signal change versus composition of this mixture. However, no experimental data were available at the time for either gaseous or liquid mixtures. The plotted property data in **Figure 16** were normalized by dividing all data by the value of the property at 0% C_2H_6 or 100% CH_4.

One property not displayed in **Figure 16** is dynamic viscosity, which has values of 91.32 and 108.64 μP for C_2H_6 and CH_4, respectively, or a change of 19% or only slightly stronger than Prandtl number. However, viscosity has other special properties for feedforward combustion control and can be measured with actuation-based sensors, to be discussed next.

Gas viscosity sensing: The measurement of gas properties such as viscosity (Altemark *et al.* 1979, Bonne 1994, Durst *et al.* 1999, Sommers *et al.* 1979, Wakeham *et al.* 1991), density (Altemark *et al.* 1979, Sommers *et al.* 1979, Wakeham *et al.* 1991), thermal conductivity (Bonne 1992, 1994a, 1996), and specific heat (Bonne 1992, 1996) has been under study for some time in order to enable the combustionless determination of fuel gas properties such as heating value, oxygen demand, and Wobbe number via correlation with those properties. In accurate viscosity measurements via vibrating quartz crystals, wires, or Lamb waves, sensor outputs typically provide combinations of viscosity and density (Wakeham *et al.* 1991). Most importantly, such property measurements, if they are to be widely used in small combustion systems, need to be affordable in that market, i.e., manufacturable at a cost of $5–10 or less.

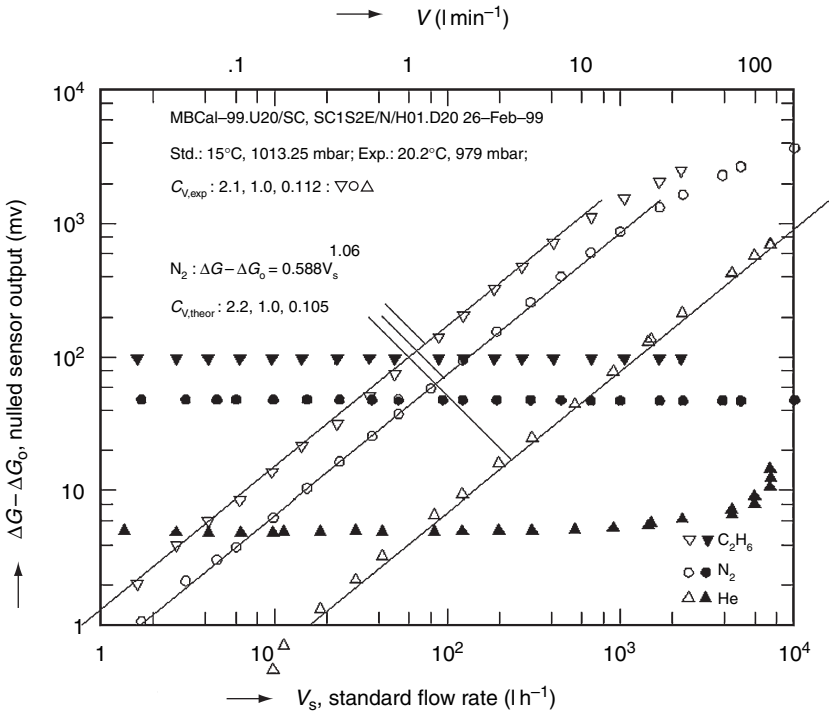


Figure 15 Demonstration of composition correction via the oscillatory sensor approach with N_2 as reference gas, C_2H_6, and He. Actuator driven with 3 V_{p-p} at 60 Hz. As intended, the oscillatory sensor signal is unaffected by the flow rate (for He up to ~1000 l h^{-1}). Measured with the 200 l min^{-1} flow sensor system. ΔG_o: 1027.6, 1029.0, and 1035.0 mV for N_2, C_2H_6, and H_2, respectively. (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission. Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *Proceedings GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002, © 2002 reproduced with permission from Gas Technology Institute.)

Such affordability had not been achieved before by the approaches based on sensing viscosity and density via flow through a capillary and an orifice, respectively, as proposed by the Ruhrgas group (Altemark *et al.* 1979, Sommers *et al.* 1979), by our own attempts based on the use of quartz crystal oscillators (Bonne 1994a), or by the specially designed micropump to drive gas through a capillary as practiced at the University of Erlangen (Durst *et al.* 1999). Our early 1997 experiments with thermal-expansion-based gas pumps or drives were encumbered by the associated thermal gas properties and long thermal response times.

Figure 17 shows the actuator-based setup 3, designed and built to sense viscosity, using the same EM speaker (see left-hand side) as in **Figure 11**, but is now used to drive fluid through either capillaries or porous frits. The resulting AC pressure signals were then measured by a differential pressure (Δp) sensor, for which we used a piezoelectric (PE) speaker (Bonne and Kubisiak 2001, 2003, 2004; see right side of **Figure 17**), because of its higher S/N ratio and lower cost when compared with silicon-based Δp sensors available at the time (the PZT actuator by itself was only \$0.12 in volumes over ~50 000 units). In order to maximize the signal, we minimized the volume of the cavity between the driver and sensor membranes shown to either side of the frit.

We had calculated that we should be able to generate AC Δp signals on the order of 0.2–2 cm WC, with the setup shown in **Figure 17**, and obtain pressure signals shaped as in the curves shown in **Figure 18**. These were obtained first via numerical integration and later analytically from Poiseuille's law of capillary flow (Bonne 1994a)

$$\frac{dV}{dt} = \frac{\pi \Delta p r_c^4}{8 L_c \eta} \qquad [4]$$

where dV/dt is the laminar volumetric flow, and Δp is the pressure drop in a capillary of radius r_c and length L_c. Because we assume the displacement rate, time ($\Delta t = 1/(2f)$, with f as the frequency), and volume

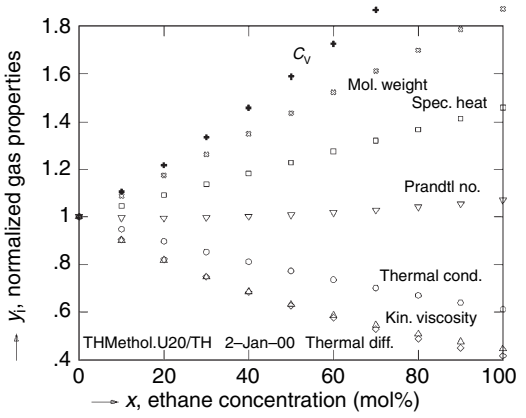


Figure 16 Comparison of relative physical property signal strength for the determination of the composition of the gas binary mixture, C_2H_6–CH_4 at 15°C and 1 atm. The large change in C_V shows it to be the most effective indicator of this mixture, whereas the Prandtl number is the parameter least affected by changes in that composition. (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission. Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002, Proceedings; © 2002 reproduced with permission from Gas Technology Institute.)

ΔV_c of the actuator for any gas to be constant, we may rewrite eqn [4] with $dV/dt = 2f\Delta V_c$ in the form

$$\eta = \frac{\pi \Delta p r_c^4}{4 \Delta V_c f L_c} \qquad [5]$$

which confirms that η is proportional to the measurable steady-state differential pressure signal. The calculated curves of **Figure 18** show that:

(1) Steady Δp values can be achieved toward the end of each sawtooth period, when flow through the capillary or frit is exactly balanced by the movement of the actuator membrane and therefore provides a direct and linear measure of viscosity.
(2) The increasing values of such steady Δp periods are proportional to the viscosities of the indicated gases C_3H_8, N_2, and Ar (83, 178, and 224 μP, respectively, at 20°C and 1 atm).
(3) The $\Delta V_c/V_c$, r_c, and L_c values need to be and can be chosen to both achieve a steady Δp period and a laminar flow in the capillary, as indicated by the Reynolds number, $Re = 2r_c v\rho/\eta < 2300$ for the gas with the lowest η/ρ ratio (ρ is the gas density), which was propane in this case.
(4) The peak output shows no significant dependence of absolute gas pressure, but the time constants to reach the steady Δp period are longer for higher-viscosity fluids and lower-pressure gases. This is consistent with an expression for this rise time constant, τ_p, which can be derived from eqn [4] for the final ambient pressure in the cavity, p_1:

$$\tau_p = \frac{8\eta L_c V_c}{\pi p_1 r_c^4} \qquad [6]$$

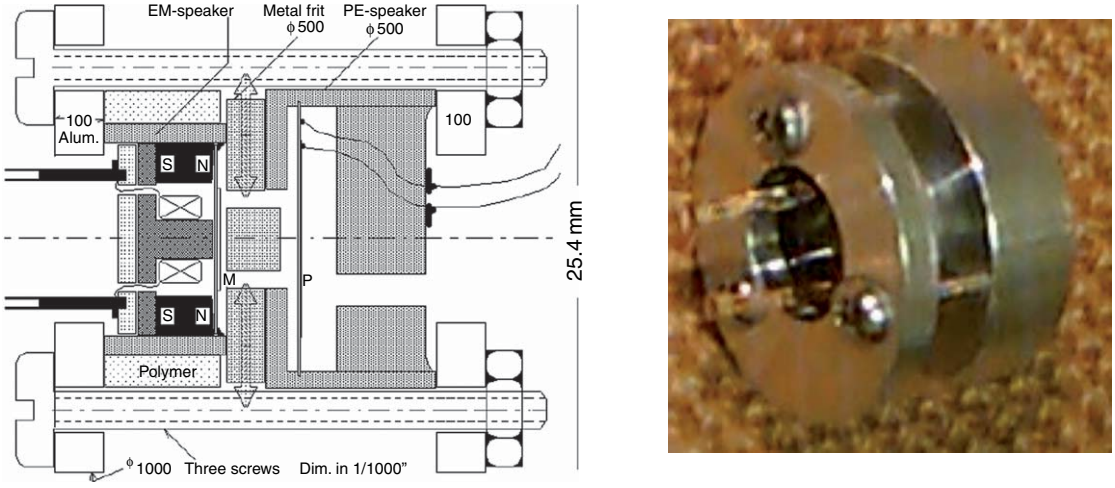


Figure 17 Gas viscosity sensor drawing and photograph, showing the electromagnetic (EM) earphone speaker membrane, M (*left*), generating pressure fluctuations sensed by the lead zirconate titanate (PZT) bimorph pressure sensor, P (*right*), while the fluctuations are partially damped by gas escaping through the metal frit pores between the speaker and the pressure sensor, as indicated by the arrows (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission. Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *Proceedings GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002, © 2002 reproduced with permission from Gas Technology Institute.)

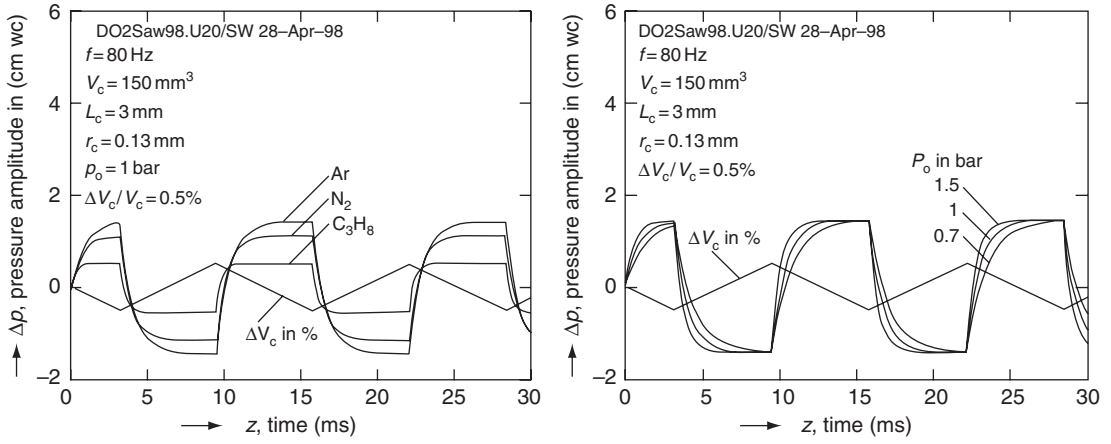


Figure 18 Calculated gas viscosity pressure signals versus time for various gases and pressures. After a few milliseconds of equilibration time, the linear change in cavity volume is balanced by the gas escaping through the frit pores via laminar flow, resulting a steady-state pressure signal, which is largely independent of gas pressure, as shown at right for three pressures of argon. (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission. Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *Proceedings GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002, © 2002 reproduced with permission from Gas Technology Institute.)

As mentioned previously, the mathematical model assumes that membrane displacement rate and stroke is constant and not affected by the relatively negligible gas mass, or by changes in gas composition or pressure. Manipulation of the parameters in the above equations taught us that for reasonable choices of f, L_c, V_c, and r_c, we can simultaneously achieve laminar flow, measurable Δp values, and fast enough time constants to fit within the linear range of the actuators.

Figure 19 presents early experimental data using absolute gas pressures from ~1 to 7 bar and gases He, N_2, CH_4, and C_3H_8. They verify the fundamental sensing approach, but also point to the fact that additional phenomena are influencing the signal: there are gas-dependent EM speaker resonances, 60-Hz line voltage pickup and frequency-dependent Δp sensor sensitivities. The storage scope data on the left side were obtained with an early setup using a combination of EM driver/0.17-mm ID capillaries/electret microphone as Δp sensor, and show significant overshoot. However, the viscosity dependence on gas type (top left set of traces in **Figure 19**) and pressure independence of the signal after stabilization (bottom left) are evident. On the right side, the setup consisted of a combination of EM driver/porous metal plate instead of the capillaries/PE speaker as Δp sensor. This porous plate or frit resulted in less overshoot and better stability at high gas pressure, and was then adopted as a preferred approach.

Choosing between capillary and frit for the laminar flow restriction is to make a trade-off between service life, sensitivity, linearity, damping, and cost, which for operation with clean gases would favor the frit. Another choice entails fastening the components of **Figue 17** with epoxy (lower cost) rather than via screw-based clamping (greater flexibility for R&D). The PE–PZT-based Δp sensor was chosen over an EM speaker or a silicon-based piezoresistive sensor due to its high immunity toward electromagnetic interference (EMI) from the solenoid driver and higher S/N ratio. **Figures 20–22** show measured Δp data at 80 Hz for several gases versus time and temperature. We found that plotting signal versus time (see **Figure 20(b)**) reveals details of a sensor's performance that would be lost by simply reporting a S/N value, and will be further discussed in Section 2.11.5. The position of the N, M, and E letterpoints included in **Figure 21** represents the temperature dependencies of viscosity expected theoretically and show that additional temperature effects, such as thermal expansion, peak actuator displacement, and electronic offset and gain, contribute to the measured dependencies.

In order to provide a user-friendly sensor output, we electronically compensated all the temperature contributions to signal error by selecting CH_4 as a reference gas and interpolating the temperature correction from a lookup table version of the experimental data (diamond-shaped points) of

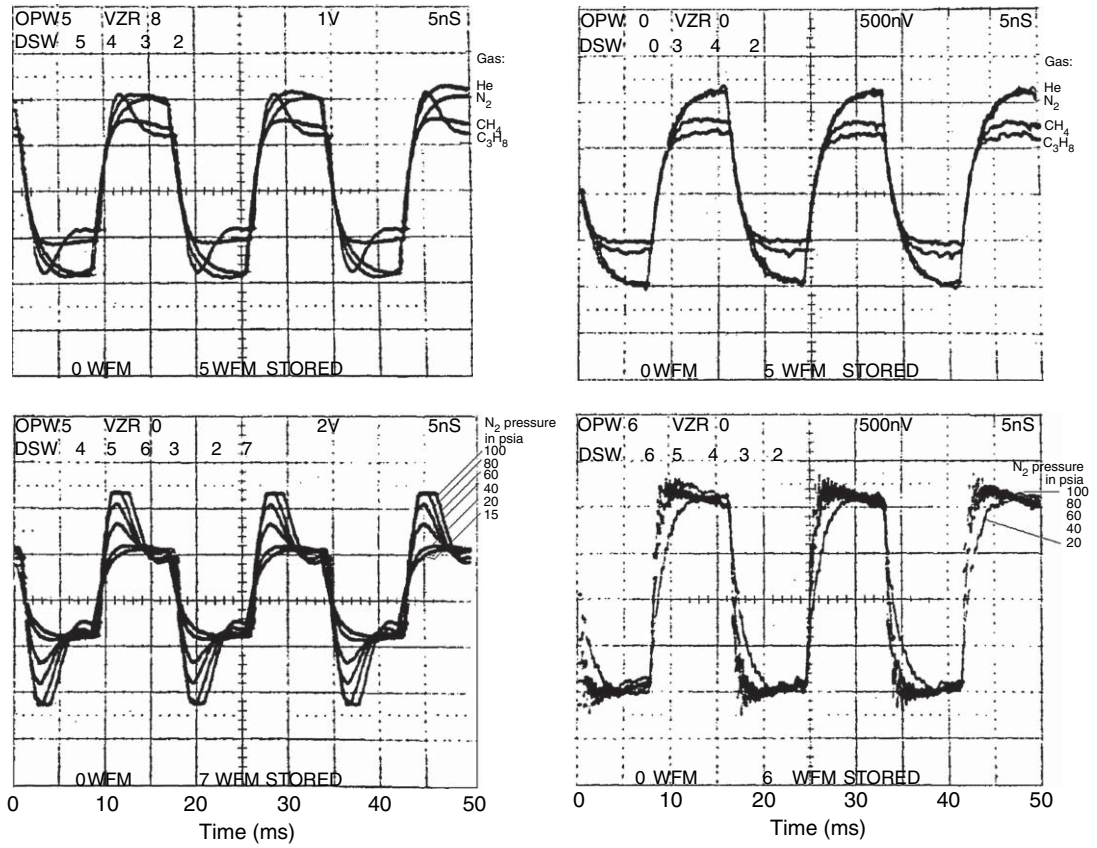


Figure 19 Experimental gas viscosity pressure signals versus time for various gases (He, N_2, CH_4, and C_3H_8) and pressures (20–100 psia or 1.3–7 bar). *Top left:* electromagnetic (EM) driver/0.275-mm capillary/electret microphone. *Top right:* EM driver/porous metal frit leak/lead zirconate titanate (PZT) pressure sensor. *Bottom:* all bottom pressure data were obtained with N_2 at the indicated pressures. (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission. Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *Proceedings GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002, © 2002 reproduced with permission from Gas Technology Institute.)

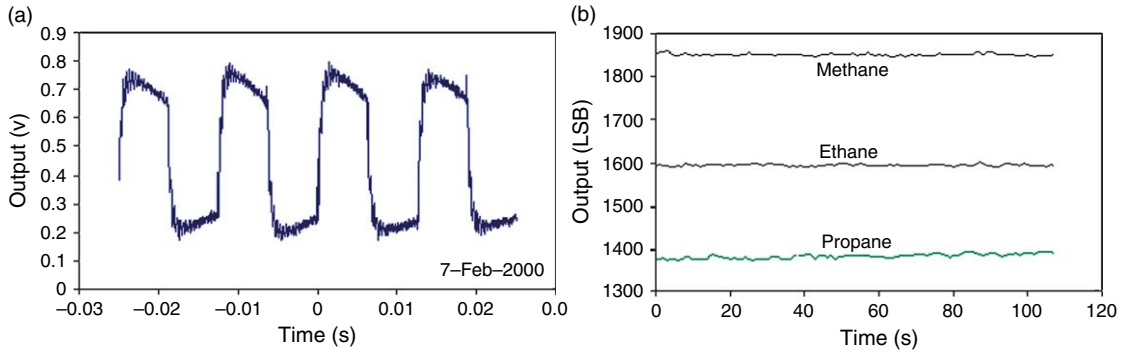


Figure 20 Experimental gas viscosity sensor digital outputs versus time and type of gas; $f = 80$ Hz. (a) The shown noise level may largely be due to driver resonance at 2.3 kHz and the lack of signal plateau flatness may be due to speaker distortion. (b) The sensor outputs versus time reveal S/N levels and remaining drift rate. (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission. Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *Proceedings GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002, © 2002 reproduced with permission from Gas Technology Institute.)

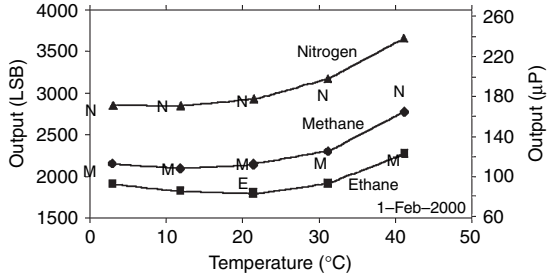


Figure 21 Experimental gas viscosity sensor digital outputs before temperature compensation versus ambient temperature and type of gas. (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission. Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *Proceedings GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002, © 2002 reproduced with permission from Gas Technology Institute.) The letters represent the intrinsic temperature dependence of N_2 (N), CH_4 (M), and C_2H_6 (E), indicating the presence of other effects.

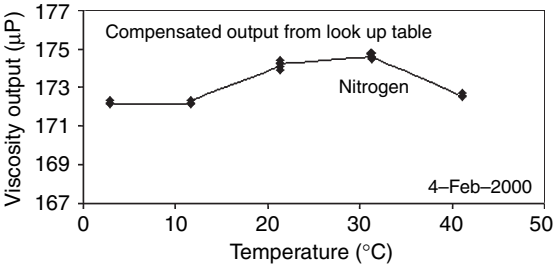


Figure 22 Measured gas viscosity sensor sensitivity to temperature changes, after the implemented temperature compensation. The temperature error is ±0.013% ($^\circ C^{-1}$). (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission.)

Figure 21. The remaining temperature sensitivity is plotted in Figure 22, showing that it decreased to about 0.026% ($^\circ C^{-1}$), which is about 10× lower than the theoretical temperature dependence of N_2 (0.25% ($^\circ C^{-1}$)) and 27× times lower than the experimental temperature sensitivity of Figure 21 (0.71% $^\circ C^{-1}$). The measured pressure sensitivity of 3.4% bar^{-1} for N_2 is much higher than the one predicted by the PPDS/NEL data of 0.14 and 0.09% bar^{-1} for CH_4 and N_2, respectively, so that some fine-tuning and adjustments would be in order. However, many near-ambient-pressure applications not requiring operation over a pressure range larger than 100–200 mbar or accuracy below ±1% may be well served by the simple sensor described above, without further refinements (Figure 23).

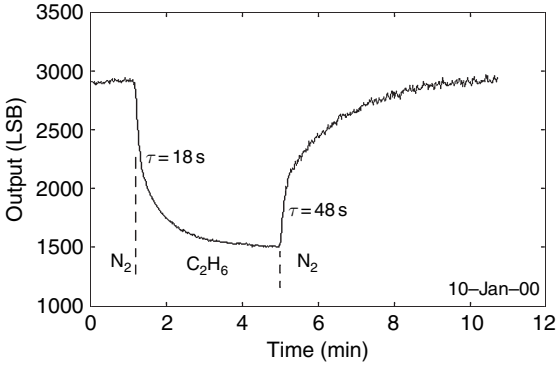


Figure 23 Measured response time of the gas viscosity sensor based on exposure to N_2, C_2H_6, and back to N_2. The 18–48 s was much longer than anticipated and may be due to the time needed for residual gas to diffuse out of the small pores of the used frit, which is associated with a longer time than the few milliseconds needed for pressure equilibration in the viscosity sensor. (Source: Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct.* **10**, 1185; reproduced with permission.)

Table 2 summarizes the data on the present performance of the developed viscosity sensor performance. It includes data on the significance of measured noise (∼30× lower at 80 than at 60 Hz) expressed in terms of a viscosity reading (±0.3%) as well as in terms of a feedforward combustion control signal in which a change in the composition of a hypothetical ($CH_4 + C_2H_6$) fuel gas mixture needs to be controlled (±2.3%). However, all other sensitivities in Table 2 are given in terms of viscosity sensitivities. Also listed are the measured response time (30 ± 15 s) resulting from rapid switching from N_2 to C_2H_6 and back; voltage sensitivity ±0.07% over the range from 8 to 23 V; and the p and T sensitivities discussed above.

2.11.2.5 Conclusions

The above results and demonstrations indicate the feasibility and cost-effectiveness of coupling fluid motion induced by simple, available, and affordable membrane actuators to thermal or pressure microsensors to provide advantageous means to determine:

(1) Accurate, submillisecond response times of different flow-sensor structures and at different flows. At flows of up to ∼1 m s^{-1}, the MB sensor flow response time decreases to 0.8 ms, but the value for the more rugged Microbrick drops to 0.3 ms, and lower still at higher flows.

Table 2 Gas viscosity sensor performance data

Sensitivity[a]	~300 LSB signal for CH_4–C_2H_6. Linear with viscosity, η
Noise	±7 LSB or ±0.3% of viscosity reading, or ±2.3% of a CH_4–C_2H_6 signal
Response time	~30 ± 15 s (63%) by switching from N_2 to C_2H_6 and back to N_2
Signal stability	
Versus temperature with compensation	$\leq \pm 0.0.16\%\ ^\circ C^{-1}$ over 2–42°C range
Versus temperature cycling	≤ ± 0.7% after six temperature cycles between 0°C and 40°C
Versus absolute pressure	$\leq \pm 1.7\%$ of η bar^{-1}, or 5 LSB psi^{-1} from 10 to 80 psia
Versus input voltage	$\leq \pm 0.07\%$ of η, or $\leq \pm 0.5\%$ of *Bn*, for $8 \leq V_{in} 23$ V
Versus EMI	TBD
Versus shock and vibration	Output did not change after tapping the sensor with a wrench from any side

[a]For reference gas

Methane	2100 LSB	≅108.6458 μP	≅2.000 B*n*
Ethane	1750 LSB	≅91.3274 μP	≅2.556 B*n*
% change	−12.5	−16	27.8

Bn or Bonne number is defined similarly to the Wobbe number, except for using volumetric or molar oxygen demand, D_{O_2}, rather than the heating value in the numerator, and CH_4 MW rather than the air density as reference in the denominator:

$$\mathrm{Bn} = \frac{D_{O_2}}{\left(M_{gas}/M_{CH_4}\right)^{0.5}}$$

The optimal correlation between *Bn* and viscosity is gas dependent as shown below. For a small set of US and European natural gases, plus test gases as defined by Gasunie in the Netherlands, the correlation

$$Bn = 10.068 - 0.7192\eta^{0.507014}$$

was found useful. However, we have found that the viscosity exponent can vary from 0.3 to over 2.3 for the optimal correlation, depending on the chosen set of gas compositions for the fit.
Source: Bonne and Kubisiak (2001a); © 2001 reproduced with permission of IOP.

Conservatively, this means that these sensors can be operated in oscillatory flows at frequencies of up to ~200 and 530 Hz, respectively. Application: Defining the actuator frequency range for the measurements discussed below, and generic response time limitations of such microstructure flow sensors.

(2) Affordable composition correction factors for thermal and other flow sensors. The laborious approach used previously for thermal anemometers, based on the measurement of thermal conductivity and specific heat, can now be replaced by one simple flow measurement to determine the volumetric composition correction factor, C_V, with a comparatively negligible effort for calibration and electronics, that achieves at least ±3% accuracy. Application: Flow sensing of fluids like gaseous or liquid fuels, whose multicomponent composition can vary versus time, grade, and source.

(3) The composition of binary and higher-order mixtures. The determination of the correction factor, C_V, adds one easy-to-measure parameter to the choice of parameters available to characterize the composition of gaseous or liquid mixtures.

(4) Viscosity and other physical properties of fluids. With the laminar flow induced by a simple oscillatory membrane (EM rather than PE earphone speaker) operated below 100 Hz, we demonstrated a compact, light weight, and affordable approach to viscosity measurement. The measurable AC pressure drop signal serves indeed as a linear indicator of viscosity according to Poiseuille's law for viscous flow, which is density independent if operated at low kinetic energy conditions and also pressure independent to the extent that the measured viscosity is pressure independent (~0.1% bar). One driver for this development was to correlate viscosity (by itself or in combination with other parameters) to oxygen demand or Wobbe number.

Two other classes of thermal sensors, also demonstrated with MB sensor chips, represent a result of efforts to minimize sensor drift by shifting sensor readout from being dependent on analog changes in some physical sensor property, to frequency or digital changes in such sensors. But here they shall only be mentioned in passing:

(1) Those geared to induce electronic RC type of self-oscillations by leveraging the temperature-dependent changes in the thin-film resistance of the sensor and the capacitance of the circuit. Such oscillations then change the sensor's output frequency signal as a result of changes in flow, thermal conductivity, or specific heat. Consistent with the thermal response time, frequencies of up to 1000 Hz were generated (Bonne and Kubisiak 2001b). The generation of a frequency output may be appealing, although does not eliminate the dependence of the underlying analog change in sensor resistance.

(2) Those based on measuring the variable time delay associated with the change in heat transfer between the sensing elements involved in the measurement of the above thermal properties. Such digitally measurable delay, if primarily dependent on changes in flow rather than in the analog resistances of the sensor elements, would bring us closer to the ideal digital sensor (Bonne and Kubisiak 2000, Bonne *et al.* 2000b, 2001b). In this promissing approach, one should still check and verify that the influence of the sensor analog resistances may still play a role in the overall sensor operation.

2.11.3 Gas Composition Sensors

2.11.3.1 Introduction

Composition sensors may be grouped into many kinds of categories, classes, or types. Because of the impact of the availability of sample mass on MDL, the two types discussed below are sensors that are either mass flow dependent or independent. This grouping has special relevance to microanalyzers because of ultimate S/N limitations when sample mass flow becomes very small. Examples for the former are those based on MS, catalytic beads, paper tape, and amperometric electrochemical (AEC), which actually consume the analyte; examples for the latter are based on nondispersive infrared (NDIR), thermal conductivity detection (TCD), and potentiometric electrochemical (PEC) sensing.

Whenever possible, we strive to report concentrations and MDLs in mole or volume fractions, such as in percent, ppm, or ppb, rather than as a minimum detectable mass flow, such as $\mu g\ s^{-1}$ or $pg\ s^{-1}$, which researchers of mass flow-dependent sensors are known to do. Concentration is a direct measure on how a compound affects our well-being – beneficially (we like to breathe 21% O_2) or adversely (e.g., when exposed to 100 ppm CO). Occupational Safety and Health Adminstration (OSHA) and material safety data sheet (MSDS) safety instructions are also issued in terms of concentrations and not mass flows.

Without unduly dwelling on how to report MDLs, it may be helpful to compare the two units for conventional (and micro) sensors with specified capability to sense a level of 1 ppm of analyte (at a $S/N = 3$). In mass flow sensor terms, sampling at $0.5\ cm^3\ min^{-1}$ to a sensor with a response time of 1 s (or 0.01 s) and an analyte pulse of 0.010-s duration, the 1 ppm MDL would correspond to an MDL of $1\ ng\ s^{-1}$ (or 0.01 $ng\ s^{-1}$ or $10\ pg\ s^{-1}$) in mass flow terms, respectively assuming an analyte MW near that of air.

The sensor composition topics selected for discussion cover a small fraction of the vast field of such sensors and only represent those that the author had some recent involvement with, i.e., it is an arbitrary selection, but one with which we hope to highlight some generic insights. In the conclusions, some typical sensitivities (**Table 3**) and relative advantages and disadvantages for a bit broader range of composition sensors are considered (**Table 4**), including widely used infrared and catalytic bead monitors (Honeywell Analytics 2006, Kramer and Mangieri 2006).

2.11.3.2 Film Sensors – Capacitive, Resistive, and Colorimetric

When thin polymer films are exposed to gases other than dry reference air, a new adsorption equilibrium establishes itself between the gas and the film phase (or stationary phase in GC terminology), whereby a fraction of the new gas component or analyte in the air carrier gas diffuses into the film. This process is reversible when the analyte disappears from the gas phase. However, while present, the addition of the analyte to the film may cause measurable changes in its physical (swelling, mass (Lu *et al.* 2005, Manginell *et al.* 2003)), electrical (resistivity (Doleman *et al.* 1998) or dielectric constant (Patel *et al.* 2003)), and optical (color) properties. Sensors based on such changes are used to sense analytes in applications such as home safety, environmental, homeland security, and regulatory compliance. Surface acoustic wave (SAW) sensors make for a preferred approach for sensing the small change in mass, especially if the film is made thin enough for the equilibration to happen fast enough, and serve as stand-alone or GC detectors (Lu *et al.* 2005, Manginell *et al.* 2003). By

Table 3 Detection ranges for some common analytes versus available sensors

			Catalytic bead	*Metal oxide*	*Electrochemical liquid*	*NDIR 1-Pass*[a]	*Paper tape*
		Sensor model	*CiTipel.*	*CiTicel*	*Midas*	*IRcel*	*CM4*
Analyte	*Formula*	*Units*	*0–ppm (max)*	*0–ppm (max)*	*0–ppm (max)*	*ppm range*	*ppm range*
Ammonia	NH_3				100		3–75
Arsine	AsH_3				0.2		0.005–0.5
Boron trichloride	BCl_3				8		
Boron trifluoride	BF_3				2		0.1–1.5
Bromine	Br_2				0.4		
Carbon dioxide	CO_2				0.4	500–20 000	
Carbon monoxide	CO			400	5–10		
Chlorine	Cl_2				2		0.007–2
Chlorine dioxide	ClO_2				0.4		0.03–1
Chlorine trifloride	ClF_3				0.4		
Diborane	B_2H_6				0.4		0.015–1
Dichlorosilane	H_2SiCl_2				8		
Difluoromethane	CH_2F_2				120		
Dimethylamine	C_2H_7N						0.3–30
Disilane	Si_2H_6				20		
Fluorine	F_2				4		0.6–10
Germane	GeH_4				0.8		0.1–2
Hexafluorobutadiene	C_4F_6				40		
Hydrogen	H_2		700–40 000		150–20 000		
Hydrazine	N_2H_4				1		
Hydrogen bromide	HBr				8		0.05–2
Hydrogen chloride	HCl				8		0.8–8
Hydrogen cyanide	HCN				20		0.5–50
Hydrogen fluoride	HF				2		
Hydrogen iodide	HI						0.1–25
Hydrogen selenide	H_2Se				0.4		0.006–0.5
Hydrogen sulfide	H_2S				40		0.5–100
n-Butylamine	$C_4H_9NH_2$						0.3–30
Methane	CH_4		800–45 000	1k–20 000		800–20 000	
Methyl fluoride	CH_3F				60		
Nitrogen dioxide	NO_2				12		0.3–30
Nitrogen oxide	NO				100		
Nitrogen trifluoride	NF_3				40		
Oxygen	O_2				25%		
Ozone	O_3			0.01–1	0.4		
Phosgene	$COCl_2$						0.007–1
Phosphine	PH_3				1.4		0.005–3
Phosphorous oxychloride	$POCl_3$				0.4		
Silane	SiH_4				2		0.5–50
Sulfur dioxide	SO_2				8		
Sulfur tetrafluoride	SF_4				0.4		
Tertiarybutylarsine	$C_4H_{11}As$						0.12–0.5
Tricloroethylene (TCE)	HC_2Cl_3						

[a]Response time 7 s.
Source: Honeywell Analytics (2006).

Table 4 Comparison of single-channel gaseous analyte-in-air sensors

Gas	*Advantage*	*Disadvantage*
Electrochemical solid	Highly specific to O_2. Potentiometric. Not sensitive to mass flow. Nondestructive. Can also be run as amperometric.	Requires <750°C, flame arrester for safety and high input power for operation. Operation down to 300°C is possible but yields poor performance/stability.
Catalytic bead	Simple, measures flammability of combustible gas–air mixtures. Low cost, proven technology.	Can be poisoned by lead, chlorine, and silicones that remain as unrevealed failure mode. Requires oxygen or air to work. Require high input power for operation, <400°C.
Metal oxide semiconductor	Mechanically robust, works well in constant high humidity/temperature conditions. Very low cost. Can give fail-safe behavior with poisoning.	Semispecific. Susceptible to contaminants and changes in environmental conditions. Nonlinear response prompts complexity (electronic correction). <300°C. Too slow (~1 s) for micro-GC application.
Thermal conductivity	Measures approximately volumetric concentrations of analyte(s) in reference gas. Nonspecific, thus good for GC detector.	High gas concentrations only. Limited range of gases. Cannot measures gases with conductivities close to air. Nonspecific. High humidity dependence.
Polymer film, change in *R*, *C*, *f*	Semispecific. Low-power operation. Easy fabrication. Sensitive as SAW	Semispecific. Slow. Frequent recalibration needed. Relatively high drift.
Electrochemical liquid (no consumable)	Measures toxic gases in relatively low concentrations. Wide range of gases can be detected.	Failure modes are unrevealed unless advanced monitoring techniques used. Requires oxygen to work.
Electrochemical liquid (consumable anode) oxygen sensor	Linear response to oxygen. Selective.	Limited lifetime (typically 2 years) due to consumable lead anode. Trade-off between sensitivity and lifetime.
Nonfocusing NDIR absorption	Nondestructive analysis to ppm level. Can be used in inert atmospheres. Optical alignment not critical. Small size.	Higher power requirement than electrochemical. Can only detect IR-active species.
Multiple-pass NDIR cavity absorption	Nondestructive analysis to ppb level. Can be used in inert atmospheres. Easy alignment.	Requires high-tech mirrors (<99.99% reflectivity) to achieve <3000 passes and drift-free operation.
NDIR-absorption with frequency modulation	Nondestructive analysis to ppm level. High reliability due to frequency modulation; % abs. immune to source or detector drift.	Requires sophisticated λ-dither mechanism. Fabry–Perot cell fabrication demands tight tolerances and alignment.
Paper tape	Cumulative/dosage sensor signal. Highly sensitive and selective for toxic gases. Leaves physical evidence of the gas exposure No drift. New sensor for easy measurement.	Can require sample conditioning and extraction systems. High sample flow needs. Requires refills of consumable paper.

Source: Hogan *et al.* (2006).

controlling the temperature of the films, the desired trade-off between speed of response and sensitivity may be achieved.

The availability of a temperature-controlled platform for metal oxide films of SnO_2, WO_3, In_2O_3, etc. (with or without doping of Pt and other metals) served to fabricate sensors that respond with a measurable change in their semiconducting properties to the change in the concentration of oxidizing or reducing analytes. Such sensors were introduced in the 1960s based on thick films of SnO_2 by Figaro. Tin oxide is still the most commonly used metal oxide gas sensing material. In the 1990s, thick-film sensors based on chromium titanate were introduced commercially by Capteur (Dawson *et al.* 1995, Niemeyer *et al.* 2002), offering improved stability and reduced humidity interference.

Metal oxides have been further developed as thin films by researchers at Carnegie Mellon (Clifford and Tuma 1982–1983, Hoffheins *et al.* 1986, Siegel 1989) and Tübingen Universities (Kleine-Benne *et al.* 2005), NIST (Boger *et al.* 2005, Cavicchi *et al.* 2000, Dable *et al.*, 2004,

Meier *et al.* 2004, Tiffany *et al.* 2000), and more recently at ETH Zurich (Barrettino *et al.* 2004), Motorola and MicroSense, in the form of single and arrays on micro-hotplate sensors, leveraging temperature-dependent sensitivity and doping to reduce response time, and increase selectivity and baseline stability.

Less well known is the use of films of chromium titanate by CTL Capteur (Dawson *et al.* 1995, Niemeyer *et al.* 2002), which claim better stability. Stability may be further improved via a differential sensing approach (Visser *et al.* 1994, Zanini *et al.* 1995), as used with other sensors (NDIR, TCD, etc.), and has successfully been implemented with the hot catalytic bead sensor (Honeywell Analytics 2006), in which the reference bead is identical to the sensing bead, except for not having been provided with the metallic doping for the catalytic response. Both these and handheld infrared detectors (Honeywell Analytics 2006, see Table 3) are used to monitor fuel gas concentration and degree of flammability.

Another method of improving the stability, response time, and/or selectivity of metal oxide sensors makes use of high temperatures to reset the sensor to a known state and then monitor the transient behavior when responding to gas at a subsequent lower temperature. **Figure 27** illustrates such an approach with semiconducting WO_3-based O_3 sensors. Such approaches are used in commercial sensors by Figaro (Massok *et al.* 1995) and Capteur (Aliwell *et al.* 2001, Pratt *et al.* 2001).

For all the above sensors, one expects that the physical changes induced by the analytes are reversible. However, as with other sensors, drift erodes reversibility, so that periodic recalibration is needed, unless one can afford to use a new sensor for each measurement.

But this is exactly what paper tape-based sensors do. Rather than emphasizing reversibility, as with some polymer films (Daws *et al.* 1997), one widely used category of paper-film sensors leverage color change with time that is irreversible, i.e., cumulative, and can directly relate to doses, e.g., doses that workers are exposed to during some working period. **Figure 28** shows the working principle of such a system. The porous paper supports the reagent that turns color when the targeted analyte is present in the sample. Everyone is familiar with the classical litmus paper to gauge the pH of liquid samples, but few would have heard about other such paper spots for sensing chemical agents in water (M8 paper used by the military) or for sensing analytes in air, which is more relevant to the subject at hand (see Chemcassette analyzers used for air monitoring in the semiconductor industry (Honeywell Analytics 2006)). In this case, the sample is drawn through the pores of the paper, as for measurements of stack gas soot concentration with the classical Bacharach smoke spot test. If the intended analyte is present in the sample gas and comes in contact with the reagent in the paper, the color of the paper changes and is quantified via change in the reflected light intensity. Such color change can be highly specific to analyte categories such as hydrides, acids, halogens (Honeywell Analytics 2006), and may be used in the form of personal badges or central monitoring stations. **Table 3** presents a partial list of compounds presently detectable with less than a dozen types of reagents embedded in the paper.

2.11.3.3 Electrochemical Sensors – Potentiometric and Amperometric

The simplest form of an EC cell sensor is of a two-electrode design, as used in potentiometric, solid-state electrolyte (YSZ) oxygen sensors, with porous Pt electrodes, now widely used for air/fuel ratio control in stationary and automotive combustion systems. The basic output signal is a log function of the available analyte (oxygen) in the sample gas, relative to the concentration in the reference gas, typically air, which is represented by the classical Nernstian cell potentiometric output of

$$E = \frac{RT}{nF} \ln\left(\frac{[O_{2,\text{sample}}]}{[O_{2,\text{ref}}]}\right)$$

where the square brackets represent concentractions or partial pressures, R is the universal gas constant, T is the absolute temperature, F is the Faraday constant (96 500 C mole^{-1}), and n is the number of charges per ion. Later innovations leveraged this basic EC cell mechanism to provide sensor outputs that are linear with absolute partial pressure of oxygen, and are based on either closed cells or cells with a purposely added leak. With the closed cell, oxygen is pumped in and out by changing the polarity of an applied fixed DC voltage, so that the time for a complete cycle becomes the linear, absolute measure for the oxygen partial pressure present outside of the cell in the sample gas (Kroot *et al.* 1990). With the leaky cell, one simply obtains the sample gas oxygen from the current input into the cell, which just uses up all the oxygen entering into the cell by diffusion, while the

applied potential drives the oxygen out via ionic oxygen conduction through the YSZ. However, ionic conduction of YSZ requires operating temperatures of $\geq$750°C. While the hazard of using such potential ignition source as a sensor can be managed via classical flame arrestors, in the form of metal frits, its high power consumption has limited its use in portable devices, despite attempts to microminiaturize (MM) such sensors (Bonne *et al.* 1995b, Cole *et al.* 1986) or to further reduce the size of the already small, 2 × 2mm × 2mm cells.

Regarding the MM effort, future researchers may benefit from our experience with fabricating unsupported thin films of YSZ, by ion sputtering thin films of YSZ onto a Si_3N_4-coated Si wafer before etching away the silicon, resulting in the films shown in **Figure 24(a)** (Cole *et al.* 1986). Severe buckling of the compressively stressed films led to breakage if the thickness was increased beyond ~400 nm, yet at that small thickness, the measured output potentials (with also-deposited thin films of Pt) did not achieve the desired stability, although temperaturewise (see **Figure 24(b)**), the resistively heated films showed lifetimes of over 2000 h at ~800°C, with input powers of only 60–70 mW during the on-periods of 10 Hz pulsed DC, to simulate and check duty cycling (50%) to extend sensor life (Cole *et al.* 1986).

In any case, the choice of solid-state, ion-conducting materials to make specific potentiometric gas sensors is much more limited than the availability of liquid electrolytes. This has led to the development and availability of a great variety and number of low-cost AEC cell sensors, with the simplest again being of a two-electrode design (Application notes by City Technology Ltd. (Div. of Honeywell)). Such an EC cell consists of a semipermeable diffusion membrane, a reservoir of acid electrolyte, a sensing electrode and a counter electrode. The electrodes generally consist of a polytetrafluoroethylene (PTFE) substrate coated with a catalytic mixture, usually platinum, which significantly accelerates the sensing gas reaction. In most cases, any gas diffusing into the cell reacts at the surface of the sensing electrode producing, as a by-product, a number of ions (H^+) and electrons (e^-). The ions travel through the acid electrolyte to the counter electrode while the negatively charged electrons travel to the counter electrode via the external circuit. Combining the electrons and ions at the counter electrode completes the reaction without any of the cells components being consumed. The amount of electrons produced by the reaction is directly proportional to the amount of gas present and measuring the current flowing through the external circuit is a basic gas monitor.

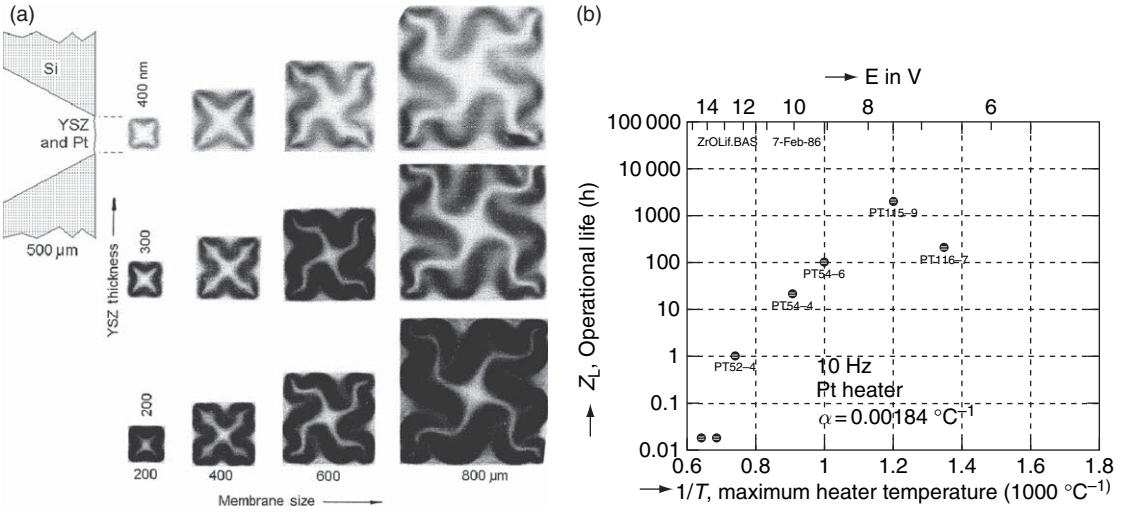


Figure 24 (a) Top-view micrograph of buckled yttria-stabilized zirconia (YSZ) thin-film membranes (with Pt thin-film heaters and electrodes), supported by a silicon chip (drawn to scale in cross section at left), showing increased buckling with membrane size, but with little dependence on film thickness. (Source: Cole B, Ule E, Schuldt S, Bonne U 1986 Oxygen microsensor development. Final Report to GRI, Contract No.: 5084-234-0984, GRI-86/0190 and -90/0170, Honeywell Labs, Plymouth, MN, March 31, 1986, revised July 21, 1986; reproduced with permission.) (b) About the life time of thin-film, self-heated YSZ sensors. (Source: Cole B, Ule E, Schuldt S, Bonne U 1986 Oxygen microsensor development. Final Report to GRI, Contract No.: 5084-234-0984, GRI-86/0190 and -90/0170, Honeywell Labs, Plymouth, MN, March 31, 1986, revised July 21, 1986; reproduced with permission.)

A key feature of this EC sensor design is the gaseous diffusion barrier (see (1) and (2) in **Figure 25**), which may take the form of such small openings, permeable membranes, or a combination of the two (Hogan 2006) that limit the flow of gas to the sensing electrode. This feature permits the electrode to react to target gases as it reaches its surface, and still has EC activity in reserve and thus a long life. The electrolyte reservoir even allows the sensor to operate for a short time in dry conditions without drying out and can contain a reserve of dissolved oxygen allowing the sensor to operate for a short time in the absence of oxygen. The diffusion barrier reduces typically large EC sensor temperature coefficients to the relatively small and predictable diffusion temperature dependence of $D = D_o(T/T_o)^{1.5}$.

Figure 25 shows details of the makeup of such a related, but a three-electrode cell, as well as its drive and sense circuit, representing a potentiostatic circuit, widely used for EC measurements. Besides the two main electrodes, sense (W in **Figure 25**, for working) and counter (C) electrodes needed for basic operation, the reference electrode (R) is added to achieve greater stability and thus lower MDL. The reference electrode is held in an environment in which it will not react with the analyte; it plays no part in the reduction/oxidation reactions in the sensor and is able to maintain a constant potential. In operation, the measurement is taken between the sensing and the reference electrodes, allowing the potential of the counter electrode to change as necessary without affecting that of the working electrode. Therefore, the performance of a three-electrode sensor is more stable, generally has a higher output than that of a two-electrode device, and can be used to monitor a wider range of analyte concentrations.

The circuit in **Figure 25** shows that the signal is generated by converting the cell current between W and R to an output voltage (bottom op-amp), while the top op-amp controls the reference potential. For the shown circuit:

- V_{OUT} = analyte concentration × cell current × R_F. If $R_F = 100\ \text{k}\Omega$, cell output is $50\ \text{nA ppm}^{-1}$ and gas concentration is 100 ppm then $V_{OUT} = 100 \times (50 \times 10^{-9}) \times (100\,000) = 0.5\ \text{V}$.
- The output polarity depends on the redox properties of the analytes being sensed.
- R_L is the cell load resistor (typically 5–50 Ω). A small value of R_L will increase noise, but give a faster speed of response; and *vice versa*.
- C_1 and C_2 help to reduce sensor signal noise and EMI susceptibility. Typical values are $C_1 = 4.7\ \text{nF}$ and $C_2 = 47\ \text{pF}$.
- R_B stops the sensor from polarizing when the circuit is powered off (typically 1–10 kΩ).

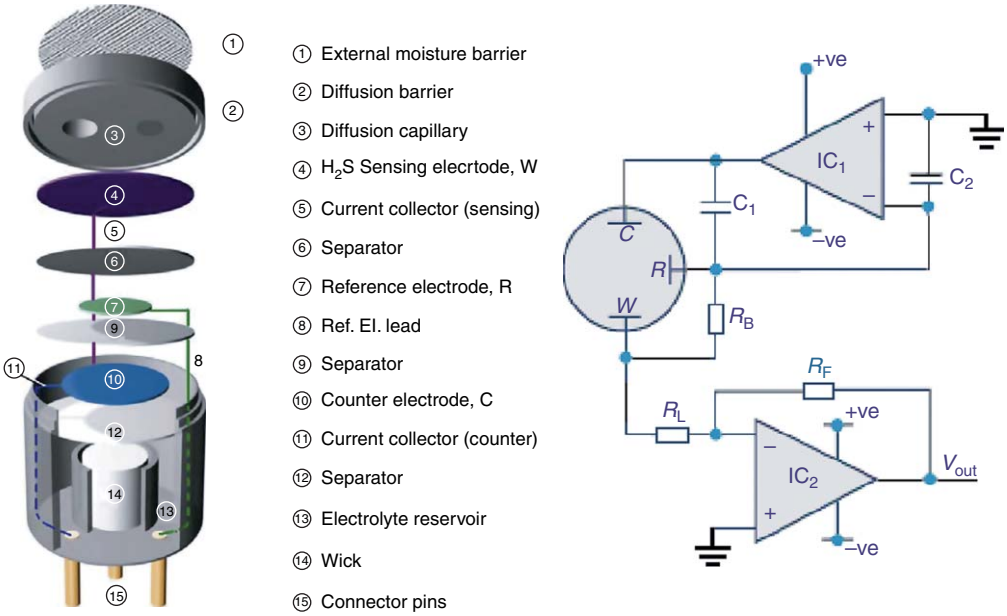


Figure 25 Component view of a three-terminal WO_3-based sensor and its circuit. The reference electrode provides greater stability than a two-electrode sensor. (Source: Application notes by City Technology Ltd (Div. of Honeywell), Walton Road, Portsmouth, Hampshire, PO6 1SZ, UK; reproduced with permission.)

- Some EC sensors require special features, such as temperature compensation circuitry for CO sensors and a small bias voltage for CO_2 sensors.

Due to the highly active nature of the catalysts used in EC sensors, such catalysts can oxidize or reduce gases other than the target gas. In these cases, the sensor exhibits a cross sensitivity, which can, on occasion, be used to extend the measurement capability of the sensor. For example, chlorine sensors exhibit a significant cross sensitivity to chlorine dioxide (ClO_2), bromine (Br_2), and ozone (O_3) and can therefore be used as a suitable sensor for detecting these gases when none of the others are present. The effect of interfering gases may be reduced by catalyst selection, use of on-board filters, or by applying a bias voltage between the working and reference electrodes.

Given the low power consumption of such sensors, intrinsic safety (IS) certification should be easy to obtain. Theoretically, it is possible for an EC sensor to produce up to 1.25 V during an open circuit, but as most of the reactions are irreversible the full potential is never reached. Sensors of Honeywell Analytics will not produce a voltage greater than 1.3 V and a current greater than 1.0 A (Application notes by City Technology Ltd. (Division of Honeywell)).

2.11.3.4 Conclusions

By way of summary, **Table 3** provides a partial list of the variety of analytes sensed by just a few of the sensors discussed above, along with sensitivities (Honeywell Analytics 2006). Paper tape-based sensors have the advantage of leveraging their cumulative, nonreversible color spot, and can therefore increase their sensitivity by increasing the sampling time. For the listed sampling times in the 10- to 60-s range their MDL is in the 2–5 ppb range, which is ~10–30× higher for the EC sensors and typically higher for the portable NDIR type, with the exception of optical cavity approaches, which have achieved NH_3, H_2O, and acetylene (Bamford *et al.* 2003, Gupta 2004) detection in the low ppb range.

Table 4 presents advantages and disadvantages of a number of individual channel sensors, beyond the above discussion, but in use today by us and others on both productized and developmental bases, borrowing heavily from and expanding on inputs received from experts at Honeywell Analytics (Hogan *et al.* 2006) and others such as B Warmack at ORNL, T Mlsna at Seacoast Science, and S Semancik at NIST-CSTL. The rows are arranged in decreasing temperature of operation, starting at the top with the YSZ O_2-sensing cell, which requires operating temperatures above ~750°C. Nonspecificity is a desirable feature (e.g., thermal conductivity) if the sensor serves as a detector with a GC, for the range of gases to be detected (ideally combined with insensitivity for the gases not of interest), as discussed in Section 2.11.4. Such joint operation eliminates a large part of the drift problem, as the GC provides a new baseline at every analysis.

Users may be willing to forgive some reversible temperature and humidity dependence, insensitivity, limited (fail-safe) service life, and cross sensitivity, but not unpredictable drift or unreliable operation. After having accepting the fact that all sensors drift, the chemical processing industry achieves the needed reliability by having test gases available for automatic calibration for their online process analyzers, as further elaborated in Section 2.11.4. Carrying calibration gas tanks with and or portable or handheld sensors for an 8-hour day can be quite burden some, although practiced in some cases as, e.g., with (hydrogen) flame ionization detectors (FIDs) for natural gas leak detection.

AEC sensors have made a virtue of their unavoidable mass flow dependence and leveraged it to generate a pulsed output from regular to zero, in accordance with periodic sample flow interruptions (Pratt *et al.* 2003), as illustrated for an ozone sensor in **Figure 26**. Of course, the pump flow rate needs to be repeatable and the stoppages need to last long enough time to allow the EC sensor to consume the analyte over its dead space, and thereby to drop its output signal to near zero. The sensor has a modest response at low gas flow rates; at intermediate flow rates the response is approximately proportional to the gas flow rate, while above the critical flow rate the response is fully developed.

To overcome drift, optical nondispersive absorption sensors (e.g., NDIR) have resorted to the dual-beam approach or, better yet, to a single beam with a chopper wheel transmitting at least two narrow wavelength bands corresponding to on and off band transmission of the analyte of interest. Such frequency modulation approach can be achieved via a Fabry–Perot interferometric scheme (Barrett 1976, 1977, Jerman *et al.* 1990, Raley *et al.* 1992), which eliminates the effect of drift in the source intensity, detector sensitivity, and/or

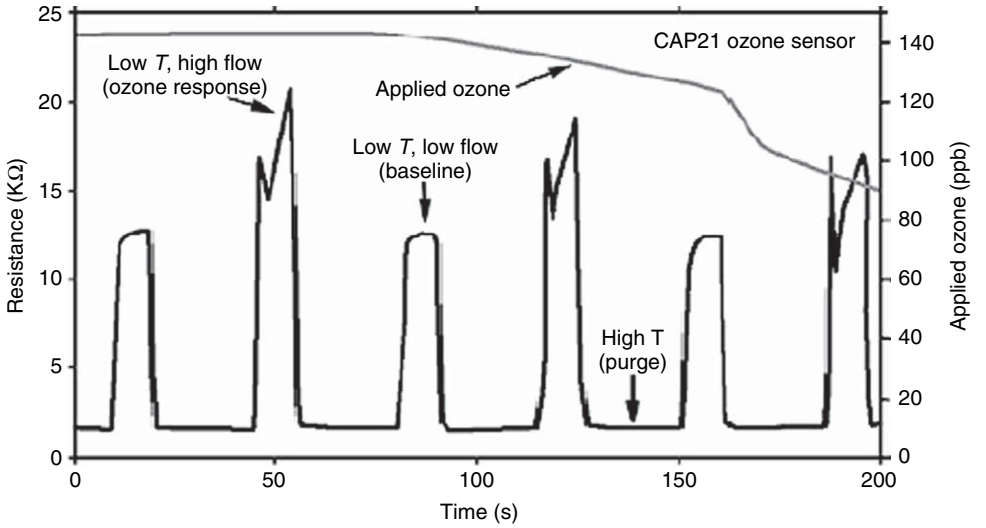


Figure 26 Measured WO_3-based ozone sensor output as its sample flow rate is periodically interrupted to reveal the signal baseline level, and thus to reduce drift and uncertainty. Also demonstrated is the benefit of periodical high-temperature (600°C) cycles to restore the sensor to reference conditions. (Source: Pratt K *et al*. 2003 The effect of flow rate on cap21 ozone sensitivity. Capteur Division of City Technology Ltd, Walton Road, Portsmouth, Hampshire, P06 ISZ, UK, June 2003, Website at http://www.citytech.com/PDF-Datasheets/MMDS_Cap21_t2ops.

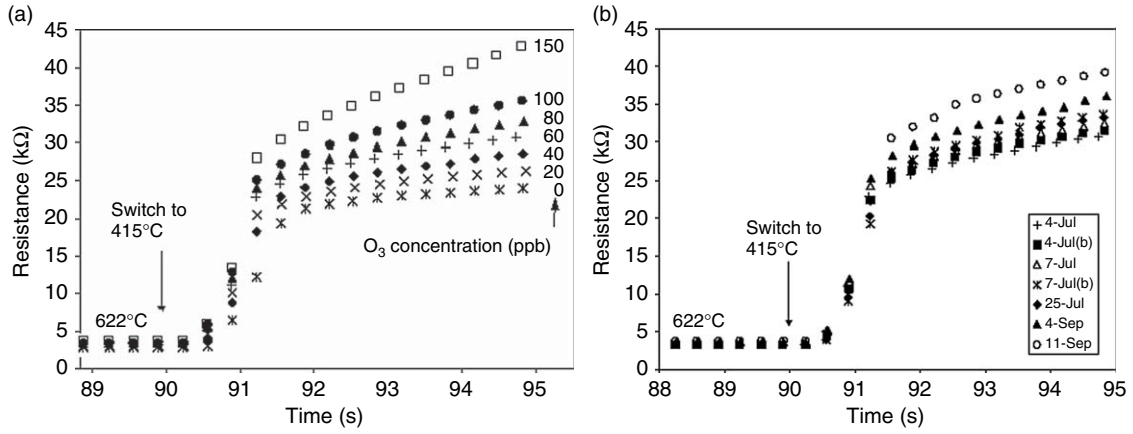


Figure 27 Calibration responses of an O_3 sensor based on a film of semiconducting WO_3, in a sample gas flow rate of 1 l/min, versus time, when switched from 622°C to 415°C. (a) For the listed O_3 concentrations. (b) After the listed days of operating time with an O_3 concentration of 60 ppb. At the higher temperature, O_3 decomposes before reaching the sensor and effectively provides a zero-O_3 baseline, thus reducing signal drift. (Source: Aliwell S R, Halsall J F, Pratt K F E, O'Sullivan J, Jones R L, Cox R A, Utembe S R, Hansford G M, Williams D E 2001 Ozone sensors based on WO_3: A model for sensor drift and a measurement correction method. *Meas. Sci. Technol.* **12**, 684–900.); reproduced with permission.

transmission of the optical surfaces, and is used, e.g., in the Vaisala CO_2 sensors (CARBOCAP CO_2 Transmitter).

The dual or differential sensing approach is used also with the popular and low-cost catalytic bead sensors, where only one of the beads is provided with the small mass fraction of catalyst, but both are exposed to the same sample gas environment (Honeywell Analytics 2006).

Last but not least, and ranking high in terms of reliability, is the paper tape sensor mentioned earlier, whereby a fixed volume of sample is drawn through a reagent-impregnated, porous paper and the irreversible color change is monitored as a measure for the amount of analyte present in the sample. After a measurement, the tape is advanced to a new and fresh paper spot location, thus providing an inherently stable reference baseline (relying on carefully

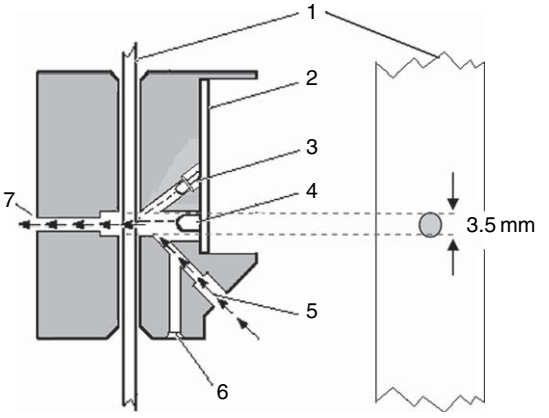


Figure 28 Cross-sectional schematic drawing of a paper tape-based gas-sensing system. Analyte reaction with the reagent in the paper results in a change in color (color spot), which is detected photoelectrically. 1, Paper tape; 2, PCB; 3, LED light source 4, detector; 5, sample gas inlet; 6, pressure-transduced port; 7, sample outlet. (Source: Honeywell Analytics 2006 *Gas Book*, May 2006, and webpage at http://www.honeywellanalytics.com, *Product Range, Analyziers and Sensors*: Honeywell-City Technology at http://www.citytech.com/technology/pellistors.asp and http://www.citytech.com/technology/irsensors.asp)

controlled paper and reagent uniformity), while the previous one remains as a permanent record (Honeywell Analytics 2006) (**Figure 28**).

Another way to provide that stable baseline is by scanning an independent variable such as wavelength, mass, temperature and time as optical spectrometers, MSs, DTAs, and GC analyzers do, respectively, which are to be discussed next.

2.11.4 Multichannel Micro Gas Analyzers

2.11.4.1 Introduction

This section focuses on advanced state-of-the art development of micro gas analyzers (MGA). As defined earlier, we view analyzers as representing multichannel sensors, which are therefore able to identify and quantify the concentration of more than one or two analytes. The larger the number of such independent (or orthogonal, see Section 2.11.5) channels, the better such analyzers can perform their identification job. Sensitivity or MDL of each channel is clearly a separate analyzer performance parameter, but equal to the MDL of single-channel sensors.

Today's commercial micro gas chromatographs or micro-MSs typically entail packaged systems of a half to one shoe box in volume or 5000–9000 cm^3, i.e., portable but still outside the range one would expect for handheld devices of below ~500 cm^3. The discussion below shall be about such analyzers, down to the 50-cm^3 level or cell phone size level, which would serve in applications of first responders, homeland security, medical and breath analyses (Sanchez and Sacks 2003), environmental emissions, and industrial *in situ* process stream sampling (shorter sampling lines and no analyzer sheds).

We shall further narrow the scope of the discussion below, by limiting it to multidimensional microanalyzers, then in Section 2.11.4.4 to using GC as the first-dimension analyzer or analyte separator, and then engage polymeric (chemical impedance detector (CID)) and optical or mass/ion mobility spectrometry (MS/DMS (differential mobility spectroscopy)) for the sample analysis in the second dimension. GC has the unique capability among all other analyzers, to make available for further analysis significant quantities of separated analytes, which then enables the use of multichannel second dimension analyzers to do a better job with analyte quantification and identification, than if that GC stage was not available. Clearly, second dimension analyzers (such as MS, ion mobility spectroscopy (IMS)/DMS, microdischarge detector or device (MDD), CID, and surface-enhanced Raman spectrometry (SERS)) are successfully being used as stand-alone detectors as well.

Venture capital and government-sponsored R&D has recently spawned a number of micro-GC-based developments falling into the 50- to 500-cm^3 range (Bonne *et al.* 2004, 2007, Lehmann *et al.* 2005, Lu *et al.* 2005, Manginell *et al.* 2003, Masel 2004, Pau *et al.* 2006, Polla and Nguyen 2004, Simonson *et al.* 2004), with the goal of achieving even further compactness and low power for battery operation and portability. The need for such MGAs and involved trade-offs were well described in Polla and Nguyen (2004), Bonne *et al.* (2007), and Bonne *et al.* (2003). Some general insights gained there were as follows:

- MM is an enabler to achieve better MGA performance in terms of response time, sensitivity (because of the enabled integration of rapid micropreconcentrators (micro-PCs)) selectivity (because of the affordability of a larger and multidimensional channel capacity), FAR (because a larger and more multidimensional channel capacity, despite the smaller time-channel capacity of a shorter micro GC column than that of a longer column of a desk-top GC), size, weight (from 50 lb to 50 g),

energy per analysis (reduction by 1000–10 000×), and cost (reduction by 10–100×).

- Optimal performance of MGAs would be achieved with units which separately optimize the preconcentration (PC), selectivity, and quantitation functions, and then closely couple them, rather than integrating them into one operational step as for example in a system comprising an array of chemiresistors or a fixed laser as light source for optoacoustic detection.
- Sample gas transport: MGA operation depends on reliable sample (or carrier) gas flow control, which is now dependent on failure-prone mechanical pumps. Alternatives to present mechanical gas pumping and flow control need to be developed, to provide low power and reliable service life (Bonne *et al.* 2003). Pumping based on thermal gas expansion and contraction (Bonne 2007, Bonne *et al.* 2006a, Gokhfeld 2006), ion drag (Garimella *et al.* 2004, Madhan *et al.* 2003, Roth 2003), as well as fluidic–diode effects (Morris and Forster 2003) are worthy of further study. The review by Laser and Santiago (2004) may be another good starting point.
- Manufacturability considerations: Beyond R&D successes, ultimate success in the eyes of potential users requires that researched MGA designs and capabilities can translate into manufacturable, high performance, and economically viable products. Wafer-level fabrication and assembly needs to be leveraged as much as possible, and is enabled by MGAs.
- The GC–MS combination is viewed as the Gold Standard for chemical analysis, because it combines the separation of analytes – with their MW identification (Bonne *et al.* 2007). Proponents of Raman spectroscopy (Haynes *et al.* 2005) would argue that MS cannot distinguish organic isomers (Heaps and Griffiths 2005) and that fragmentation during the MS ionization step, despite assisting with analyte identification, unduly complicates the mass spectrum when several analytes need to be sensed simultaneously. Thus the GC–SERS combination is clearly an analyzer of interest (Carron 1997, Carron and Kennedy 1995, Haynes *et al.* 2005, Heaps and Griffiths 2005).

2.11.4.2 Optical Spectral Absorption and Emission

Among gas analyzers to determine the concentration and identity of single or mixed analytes, NDIR, visible, and ultraviolet optical absorption analyzers and RLGAs (Rich 2001) are commercially available to measure analytes down to a 20–100 ppm range. To reach MDLs below that level with absorption analyzers requires large optical path lengths and associated bulkiness, maybe with the exception of analyzers based on optical absorption cavities featuring thousands of internal reflections (Bamford *et al.* 2003, Gupta 2004), including CRDSs but which have a limited spectral range dictated by the laser source while requiring relatively large sample volumes, polar analytes and sophisticated mirror technology. Such absorption limitations of path and wave-length, and analyte polarity are not hindering Ramas scattering approaches such as RLGA and SERS, besides exhibiting great selectivity, 10-ppm-level sensitivity and 50 ms response time, as demonstrated by ARI's analyzer (Rich 2001).

Also, the optical spectrochemical emission approach bypasses the above problems, while, of course, also having to cope with its own. It is an outgrowth of classical methods based on arc discharges (still used today as high-power atomic emission detectors (AEDs) for GCs) and modern plasma display technology using noble gas fillings. Such microglow discharges output spectral emission of lines and bands indicative of the elements and molecules in the sample. To achieve such glow discharges, well-documented conditions of applied voltage needed for breakdown, electrode spacing, gas type, and gas pressure are described by classical Paschen curves (Raizer 1991). At first, such MDDs only operated in He or Ne at subatmospheric pressures, for only a few minutes in air. But a recent effort by Herring *et al.* at Caviton Inc. resulted in extended (>5000 h of continuous and stable) operation in ambient pressure air, optional operation in 100- to 200-μm ID channels, i.e., using extremely small dead volumes, a few milliwatts of input power, and yielding emission lines and bands over the whole UV, visible, and IR spectrum. Photodetection may consist of individual channels provided by narrow ($\Delta\lambda \sim 2$–5 nm) band pass filters, or, preferably, by small spectrometers with photodetector arrays and CCD readout, as offered by, e.g., Ocean Optics, or more compact ones by Headwall (measuring less than 1 cm × 2 cm × 5 cm) and maybe others. Clearly, such MDDs are attractive detectors:

(1) By themselves, having demonstrated sub-ppm MDLs.
(2) As fast-responding and low dead volume detectors for micro-GC analyzers, as discussed in Section 2.11.4.4.

This is illustrated as follows.

- **Figures 29** and **30** present configurations of recent use for MDDs, with provision for sample gas flow, protection of the optical fiber end (leading to the PD or spectrometer) or window from metallic electrode sputter depositions, and maximum collection of optical emission intensity (Bonne *et al.* 2006a, Herring 2006b). Discharge excitation is preferably with AC in the 20- to 200-kHz range and between ~250 V for the configuration in **Figure 29** (Herring 2006b, Fiona *et al.* 2002) and 1000–2500 V_{p-p} for the one in **Figure 30** to overcome the AC impedance presented by the dielectric barriers.
- The photographs in **Figure 31** are of actual MDD discharges, with the one at left taken after logging over 4300 h of continuous operation in air, and at right an MDD operating in a channel with N_2 flow (Bonne *et al.* 2006a).
- To illustrate what the lowest operating voltages for an MDD might be, **Figure 32** presents a reproduction of classical Paschen curves (Raizer 2001). As shown, the minimum breakdown voltage requirements without the dielectric barrier are just under 200 V for Ne and under 400 V for air.
- MDD emission spectra in **Figures 33** and **34** were taken to reveal the presence of 2.5 ppm traces of Ar, and of CO and CO_2, respectively, in He with observation times of 1 s. Note that the reference He gas spectrum also shows the presence of some Ar and O_2, demonstrating the sensitivity of such spectral emission, but also hinting at the fact that heating the sample lines is needed, even for work with the above high vapor pressure gases, to minimize purge times and concentration reading errors.
- To better visualize and separate emissions of targeted analytes from background emission lines and bands, the formation of spectral differences is helpful, or as **Figure 35** shows an even more effective approach may be the formation of spectral ratios.
- If an MDD is fitted to the output of a GC, wavelengths of special interest may identify the analyte peaks as they elute, as shown by the traces in **Figure 36**, which demonstrates the 2D approach of a GC–MDD combination. The MDD had been placed at the exit of a temperature-ramped, 5-m/100-μm-ID/400-nm DB-5-coated column of a Hewlett-Packard HP 6890 GC, with He as carrier gas.
- **Figure 37** compares the speed of response and detection limit of a standard GC FID and MDD emission by CN at 388.34 nm of a similar set of analytes as in **Figure 36**, but now carried by N_2 flow. Further processing of the data shown revealed MDLs at and below the 1 ppm range. The speed of response of the MDD limit was not taxed in this experiment; it was determined separately that the MDD response time may be even shorter than the 1- to 2-ms timescale available with that setup.
- As analyte concentration increases, so does generally its emission. **Figure 38** shows this for increasing concentrations of CH_4 in N_2, by monitoring the emission at 388 nm due to the formation of CN from N_2 and CH_4 in the discharge plasma. As shown, the increase is less than linear with concentration, and may be caused by the particular chemical kinetics taking place in the plasma.

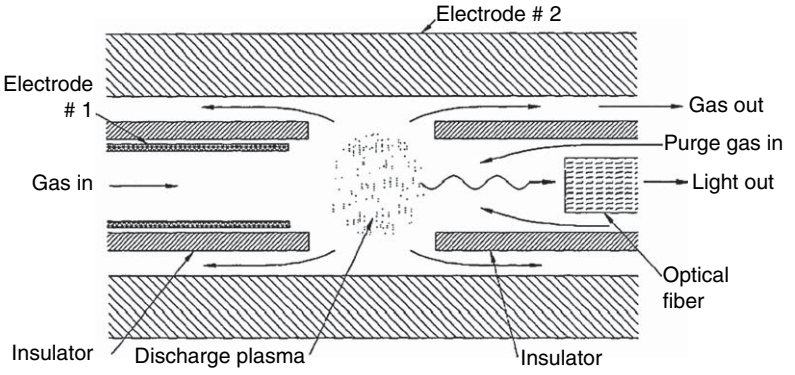


Figure 29 Microdischarge support structure with concentric gas flows (sample and purge gases) and bare electrodes. Operating voltage ~250 V. (Source: Herring C M 2006b Micro-discharge gas detector. *US Pat.* 7 100 421 B1, September 5, 2006, assigned to Caviton Inc.; reproduced with permission.)

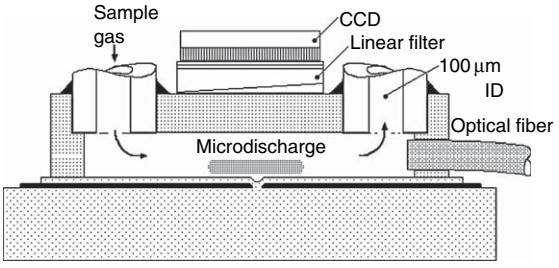


Figure 30 Microdischarge support structure with plug-flow sample gas in one channel, dielectric-covered electrodes, and an attached simple spectral detector array behind a linear filter. (Source: Bonne U, Higashi R, Marta T, Nusseibeh F, Rezachek T (Honeywell Labs), Herring C, Kellner D, Kunze K, Castelein M (Caviton Inc.) 2006a Microgas analyzer for NeSSI and DHS: Measurements and simulations. *PittCon 2006*, Orlando, FL, USA, March 13–16, 2006, Paper #2020-6; reproduced with permission.) Overall dimensions ~10 mm × 10 mm × 5 mm. The optical fiber is optional.

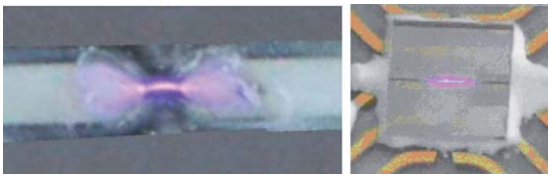


Figure 31 Photographs of microdischarges with dielectric-covered electrodes with ~100-μm gaps. (Source: Bonne U, Higashi R, Marta T, Nusseibeh F, Rezachek T (Honeywell Labs), Herring C, Kellner D, Kunze K, Castelein M (Caviton Inc.) 2006a Microgas analyzer for NeSSI and DHS: Measurements and simulations. *PittCon 2006*, Orlando, FL, USA, March 13–16, 2006, Paper #2020-6.) *Left:* In ambient air, without sample channel, after over 4300 h of continuous operation and *right:* in N_2 flow, inside a 150 μm × 150 μm channel as in **Figure 21**.

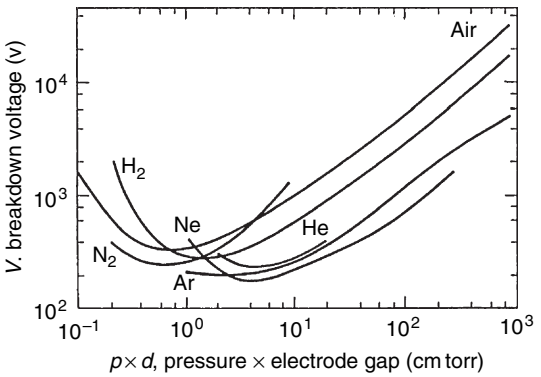


Figure 32 Glow discharge breakdown voltage versus gas pressure and electrode gap, for several gases, known as Paschen curves. (Source: Raizer YP 1991 In: Allen JE (ed.) *Gas Discharge Physics*. Springer-Verlag, Berlin, Heidelberg; reproduced with permission; Herring C M 2006; Kunze K 1997.)

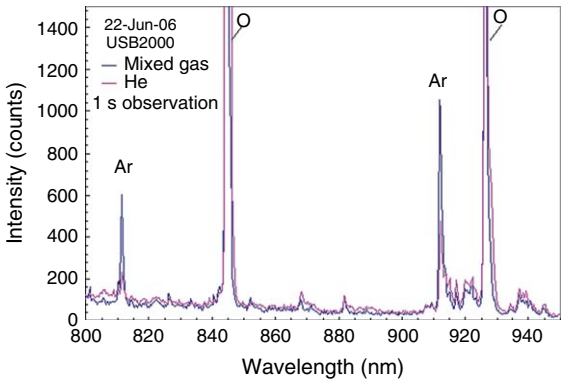


Figure 33 Microdischarge detector (MDD) emission spectrum to show 2.5 ppm level of Ar traces in a special He calibration gas (compliments of Air Products), versus the more pure He reference gas. Recorded with an Ocean Optics USB2000 spectrometer. (Source: Herring C M 2006a Caviton Inc., Private communication, June 2006.)

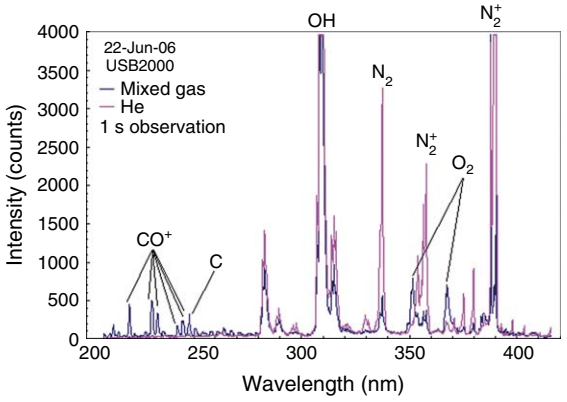


Figure 34 Microdischarge detector (MDD) emission spectrum to reveal the presence of 2.5 ppm CO and CO_2 traces in the He sample gas mixture, compared to pure He reference gas. Recorded with an Ocean Optics USB2000 spectrometer. (Source: Herring C M 2006a Caviton Inc., Private communication, June 2006; reproduced with permission.)

The above points show the versatility, speed of response, sensitivity, multichannel, small dead volume, and compactness of MDDs. Stand-alone detection of NO_x, and ppb level detection of Hg and other metals have been demonstrated (Herring 2006a). Future R&D effort is needed to better understand glow plasma kinetics, which often results in emission output sensitivities to the uncontrolled change in the concentration of gases such as O_2 and H_2O, which at present, are worked out empirically.

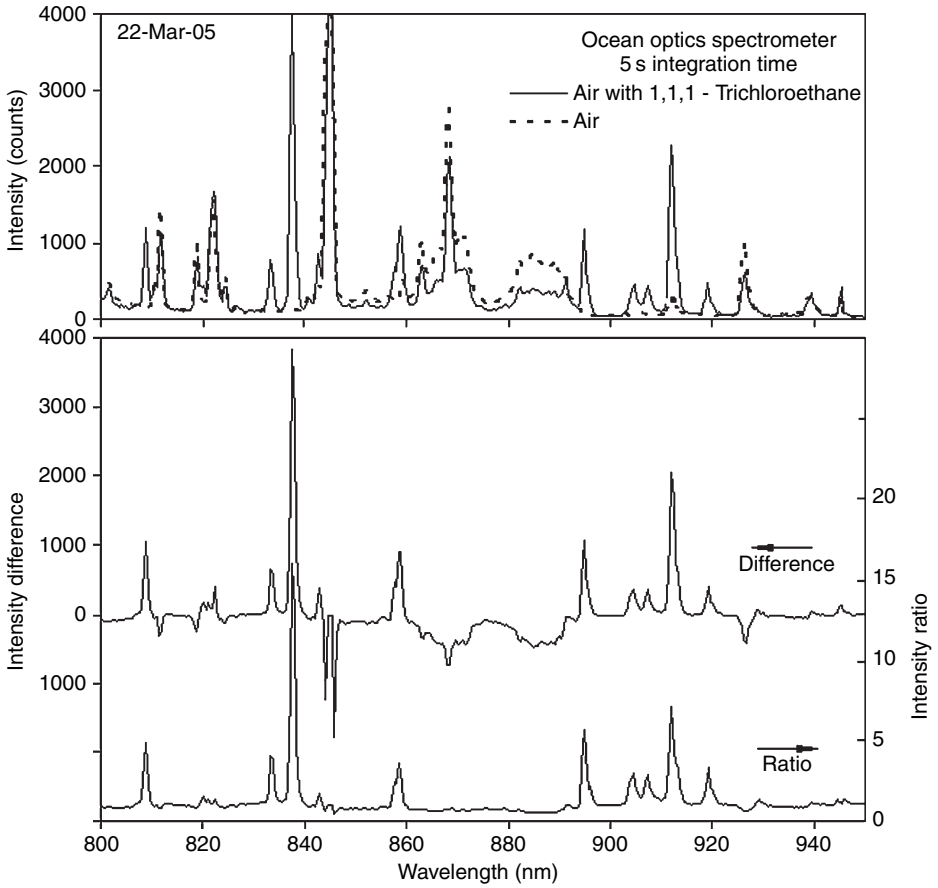


Figure 35 Data processing of a microdischarge detector (MDD) emission spectrum of air with trichloroethane. (Source: Bonne U, Higashi R, Marta T, Nusseibeh F, Rezachek T (Honeywell Labs), Herring C, Kellner D, Kunze K, Castelein M (Caviton Inc.) 2006a Microgas analyzer for NeSSI and DHS: Measurements and simulations. *PittCon 2006*, Orlando, FL, USA, March 13–16, 2006, Paper #2020-6; reproduced with permission. Bonne U, Nguyen T-C, Polla D L 2007 Developments in ultra micro gas analyzers. Chapter 9.3 in: Koch M V, Van den Bussche K M, Chrisman R W (eds.) *Micro-Instrumentation for High Throughput Experimentation and Process Intensification – a Tool for PAT*. Wiley-VCH, Weinheim; reproduced with permission from Wiley-VCH Europe Ltd.). *Top:* plot of unprocessed data of spectral emission of air, with and without tricholoroethane. *Middle:* plot of the difference between the two above spectra. *Bottom*: plot of the ratio of the two above spectra.

2.11.4.3 Mass Spectrometry

One of the most attractive aspects of mass spectrometry is its ability to provide the user with the direct identification of the analytes by way of their MW. The most problematic aspects (preventing their use in portable or handheld gas analyzers) are the requirement of operating at absolute pressures below a millitorr (or below $\sim$1 dyne cm^{-2}, 1 μbar or 0.1 Pa) and the bulkiness (size and weight) of present pumps. Among the different types of MSs in use today (magnetic deflection, time-of-flight, quadrupole, and ion trap) only ion trap MS (ITMS) shall be discussed in more detail, because of its selection to develop a chip-level GC–MS, based on ORNL's ITMS technology (Bonne *et al.* 2004, 2007, Pau *et al.* 2006).

The principle of operation of such ITMSs can be briefly described as follows: after ions of the sample gas are created by any of a number of approaches (electron beam, corona discharge, charge transfer, etc.) and ion groups (small enough to avoid space charging) or pulses are transported into the cavity such as the one shown in **Figure 39**, they are trapped by the quadrupole electric field obtained by the application of just the right AC + DC field amplitude to the end caps and the fundamental radio frequency (RF) signal applied to the ring electrode. The mass spectrum results from scanning the amplitude of this fundamental RF signal, V, which causes sequential ejection of trapped ion masses, whose trajectories become unstable, through the holes in the end-cap

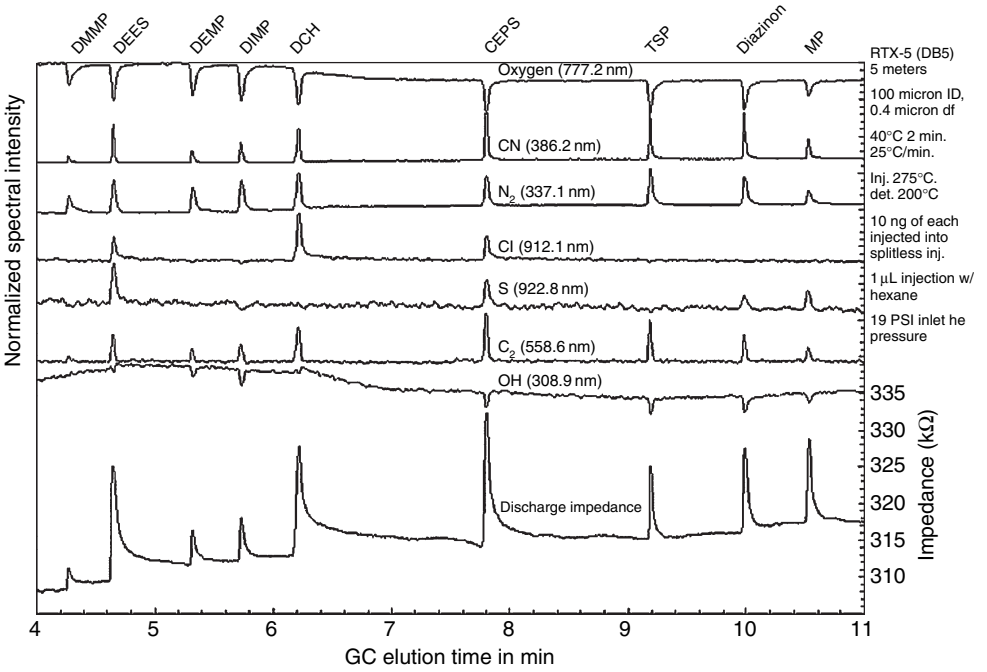


Figure 36 Microdischarge emission outputs of a mixture of analytes, with seven optical wavelength channels, in addition to an electronic channel monitoring microdischarge detector (MDD) impedance, all downstream of a commercial gas chromatagraph (GC) (5-m/100-μm/400-nm DB5 with He carrier gas). (Source: Herring C M 2006a Caviton Inc., Private communication, June 2006; reproduced with permission.)

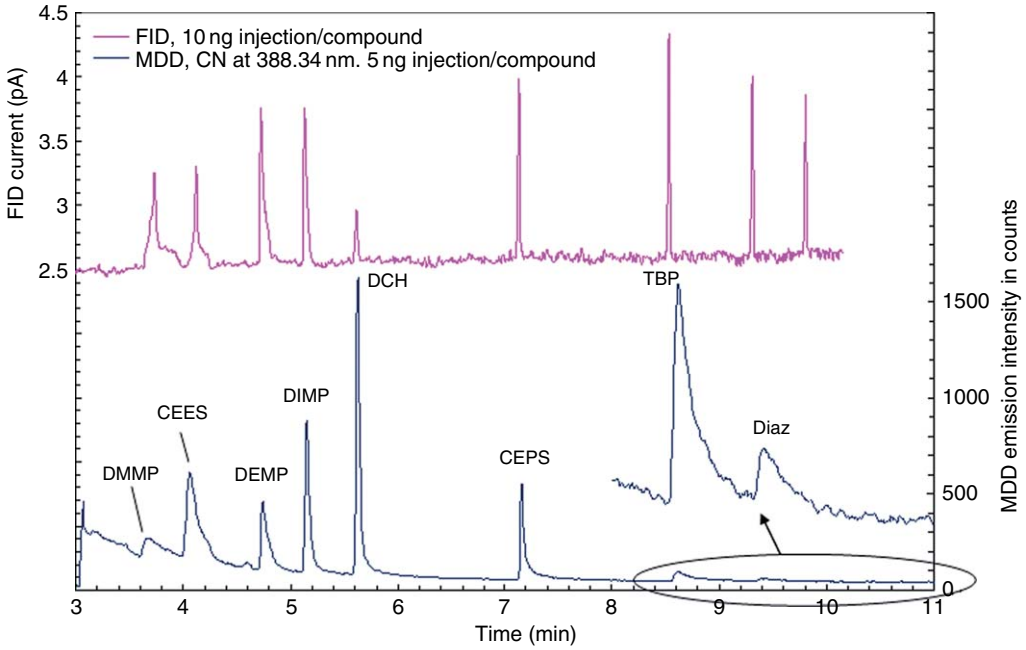


Figure 37 Comparison between flame ionization detector (FID) (*top*) and microdischarge detector (MDD) (*bottom*) outputs for a mixture of analytes separated by a commercial gas chromatagraph (GC) (5-m/100-μm/400-nm DB5 with N_2 carrier gas). (Source: Herring C M 2006a Caviton Inc., Private communication, June 2006; reproduced with permission.)

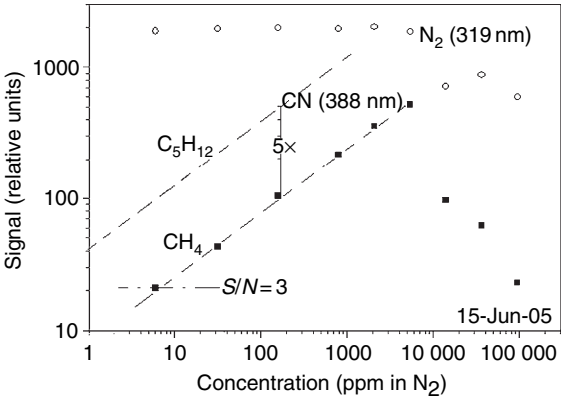


Figure 38 (a) Calibration curve of microdischarge detector (MDD) emission from CH_4 in N_2 in the form of the CN band at 388 nm, demonstrating ppm level detection of CH_4, and (estimated) sub-ppm level for pentane at $S/N > 3$. Observation time: 100 ms. Also shown is N_2 emission at 319 nm for reference. (Source: Herring C M 2006a Caviton Inc., Private communication, June 2006; reproduced with permission. Kunze K 1997 Miniaturized discharges – Prospects and limits for quantitative analysis. Ph.D. thesis, University of Dortmund, Germany; reproduced with permission.)

electrode. The end cap is close to a Faraday cup or electron multiplier (not shown in **Figure 39**), which counts the ejected ions as V is scanned. Textbooks and websites abound with more details, which need not be repeated here.

Achieving mass resolution of 1 atomic mass unit (amu) is generally expected of and achieved by the above-mentioned MSs. A resolution of $\leq$0.25 amu was achieved with a 1-mm ID trap ITMS (see **Figure 39**, which also details the structure of such an analyzer). Such resolution sets ITMSs apart from ion mobility and differential (ion) mobility spectrometry (IMS and DMS), which have achieved the equivalent of mass resolutions of 20 amu (Eiceman *et al.* 2004) and a shorter mass range, although operating at ambient pressures and excelling with achieving low ppb levels of detection of large MW and low ionization energy compounds, which are characteristic of those of interest to civilian security and defense. The size reduction has been and continues to be addressed by many researchers (Akinwande 2004, Blain *et al.* 2004, Ferran and Boumsellek 1996, Pai *et al.* 2005, Pau *et al.* 2006, Whitten 2004), and is tied to mitigating the pressure issue, because increasing pressure reduces the mean free path and forces a reduction of internal MS structures and molecular paths. The increase in gas density associated with increased pressure may increase ionization efficiency for IMS (Eiceman and Karpas 1994) and for ITMS, at least partly making up for the reduced number of ions trapped in the smaller ion trap, besides enabling the fabrication of trap arrays to generate higher S/N outputs, as being pursued by the ORNL–UNC–UA team (Pau *et al.* 2006). Theoretically, an array of n traps with $d = d_o/n$ should give the same performance as an individual trap with a trap diameter of d_o, provided the ring frequency is increased by the same factor of n and yielding an overall reduction of MS volume by a factor of n^2. However, producing the same RF voltages while increasing the frequency has practical limits; yet lower RF voltages and higher frequencies may be used to produce a viable ion trap MS with some distinct advantages (Witten and Ramsey 2002–2006).

To illustrate the ITMS micromachining approach taken by this team, **Figure 40** presents the design and

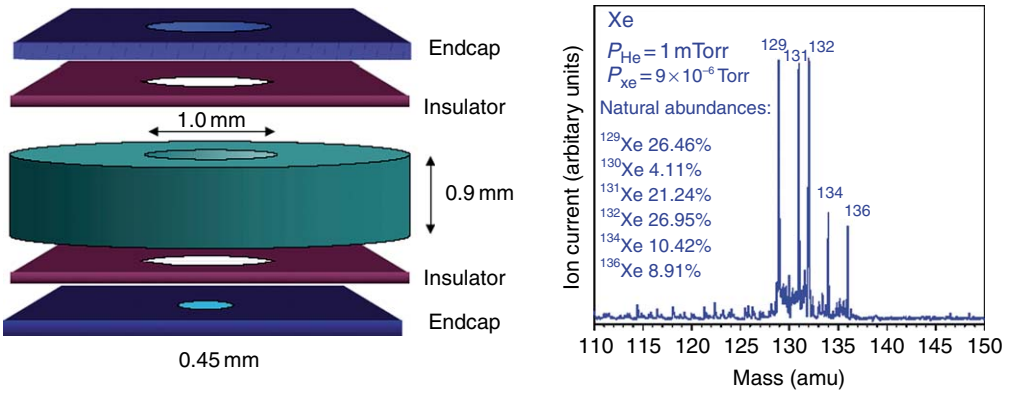


Figure 39 Assembly of a 1-nm ID cylindrical ion trap for an ion trap mass spectrometry (ITMS) (*left*) and obtained resolution of 0.1–0.25 amu of Xe isotopes in 1 mtorr of He, using electron impact ionization and a single scan of ~5 ms for data acquisition. (Source: Kornienko A, Reilly PTA, Whitten W B, Ramsey J M 1999 *Rapid Commun. Mass Spectrom.* **13**, 50; reproduced with permission from Wiley-VCH Europe Ltd.)

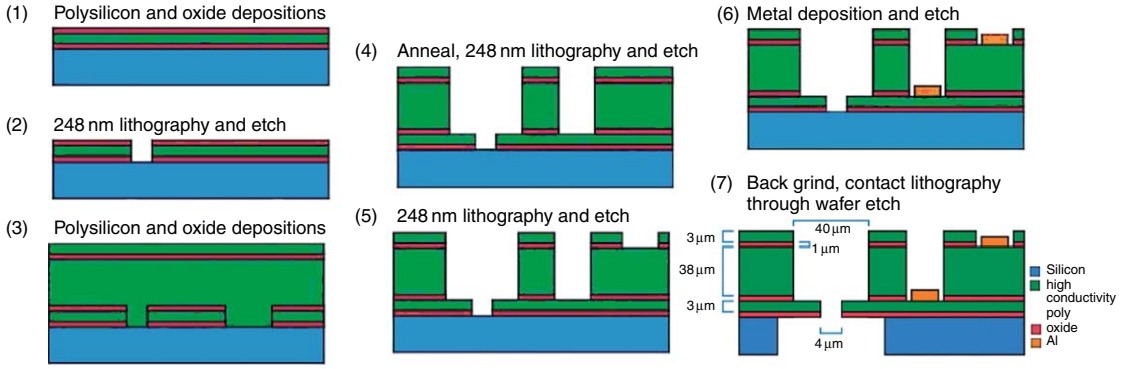


Figure 40 Design and fabrication steps of a 40-μm ID ion trap array, consisting of open cap cylindrical traps. Step 7 also shows the dimensions of the completed trap. (Source: Raizer Y P 1991 In: Allen J E (ed.) *Gas Discharge Physics.* Springer-Verlag, Berlin, Heidelberg; reproduced with permission.)

micromachining steps of a 40-μm ion trap array, which has resulted in measurable ion currents but still exhibits undesirably high capacitance, that lead to impractically high displacement currents (Witten and Ramsey 2002–2006).

The limit in ITMS size reduction and associated pressure increase appears to be governed by the ability to fabricate well-defined and uniform ion traps in the 1- to 10-μm range, and operate in the trapping frequency range of gigahertz (at trap potentials above those corresponding to thermal energy) without incurring excessive displacement currents resulting from the smallest possible capacitance of the ion trap array (Pau *et al.* 2006). If we can assume that the relations between mean free path, trap diameter, 1/(pressure), and 1/(trapping frequency) hold when extrapolating from the millitorr range to ambient pressure, and taking experimental evidence of 1-torr operation with a 1-mm trap at 6 MHz (and 1.7 amu, 3 ppm Xe in He, and 20 amu ms^{-1} scan rate performance) (Witten and Ramsey 2002–2006) as a guideline, one can begin to appreciate the challenges that may still lie ahead. **Figure 41** presents the overall trends in graphical form, showing that operation at

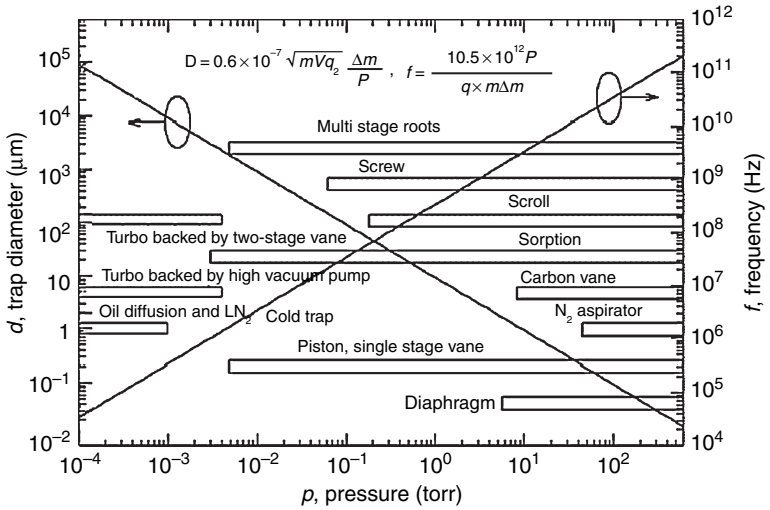


Figure 41 Relationships between ion trap mass spectrometry (ITMS) operating pressure, field frequency, trap radius, and pump capabilities. *m*, atomic mass; *V*, applied trap voltage. (Source: Raizer Y P 1991 In: Allen J E (ed.) *Gas Discharge Physics.* Springer-Verlag, Berlin, Heidelberg; reproduced with permission. Bonne U, Nguyen T-C, Polla D L 2007 Developments in ultra micro gas analyzers. In: Koch M V, Van den Bussche K M, Chrisman R W (eds.) *Micro-Instrumentation in High Throughput Experimentation and Process Intensification*. Wiley-VCH, Weinheim, in press; reproduced with permission from Wiley-VCH Europe Ltd.)

76 torr might require an operational frequency of 30 GHz; however, this frequency would drop to ~460 MHz if one were to extrapolate from the 6-MHz data point at 1 torr.

2.11.4.4 Gas Chromatography and Multidimensional Analysis

To illustrate the essential functions of the type of GC analyzer to be discussed in this section, **Figure 42** shows how in a most general industrial setting, sample gas is conditioned for the analyzer: a filter removes particles and a flow sensor verifies that sample gas from the process is flowing. A pump draws a small fraction of the sample flow through the analyzer, at a rate controlled by the analyzer and not by the main sample flow.

In an effort to standardize on such sample conditioning, an initiative at the Center for Process Analytical Chemistry (CPAC) has led to guidelines for NeSSI (New Sampling and Sensing Initiative) systems, which consist of devices (filters, on–off and modulating valves, flow sensors, and composition sensors and analyzers) mounted on ANSI Std. SP-76 building-block substrates that are networked and in its Generation-2 certified for IS, so that their replacement can be made in the presence of flammable gas mixtures (Bonne *et al.* 2005b, 2006b, NeSSI 2006).

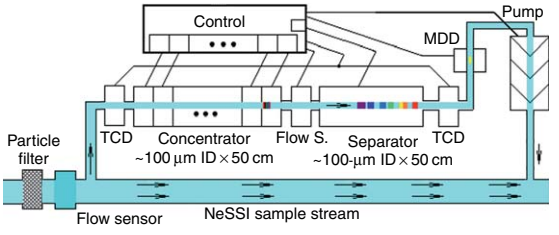


Figure 42 Micro gas analyzer (MGA) functional block diagram, showing a preconcentrator, gas chromatography (GC) separator, detectors, and pump to draw a small flow from the sample stream. (Reproduced with permission from Wiley-VCH Europe Ltd.)

The analyzer itself may consist of a PC (see further details below and Tenax PC tables from Scientific Instrument Services), a GC separation column (or channel in the case of a micromachined GC), its own flow sensor, differential TCD, or other detectors to be discussed below. Not drawn in **Figure 42** but not to be forgotten is the need for a sample gas injection system, which injects a small sample into the (also omitted) carrier gas stream. Those omissions were intentional in **Figure 42** and reflect the ability of the PHASED MGA heater array structure for enhanced detection (PHASED) MGA (Bonne and Higashi 2002, Bonne *et al.* 2004, 2006b) to operate by using air or other background gas as the continuously flowing carrier gas and to generate the injection via rapid thermal pulse to the last element of the integrated PC of **Figure 42**, which desorbs and thus injects the analyte(s) into the separator, which then functions as a traditional GC.

Whereas traditional PCs were based on a single element of packed Tenax or other materials and required watts of heating power to desorb within ~3–10 s, a simple trap in the capillary sample loop (**Figure 43**) (Staples 2004) can provide subsecond desorption pulses as demonstrated by the portable z-Nose GC. An example of its analytical capability is depicted in **Figure 44** (Staples 2004), obtained for the gaseous signature of paper money, and indicating a high peak capacity for the portable GC. **Figure 44** also presents a creative way to present the GC data to

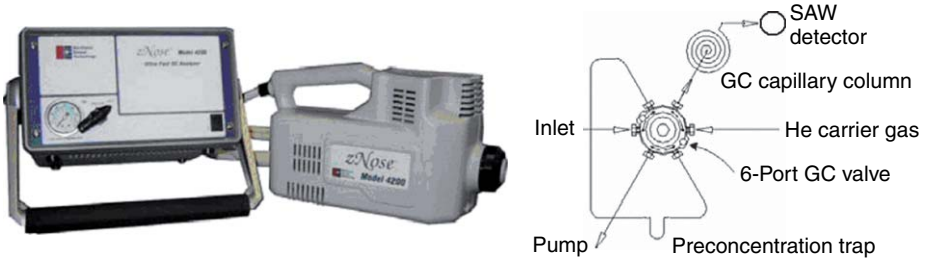


Figure 43 Portable gas chromatography (GC) system (z-Nose by EST) for 10-s chromatograms. Photograph (*left*) and system flow diagram (*right*), showing preconcentration trap and a six-port GC valve. (Source: Staples E J 2004 Chemical profiling cargo with an ultra-high speed gas chromatograph, olfactory images, and virtual chemical sensors. *3rd Microsensors (DHS-ATF)Workshop*, Scottsdale, AZ, USA, April 19–22, 2004 on website, at http://www.estcal.com/TechPapers/CargoContainerOdors.doc and http://www.znose.com/modelsection.htm (Electronic Sensor Technology, Newbury Park, CA); reproduced with permission from IEEE.)

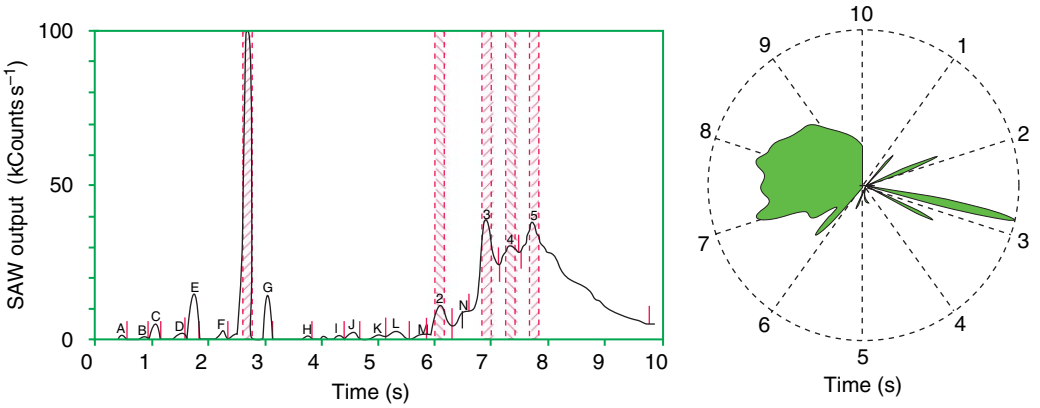


Figure 44 z-Nose gas chromatography (GC) analyzer output of the vapors given off paper money, presented at *left:* as (traditional) surface acoustic wave (SAW) detector intensity versus elution time and *right:* As SAW intensity (radial) versus elution time (angular) olfactory image. (Source: Staples E J 2004 Chemical profiling cargo with an ultra-high speed gas chromatograph, olfactory images, and virtual chemical sensors. *3rd Microsensors (DHS-ATF)Workshop*, Scottsdale, AZ, USA, April 19–22, 2004 on website, at http://www.estcal.com/TechPapers/CargoContainerOdors.doc and http://www.znose.com/modelsection.htm (Electronic Sensor Technology, Newbury Park, CA); reproduced with permission from IEEE.)

a user in the form of an angular plot, generating patterns that may be easier to remember and to recognize (Staples 2004). The speed and creativity of the z-Nose GC stands out among the shoe box-size GCs by traditional GC firms such as Agilent, Shimadzu, Varian, and even Siemens and Yamatake, as described on their webpages, but without getting down to handheld size.

In the quest to leverage the benefits of miniaturization pointed out in Section 2.11.4.1, the achieved advances to date enable handheld GC operation and compatibility with a NeSSI plug-and-play sampling system. Noteworthy in this respect are:

- Sandia's μChemLab, the first hybrid micro-GC, with micromachined ~3-m columns and discrete components connected and supported on a ceramic substrate, also using SAW detectors (Manginell *et al.* 2003). It has performed field tests and is ready for productization.
- SLSs hybrid micro-GC on a $\sim 4'' \times 5''$ circuit board, individually supporting the injector, heatable and Si-etched column, TCD, and data processing and control electronics (Bonne *et al.* 2006a), which is commercially available.
- The University of Michigan's hybrid micro-GC (3-m Si-etched columns), with three types of PC materials, and chemiresistors and nanoparticle films as detector (Lu *et al.* 2005), combined many interesting GC technologies during its development, including polymer films on SAW detectors similar to Sandia's (Manginell *et al.* 2003).
- Honeywell Labs' developmental PHASED micro-GC, which may be the first version of a monolithic, Si-integrated MGA. It has integrated a multistage PC, injector, heatable separation column (0.2-, 0.3-, and 0.5-m versions), and flow sensor, differential TCD, and temperature sensors (Bonne and Higashi 2004, Bonne *et al.* 2004, 2006a, b). The sample, which is preconcentrated from the sample stream, is injected via a thermal desorption pulse, so that no mechanical valving is needed.

Before the above micro-GCs can be merged with other analyzer–detectors to form 2D or 3D analyzers, the design and fabrication of several components need to achieve the needed performance: PC gain and pulse width, carrier gas selection, flow control, sample gas transport (pump), stationary-phase film material selection and processing, and detector speed and sensitivity.

In the following we will discuss such a performance, while borrowing heavily from the author's first-hand experience with Honeywell's PHASED MGA development, but using it mainly as an example to illustrate generic insights. As principal investigator (PI) of that effort since its inception in 1998 through 2006, he tried hard to keep the overall MGA structure and operation as simple as possible, in order to facilitate future productization, by enabling as much fabrication and assembly at the wafer level. In that spirit, we decided to rely on air as carrier gas to eliminate the need for managing and controlling the availability of bottled and pressurized

carrier gas. Similarly, the thermal injection idea was appealing as a way to eliminate the need for mechanical valving.

We realized early on that PC would be needed in order to achieve the less than or equal to ppb-level sensitivities needed even for simple environmental applications. To hold down on energy consumption, our experience with thermal microflow sensor membranes naturally led us to use that technology platform to fabricate heatable PC and separation elements. **Figure 45(a)** shows the cross section of the ~1-μm-thick silicon nitride membrane support (with embedded thin-film Pt heaters) of the stationary phase, a SEM photo of one (of 40–100) turnarounds etched in to the channel wafer (**Figure 45(b)**), and a photo (bottom view, see **Figure 45(c)**) of a completed PHASED MGA chip, with the integrated components listed above. In the first version, we had the PC and separator on separate chips (Bonne and Higashi 2002), but quickly realized the advantage of merging the two, despite the associated implication of then favoring the use of the same stationary phase for both PC and separator.

Figures 46–48 show details of the integrated components: **Figure 46** presents a top view of a Generation-2 PHASED chip layout version, in which all the 20 + 20 PC and separator elements reside on one chip and can be heated individually, although for most uses the separator elements are all energized in parallel. **Figure 47** depicts a magnified view of the TCD, hosted by the ~1-μm nitride membrane, which in turn is anchored to the Si chip wall, which influences the response time of ~0.8 ms (shorter than the ~1 ms of an MB in a 1 m s^{-1} flow (see **Figure 14**) and heat loss, which was reduced by ~2× by improving the thermal isolation (Bonne *et al.* 2006b). **Figure 48** shows the heater and sensing elements of the thermal flow sensor. Its elements are like the one from **Figure 1(a)**, but reduced 2.5× in size to fit onto the MGA channel, together with a calibration curve. Its remarkably low-noise performance may be due to the smaller Reynolds number associated with the >25× narrower flow channel. Individual element heating (**Figure 46**) enables the PC operation, which comprises three steps (Bonne and Higashi 2002, Bonne *et al.* 2005b): while the sample gas flows continuously, the following occur:

(1) Analyte adsorbtion up to a maximum concentration in the stationary-phase film or up to saturation at which point it is in equilibrium with the sample gas, and is achieved in a time ranging from 0.1 to 4 s.
(2) Pulsed desorption, with 3- to 10-ms-long heater pulses and in sync with the flow, starting with the first element, which injects desorbed analyte into the sample stream, followed by the second, third, and $(n-1)$th element when the gas from the previous desorption passes by, thus achieving a gas-phase concentration gain of approximately

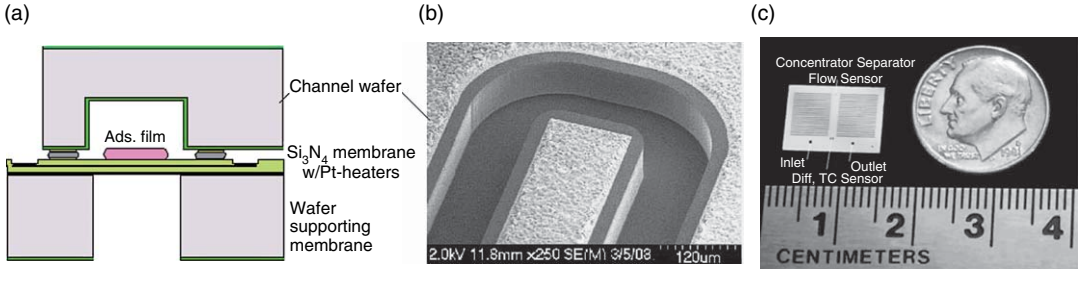


Figure 45 Views of phased heater array structure for enhanced detection (PHASED) micro gas analyzer (MGA) with channels: (a) cross section. (b) Scanning electron micrograph (SEM) photo of top wafer channel turn, and (c) photo of completed chip showing bottom wafer with etched areas to free up the membrane. (Source: Bonne U, Higashi R E, Johnson K, Iwamoto N, Sacks R D, Synovec R E 2005a Stationary phase films for microanalytical measurements. *PittCon 2005*, Orlando, FL, USA, February 27–March 4, 2005, Paper No. 420-1; reproduced with permission. Bonne U, Johnson K, Higashi R, Nusseibeh F, Newstrom-Peitso K, Marta T, Cabuz E, Wang T-Y, Mosher J, Iwamoto N, Lytle N (Honeywell International), Synovec R, Reid V, Prazen B, Gross G, Veltkamp D (Chemistry Department/CPAC, University of Washington, Seattle, WA), Herring C, Kunze K, Kellner D (Caviton Inc., Champaign, IL) 2005b PHASED*: Development of a μGC-on-a-chip with interface to NeSSI. *Sensors Expo*, Chicago, IL, USA, June 6–9, 2005, Session on 2005 DOE-ITP Sensors and Automation Annual Portfolio Review, http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0605/bonne_0605.pdf; reproduced with permission. Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *Proceedings GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002, © 2002 reproduced with permission from Gas Technology Institute.)

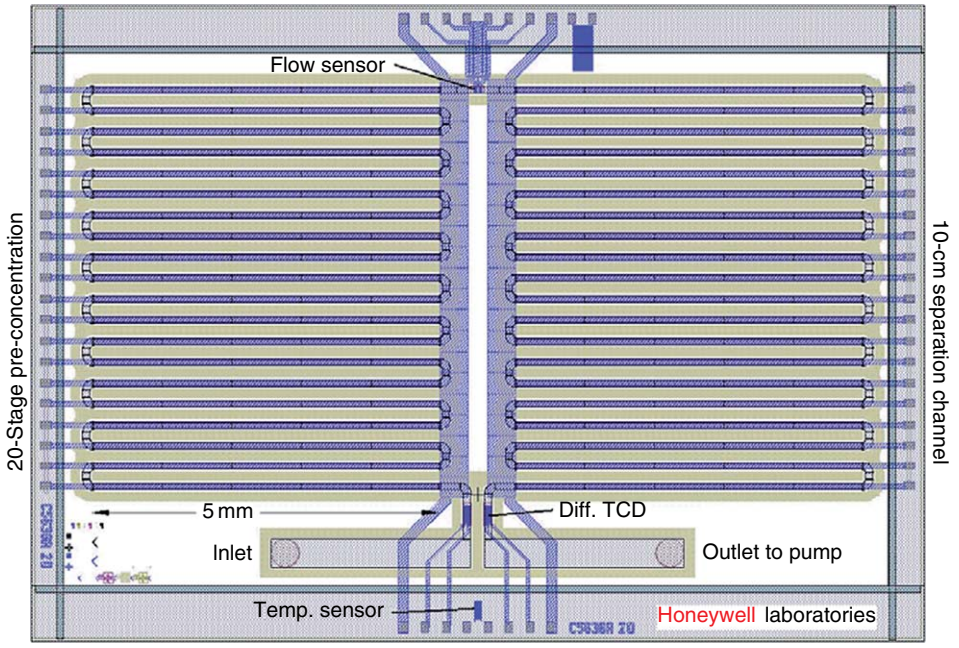


Figure 46 Layout of phased heater array structure for enhanced detection (PHASED) micro gas analyzer (MGA) Generation-2 chip, with integrated 20-stage preconcentrator, flow and temperature sensors, and differential thermal conductivity detector (TCD). (Source: Bonne U, Higashi R E, Johnson K, Iwamoto N, Sacks R D, Synovec R E 2005a Stationary phase films for microanalytical measurements. *PittCon 2005*, Orlando, FL, USA, February 27–March 4, 2005, Paper No. 420-1; reproduced with permission. Bonne U, Nguyen T-C, Polla D L 2007 Developments in ultra micro gas analyzers. In: Koch M V, Van den Bussche K M, Chrisman R W (eds.) *Micro-Instrumentation in High Throughput Experimentation and Process Intensification*. Wiley-VCH, Weinheim, in press; reproduced with permission from Wiley-VCH, Europe Ltd.)

$2\times$ and up to $(n-1)\times$ relative to using only one PC stage, by the time the flow reaches the nth element (Bonne and Higashi 2002). The total achievable concentration gain (**Figure 49**) may range from $10\times$ to $>1000\times$ (see yellow highlights) depending on the number of elements used, the volumetric gas/film ratio, adsorption, and desorption temperatures, and most importantly, on the heat of adsorption of the targeted analyte (Bonne and Higashi 2002, Bonne *et al.* 2006a). The total time to so desorb 20–50 elements may not take more than 60–500 ms.

(3) Pulsed desorption of the nth element, which injects the total PC analyte(s) into the separator, from where separation proceeds as in a traditional GC, except for one feature.

This unique feature is the PHASED coating, which to date has covered only one of the four walls of the approximately square channel. While it would be desirable to have all four sides coated as others do (Lu *et al.* 2003, Manginell *et al.* 2003, Masel 2004, Simonson *et al.* 2004), the penalty in GC resolution, according to Poppe (2002), is only factors of $2\times$ or 1.41, when comparing between fully coated circular on one hand and one- or four-side-coated square channels, respectively. The effect of the associated channel transverse temperature gradient certainly deserves further study. However, coating only one side eases fabrication, as it enables application of the stationary phase via spin-coating and subsequent patterning, so that problems with thick films in channel corners or in cold spots can be avoided. With PHASED element channels, only the middle ~80% of the heatable membrane is coated with the stationary phase, to also achieve a relatively uniform temperature throughout the width of the coated part of the membrane.

Multistage PC gain is of course sensitive to close synchronization between gas flow velocity and heater pulse timing, as shown in **Figure 50**, which Bonne and Rezachek obtained with a constant set of 5-ms pulses to a 20-stage PC, while sample gas (room air) flow varied from 105 to 117 cm s^{-1}, with the optimal value at 115 cm s^{-1} (Bonne *et al.* 2003b). As shown, a

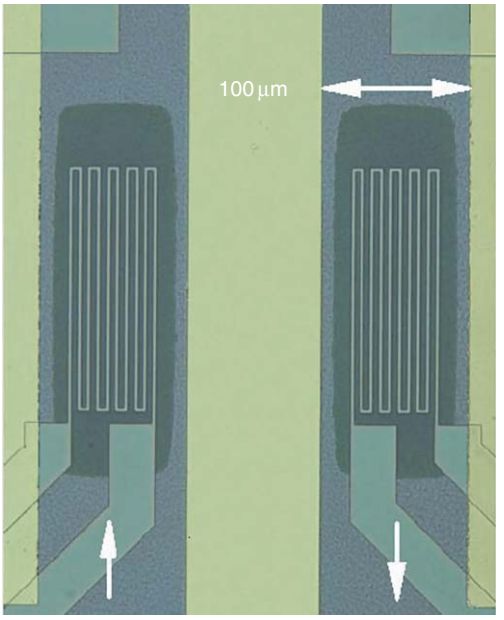


Figure 47 Microphoto of the differential thermal conductivity detector (TCD). (Source: Bonne U, Johnson K, Higashi R, Nusseibeh F, Newstrom-Peitso K, Marta T, Cabuz E, Wang T-Y, Mosher J, Iwamoto N, Lytle N (Honeywell International), Synovec R, Reid V, Prazen B, Gross G, Veltkamp D (Chemistry Department/CPAC, University of Washington, Seattle, WA), Herring C, Kunze K, Kellner D (Caviton Inc., Champaign, IL,) 2005 PHASED*: Development of a µGC-on-a-chip with interface to NeSSI. *Sensors Expo*, Chicago, IL, USA, June 6–9, 2005, Session on 2005 DOE-ITP Sensors & Automation Annual Portfolio Review, http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0605/bonne_0605.pdf; reproduced with permission. Bonne U, Nguyen T-C, Polla D L 2007 Developments in ultra micro gas analyzers. In: Koch M V, Van den Bussche K M, Chrisman R W (eds.) *Micro-Instrumentation in High Throughput Experimentation and Process Intensification*. Wiley-VCH, Weinheim, in press; reproduced with permission from Wiley-VCH Europe Ltd.)

10% velocity mismatch can lead to a twofold width increase of the PC analyte peak, which was sensed by the integrated TCD and only later was identified as water from the sampled room air with the help of a time-of-flight MS (Leco Pegasus-III).

The PC process steps are clearly shown in **Figure 51**, with a sample gas containing 720 ppm of hexane, which shows as the baseline before and after PC, as sensed by the time-of-flight MS, as indicated in the inset, by recording mass 57, a fragment of hexane. Note that the peak half-width (HW) of 91 ms is wider than the ones in **Figure 50**, and is because of the longer transit time to the nonintegrated MS, and also because of nonoptimal synchronization. Later, optimized operation with gas velocities up to 2.5 m s^{-1} resulted in peak widths down to 3 ms and corresponding increases in PC gains to near 100× with hexane (see bold dots in **Figure 49**), over 400× with nonane (Bonne *et al.* 2006a), and higher yet with analytes of higher adsorption enthalpies or retention indices.

The desorption temperature needs to be high enough to desorb all the analytes, and keep them together through the n stages. As **Figure 52** shows for the calculated example of water and heptadecane ($C_{17}H_{36}$), >160°C is sufficient for water vapor, but $C_{17}H_{36}$ needs a temperature ≥250°C.

The choice of carrier gas has some influence on the separation performance of a GC and on the MDL, and is worth some discussion. The widely used Golay equation (Bonne *et al.* 2005a, Kenndler 2004) well describes the relation between capillary length and diameter, coating thickness, carrier gas velocity, pressure drop, and analyte diffusivities in the mobile phase (carrier gas) and stationary phase (coating). However, less well known is the influence of carrier gas on GC resolution, pressure drop, and HW, which are addressed in **Figures 53** and **54**. According to some opinions, using N_2 rather than H_2 as carrier gas limits the useful gas velocity range. Yet, the relative velocity range is the same for both, as **Figure 53** shows, because the three R^* curves would overlap if just displaced parallel to the x-axis, due to the change in the involved gas diffusivities. Despite the lower viscosity of H_2, relative to N_2, the pressure drop needed to achieve the same resolution with H_2 is almost 2× higher. However, separations with H_2 result in chromatograms with peak widths about 2.2× narrower, which is important if the total analysis time is to be made as short as possible.

Figure 54 shows that the retention factor, k', has a sizeable influence on resolution and HW: As k' and retention time increase, resolution and optimal gas velocity drop while peak width increases.

The key ingredient of separation columns is the stationary phase. It is generally taken for granted, because off-the-shelf capillaries are available in different lengths, inner diameters, and coating types and thicknesses from GC supply houses such as Restek, Agilent, and Supelco. Only when facing the steep learning curve of having to select a film material, its solvent, and the polymerization catalyst; ensuring proper wetting and adhesion; and working out a baking routine does one appreciate the involved developers' work and their reluctance to part with the details of their achievements. At start-up, the Honeywell team were

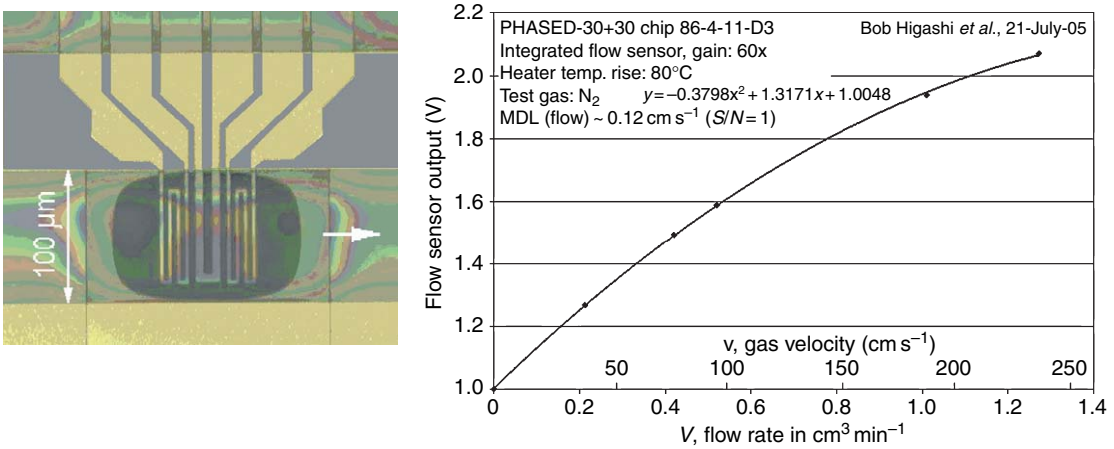


Figure 48 Photo of the integrated flow sensor on a ~1-μm membrane, positioned in the phased heater array structure for enhanced detection (PHASED) micro gas analyzer (MGA) as indicated in **Figure 46**; and its low-noise calibration curve (Source: Bonne U, Higashi R E, Johnson K, Iwamoto N, Sacks R D, Synovec R E 2005a Stationary phase films for microanalytical measurements. *PittCon 2005*, Orlando, FL, USA, February 27–March 4, 2005, Paper No. 420-1; reproduced with permission. Bonne U, Johnson K, Higashi R, Nusseibeh F, Newstrom-Peitso K, Marta T, Cabuz E, Wang T-Y, Mosher J, Iwamoto N, Lytle N (Honeywell International), Synovec R, Reid V, Prazen B, Gross G, Veltkamp D (Chemistry Department/CPAC, University of Washington, Seattle, WA), Herring C, Kunze K, Kellner D (Caviton Inc., Champaign, IL) 2005b PHASED*: Development of a μGC-on-a-chip with interface to NeSSI. *Sensors Expo*, Chicago, IL, USA, June 6–9, 2005, Session on 2005 DOE-ITP Sensors and Automation Annual Portfolio Review, http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0605/bonne_0605.pdf; reproduced with permission.)

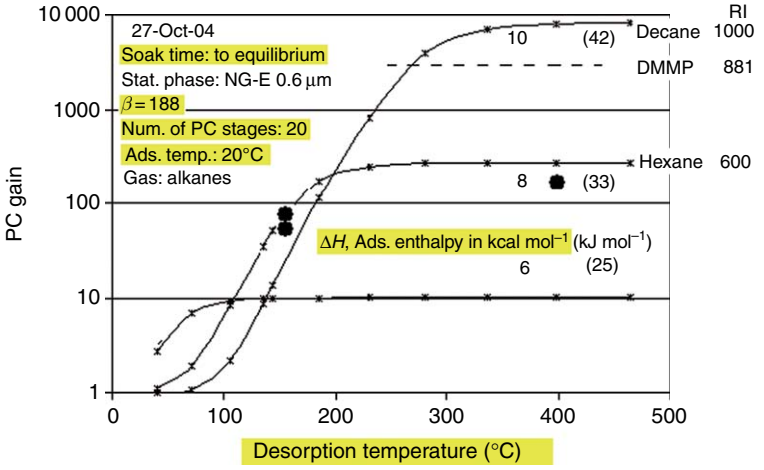


Figure 49 Calculated multistage, saturated, and preconcentration (PC) gain versus desorption temperature and adsorption temperature, showing agreement with a few experimental points for hexane, and expected PC gains with dimethyl methylphosphonate (DMMP) and decane, consistent with their standard gas chromatography (GC) retention indices (RI) (Source: Bonne U, Higashi R, Marta T, Nusseibeh F, Rezachek T (Honeywell Labs), Herring C, Kellner D, Kunze K, Castelein M (Caviton Inc.) 2006a Microgas analyzer for NeSSI and DHS: Measurements and simulations. *PittCon 2006*, Orlando, FL, USA, March 13–16, 2006, Paper #2020-6; reproduced with permission.)

fortunate to find an in-house commercial fabrication capability of open-pore, nanoporous organic silicate (Nanoglass™) films intended to serve as low-k dielectric film to fabricate microelectronic circuits. Synovec *et al.* demonstrated that such Nanoglass films also perform well as GC films to separate alkanes (Huseth *et al.* 204), as shown in **Figures 55** and **56**, relative to a commercial DB-5 column, whether used with H_2 and N_2 as carrier gas, respectively, with astounding, subsecond elution times. As processing expertise grows, more materials may be added such as polydimethylsiloxane (PDMS) from

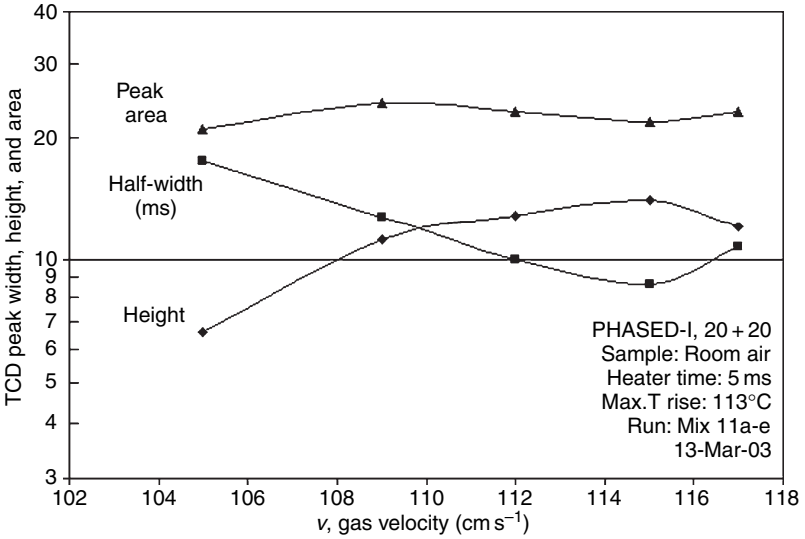


Figure 50 Experimental phased heater array structure for enhanced detection (PHASED) preconcentration peak height and width, showing their sensitivity to proper synchronization between sample gas and heater array pulse progression velocities: the minimum achieved peak half-width is near 8 ms, i.e., excellent, considering the 5-ms widths of the heater pulses, piggybacking desorbed analyte pulses from 20 elements, and peak broadening during separation and thermal conductivity detector (TCD) detection. As shown best with a log plot, the peak area is and should be unaffected. (Source: Bonne U, Higashi R E, Johnson K, Iwamoto N, Sacks R D, Synovec R E 2005a Stationary phase films for microanalytical measurements. *PittCon 2005*, Orlando, FL, USA, February 27–March 4, 2005, Paper No. 420-1; reproduced with permission.)

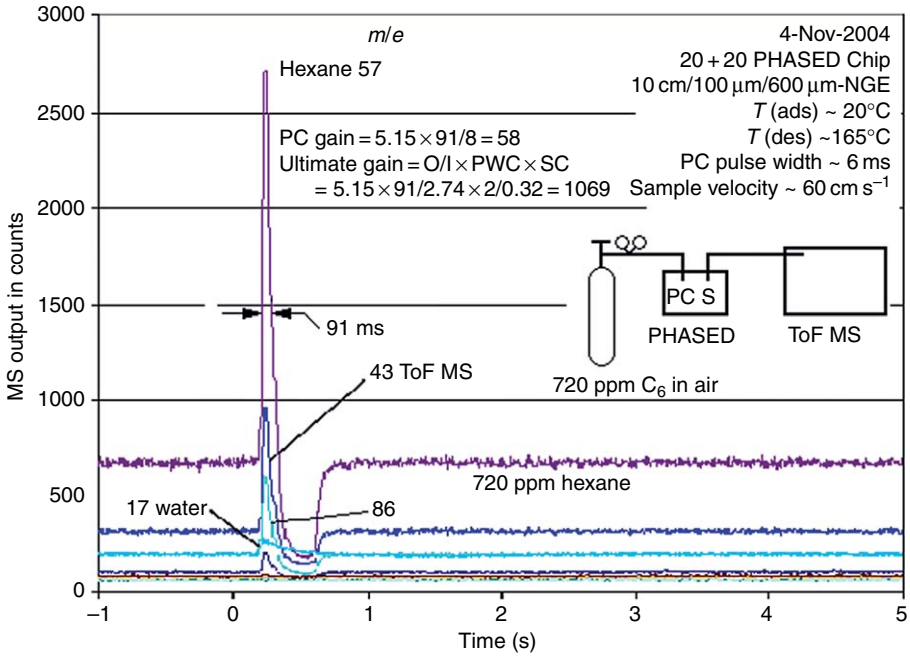


Figure 51 Twenty-stage sampling, desorption, and preconcentration of 720 ppm hexane in air with a phased heater array structure for enhanced detection (PHASED) chip, showing that the adsorption step only takes ~0.3 s or about two sample volumes to refill and saturate the stationary phase with hexane. (Source: Bonne U, Higashi R E, Johnson K, Iwamoto N, Sacks R D, Synovec R E 2005a Stationary phase films for microanalytical measurements. *PittCon 2005*, Orlando, FL, USA, February 27–March 4, 2005, Paper No. 420-1; reproduced with permission from IEEE.)

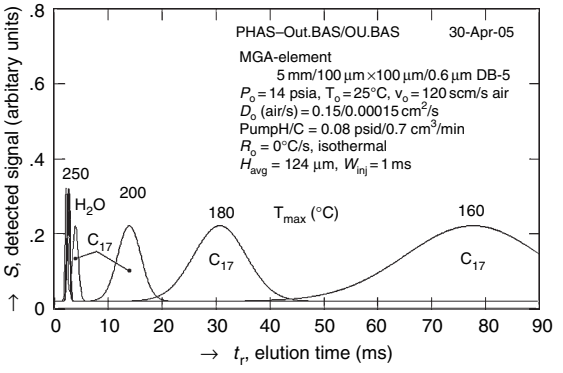


Figure 52 Calculated influence of desorption temperature of a 5-mm phased heater array structure for enhanced detection (PHASED) heater element on the ability to keep all analytes together, represented here by water (H_2O) and heptadecane ($C_{17}H_{36}$). As shown, desorption pulses of ~250°C or higher are needed to desorb and elute heptadecane within the allotted 5–6 ms of each of the 5-mm elements of the multistage preconcentrator, compared to water, for which only a desorption temperature of <160°C is needed. The stationary film thickness was 600 nm of DB-5.

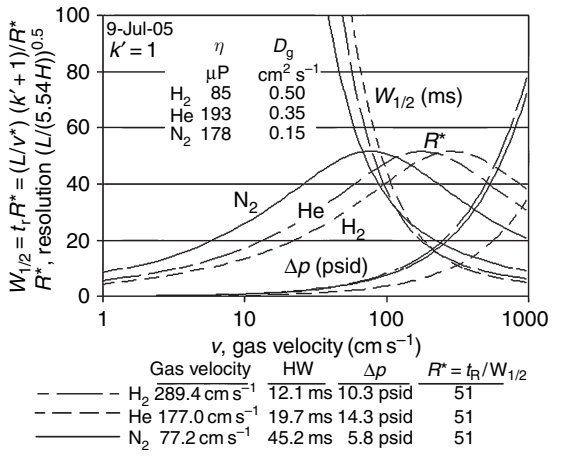


Figure 53 Calculated influence of carrier gas type on best GC separation performance, according to Golay's equation, showing: (1) no change in v-range or resolution, (2) lowest pressure drop for N_2, but (3) lowest half-width for H_2, which is even still 2.2× lower than N_2 at equal pressure drop.

Ohio Valley Inc., carbosilanes from Seacoast Science, and even CNTs (Saridara and Mitra 2005, Stadermann *et al.* 2006). **Figure 57** shows the result of separating alkanes, with temperature ramping, with a first-generation film of ~1-µm-long CNTs grown by LLNL at the bottom of an etched 50-cm-long channel of 100 µm × 100 µm in cross section in silicon, and evaluated by the University of Washington/CPAC (Stadermann *et al.* 2006), who pioneer high-speed GC technology, as needed also for MGAs.

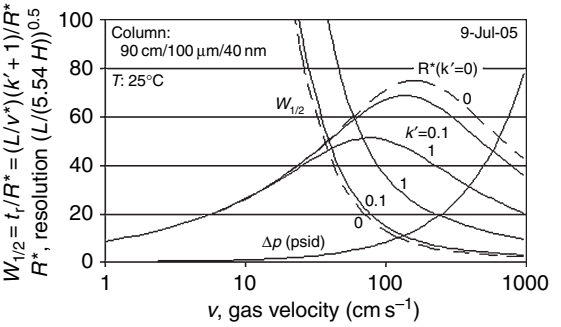


Figure 54 Calculated influence of retention factor $k' = t_R/t_0$-1 on the resolution and half-width of gas chromatography (GC) separation with a N_2 carrier gas stream, according to Golay's equation.

As the quest for a smaller GC analyzer continues, and separation columns shrink below the 1-m length and 100-µm ID size range, it may be good to remember that (Bonne *et al.* 2003):

- Micro-GC peak capacity and analysis time appear as the most important GC performance parameters. As needed analysis times shrink and column lengths shorten below the 1-m mark, peak capacities of only about 20–50 may be achievable with stationary-phase film materials approaching theoretical performance of $H/d = 1$, where H is the theoretical plate height and d is the diameter of the capillary.
- Resolution and temperature ramping: ramping of the separation column increases the range of boiling points of separable analytes but not the full-width peak capacity, C,

$$C = \frac{t_R - t_0}{1.669\, W^*_{1/2}}$$

where the factor 1.669 converts the actual Gaussian (rather than triangular) shape of the GC peaks from average HW, $W^*_{1/2}$, to full-width. Temperature ramping reduces elution time by about 2× for each 16–20°C rise in temperature. The ramp rate should be such that the boiling point temperature of the eluting compound stays in about a constant ratio to the column temperature (Sacks 2004).

- Pressure drop may impose a practical limit to ultimately achievable compactness with ever smaller µGCs, as it limits the column length and thus the achievable resolution (Bonne *et al.* 2003) most of the time. However, pressure drop is predictable via Poiseuille's viscous flow eqn [4], as given in Section 2.11.2.4.

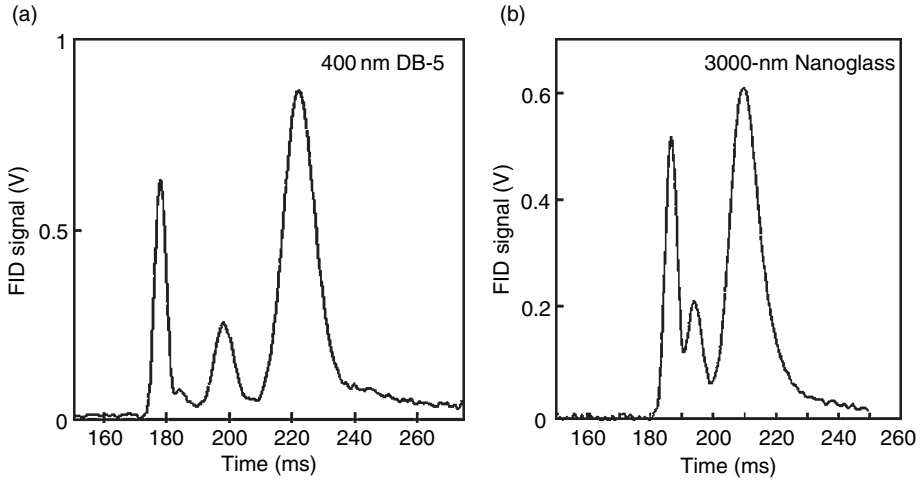


Figure 55 Experimental separation of CH_4, propane, and butane, H_2 carrier gas and flame ionization detector (FID). Comparison between a commercial DB-5 and a developmental Nanoglass film. (a) Column: 1-m/100-μm/400-nm DB-5, ~700 cm s^{-1} H_2 at 25°C. (b) Column: 1-m/100-μm/3000-nm Nanoglass, ~700 cm s^{-1} H_2 at 25°C (Courtesy: G Gross and R Synovec, CPAC; Huseth S, Gonia P, Bonne U 2004 Wireless and sensing solutions advancing industrial efficiency, Part II: PHASED microanalyzer. DOE S&A Annual Review, June 21, 2004, and website at http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0604/presentations/04_honeywell_cps_14227.pdf; reproduced with permission.)

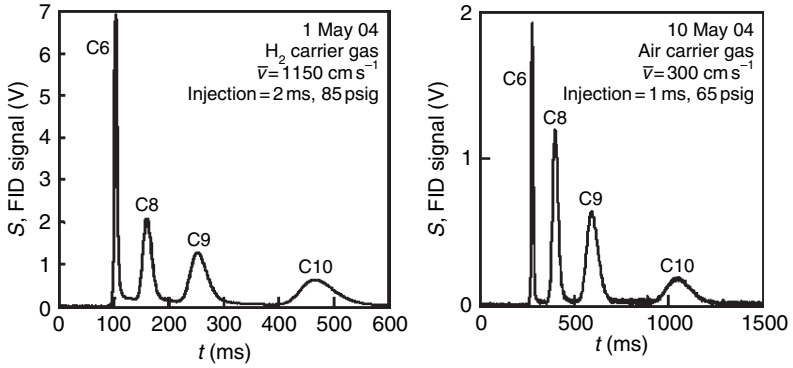


Figure 56 Experimental separation of hexane, octane, nonane, and decane with (a) H_2 and (b) N_2. Capillary #3: 1-m/100-μm/3000-nm Nanoglass at 100° C and flame ionization detector (FID) at 250°C and 20 kHz. (Courtesy: G Gross and R Synovec, CPAC; Huseth S, Gonia P, Bonne U 2004 Wireless and sensing solutions advancing industrial efficiency, Part II: PHASED microanalyzer. DOE S&A Annual Review, June 21, 2004, and website at http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0604/presentations/04_honeywell_cps_14227.pdf; reproduced with permission.)

- Predicting retention time of new analytes is difficult, but approximations may be made on the basis of their vaporization enthalpy or boiling point, as exemplified with **Figure 58** (Bonne *et al.* 2005a). Equating vaporization enthalpy with adsorption enthalpy, ΔH, gets us the retention factor, $k' = (t_R/t_o - 1)$, and retention time, t_R, by leveraging the equilibrium constant, $K = k'\beta$, with $\beta =$ (gas volume/stationary film volume), into the classical relation $\ln(K) = (\Delta H/RT) - (\Delta S/R)$, where, according to Trouton's rule (strictly only for nonpolar compounds), the vaporization entropy is approximately $\Delta S = \Delta H/T_b$ at the boiling point, T_b. From there, the temperature dependence of k' emerges with the remarkably simple van't Hoff's relation, $\log(k'_2/k'_1) = -(\Delta H/R)(1/T_2 - 1/T_1)$, where R is the universal gas constant. Prediction of k' and K may be improved by introducing additional analyte and stationary-phase properties such as their polarizability, dielectric constant (Bonne *et al.* 2005a), and others (Grate 1998, McGill *et al.* 1994).

- Predicting GC peak capacity under given constraints of pressure drop and analysis time may be accomplished iteratively with the help of the expression for HW derived from Golay's

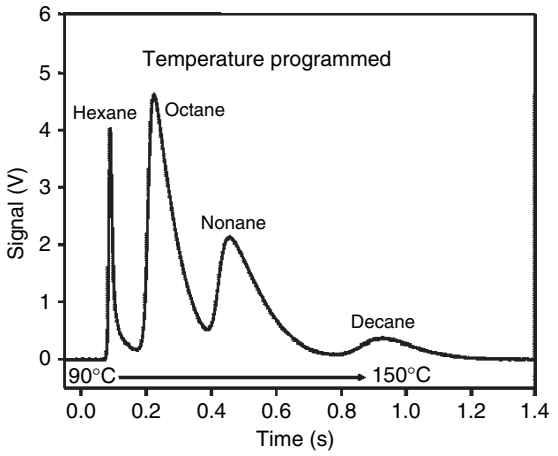


Figure 57 Experimental separation of C_6–C_{10} alkanes on 50-cm/100 μm × 100 μm Si microchannels, coated on one side with single-walled nanotubes (SWNTs) of ~1 μm in length, and temperature ramped from 90°C to 150°C. Achieved peak capacity: ~16 between t_o and 4 s. (Source: Stadermann M, McBrady A D, Dick B, Reid V, Noy A, Synovec R E, Bakajin O 2006 Fast gas chromatography on single-wall carbon nanotube stationary phases in microfabricated channels. *Anal. Chem.* **78**, 5639 (University of Washington and LLNL); © 2006 reproduced with permission from the American Chemical Society. Bonne U, Nguyen T-C, Polla D L 2007 Developments in ultra micro gas analyzers. In: Koch M V, Van den Bussche K M, Chrisman R W (eds.) *Micro-Instrumentation in High Throughput Experimentation and Process Intensification*. Wiley-VCH, Weinheim, in press; reproduced with permission from Wiley-VCH Europe Ltd.)

equation, $W_{1/2} = t_R(5.54H/L)^{0.5}$, and the above retention time and pressure drop equations, assuming some ideal or reference conditions of column geometry, temperature, ramping, gas velocity, and injection pulse width. A few examples shall illustrate this approach. **Figure 59** shows the result of using an experimental gas chromatogram to calibrate a physical GC model, by adopting and incorporating measured retention times and their previously determined temperature dependency. This model then served to also visualize the effect of injection pulse width on achievable peak capacity (**Figure 60**), when the analysis up to dodecane was constrained to be completed within 4 s, whereby only column temperature, length, and inner diameter were held to the indicated values, while adjusting film thickness and pressure drop to the gas velocity of the x-axis. When pressure drop and analysis time were also constrained to $\Delta p \leq 5$ psid and 2 s, respectively, while column diameter was stepwise reduced from 95 to 50 μm, **Figure 61** reveals that smaller columns can generate increased peak capacity. This was unexpected and caused by the increase in peak capacity provided by the reduced zero retention time resulting from the reduced column length.

- Stationary-phase film uniformities of transverse thickness and temperature are the key to success and present significant challenges to micro-GC design and fabrication. For example, a systematic 5% thicker film at the center of a (one wall) stationary-phase coating would potentially increase retention time by 5%, which may theoretically amount to an increase in peak width of 5% × R^*, if diffusion cannot even it out, with the resolution, $R^* = t_R/W_{1/2}$. Transverse temperature gradients may have an even greater effect were it

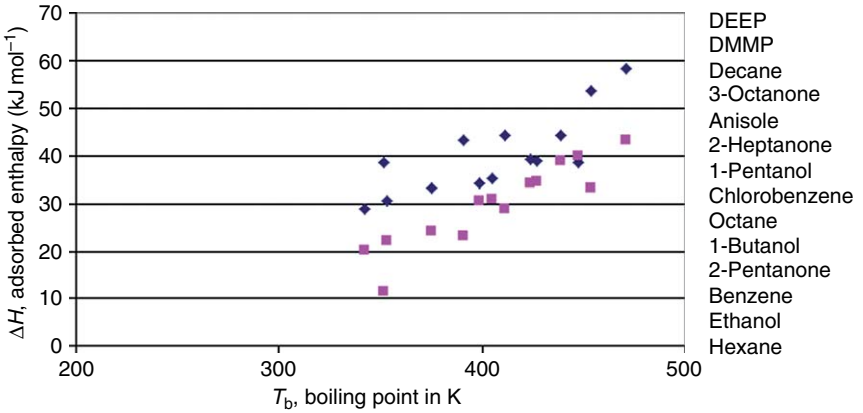


Figure 58 Comparison of experimental adsorption enthalpy of analytes (squares). (Source: NIST WebBook webpage at http://webbook.nist.gov) relative to their experimental vaporization enthalpy (diamonds). (Source: Bonne U, Higashi R E, Johnson K, Iwamoto N, Sacks R D, Synovec R E 2005a Stationary phase films for microanalytical measurements. *PittCon 2005*, Orlando, FL, USA, February 27–March 4, 2005, Paper No. 420-1; reproduced with permission.)

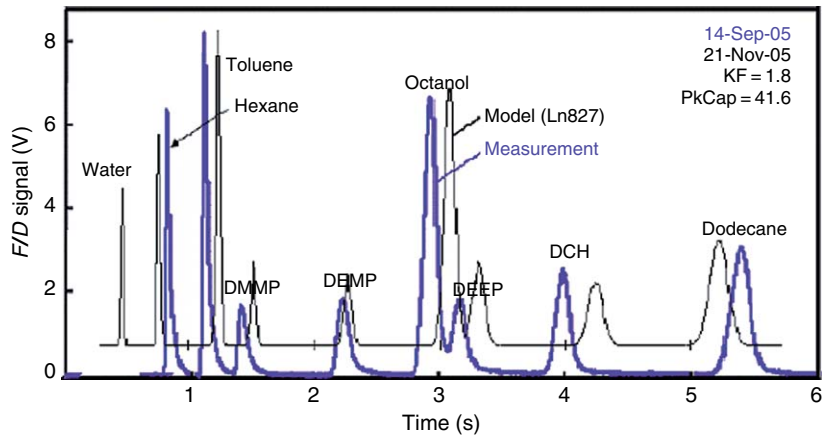


Figure 59 Physical gas chromatography (GC) model (black trace), calibrated with experimental GC retention times on DB-5 by Synovec *et al*. (blue), showing good replication of half-widths. (Source: Bonne U, Higashi R, Marta T, Nusseibeh F, Rezachek T (Honeywell Labs), Herring C, Kellner D, Kunze K, Castelein M (Caviton Inc.) 2006a Microgas analyzer for NeSSI and DHS: Measurements and simulations. *PittCon 2006*, Orlando, FL, USA, March 13–16, 2006, Paper #2020-6; reproduced with permission.) Capillary column: 100-cm/100-μm/400-nm DB-5, isothermal at 125°C, flame ionization detector (FID), 146 cm s^{-1} H_2, and $W_{inj} = 2$ ms.

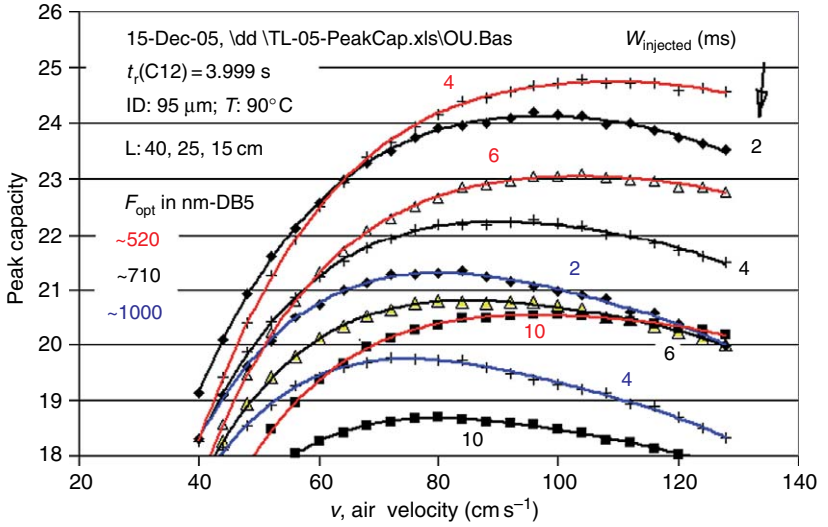


Figure 60 Maximizing peak capacity within the only constraint that analysis time ≤4 s, versus air gas velocity, for four injection pulse half-widths and three column lengths, variable film thickness, and indicated optimal value, F_{opt}. All column diameters at 95 μm at 90°C. DB-5 on one side.

not for the much higher thermal conductivities of the stationary phase and its substrate than that of any gas. **Figure 62** shows that while transverse gas temperature gradients with one-sided column heaters may appear large, the important gradient is the one right at the surface, under the middle 80-μm coating location, which needs to be kept as uniform as possible. Also visualized is the rapid rise in temperature ($\tau \sim 0.8$ ms) of the heater + coating assembly.

As the analytes elute from the micro-GC, they are led to the detectors, single or multichannel, as discussed in Sections 2.11.3, 2.11.4.2, and 2.11.4.3, as rapidly as possible, to minimize peak broadening and loss of concentration and to enable peak identification and quantitation.

As one gathers from the above-mentioned points, new challenges can be tackled with the help of estimates and/or predictive modeling, which all clever developers utilize extensively. **Figure 59** and

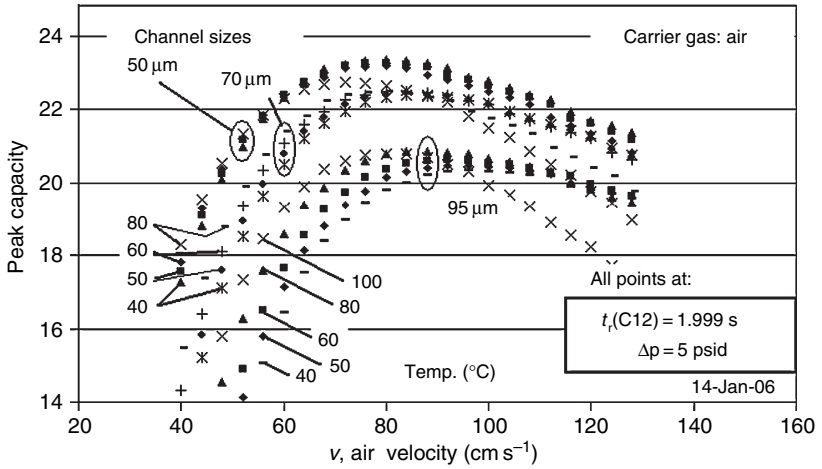


Figure 61 Maximizing peak capacity within the constraints of analysis time $\leq$2 s and $\Delta p \leq 5$ psid, versus air gas velocity, column temperature, and three diameters, variable film thickness, and column lengths. All columns have DB-5 film on one side.

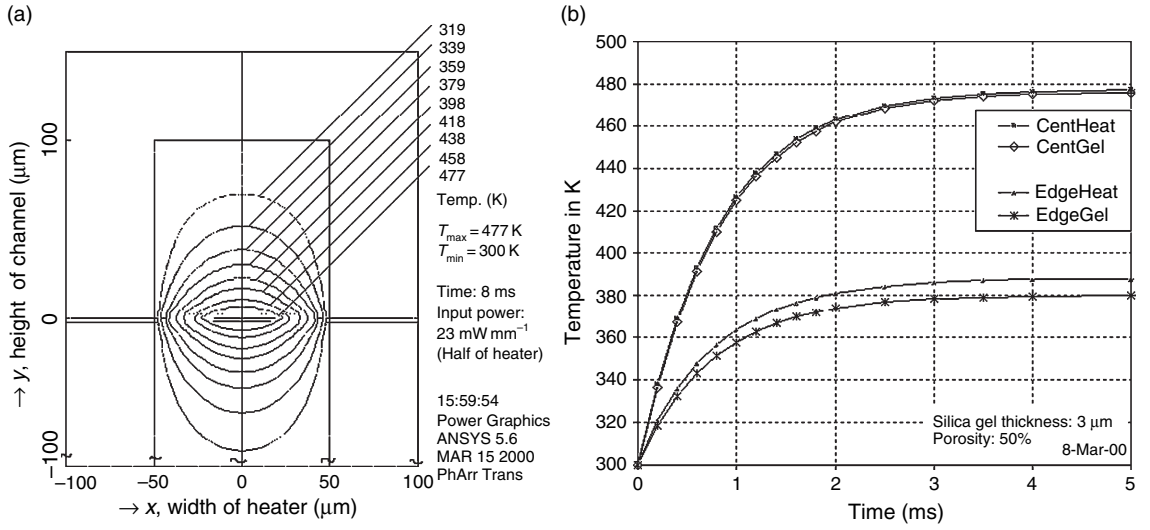


Figure 62 (a) ANSYS-computed transverse gas temperature isotherms in a 100 μm × 100 μm channel over a 100-μm-wide thin-film heater with stationary phase, at zero flow, 8 ms after start from cold, and constant 46 mW mm^{-1} input power to achieve a 200°C steady-state heater temperature. (b) Temperature rise versus time for the same heater, amounting *t* an exponential time constant of 0.8 ms. (Courtesy: Burgass Johnson. Bonne U, Detry J, Higashi R, Newstrom-Peitso K, Pham H, Rezacheck T, Swanson S 2003 PHASED, a faster, smarter and more affordable gas analysis device – Update. *17th Int. Forum Process Analytical Chemistry (IFPAC)*, Scottsdale, AZ, USA, January 21–24, 2003; reproduced with permission. Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *Proceedings GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002 © 2002 reproduced with permission from Gas Technology Institute.)

additional comparisons between simulations and experiment at other temperatures (Bonne *et al.* 2005a) had further validated the model, so that predictions could be made with confidence, for different geometries, gas types, and temperature programming. **Figures 63** and **64** show two simulated results of PHASED geometries and air as carrier gas, which achieved peak capacities >20, with much shorter channels than the 100 cm simulated in **Figure 59** and achieving shorter analysis times. Whether micromachined GCs will be able to compete with the somewhat higher peak capacity, capillary-based micro-GCs, only time will tell. But those that can leverage the potential cost savings of wafer-level fabrication and assembly would surely increase their odds.

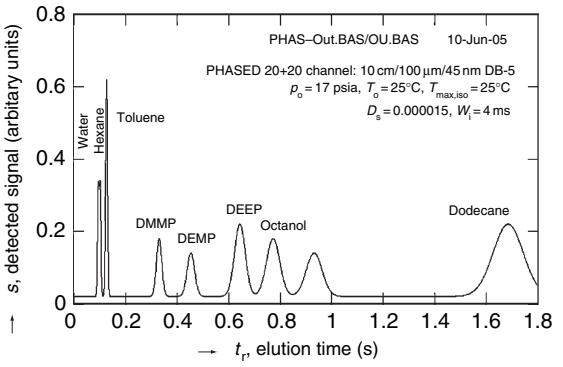


Figure 63 Simulated separation of nine analytes up to dodecane with a 20-element, 10-cm phased heater array structure for enhanced detection (PHASED) channel, over a thin 45-nm film of DB-5, in an 120 cm s^{-1} air carrier gas, eluting in less than 2 s, at room temperature, and with $W^*_{1/2} = 40$ ms resulting in a peak capacity of

$$C = \frac{t_{R,\,C12} - t_0}{1.669\,W^*_{1/2}} = 24$$

(Source: Bonne U, Nguyen T-C, Polla D L 2007 Developments in ultra micro gas analyzers. In: Koch M V, Van den Bussche K M, Chrisman R W (eds.) *Micro-Instrumentation in High Throughput Experimentation and Process Intensification*. Wiley-VCH, Weinheim, in press; reproduced with permission from Wiley-VCH, Europe Ltd).

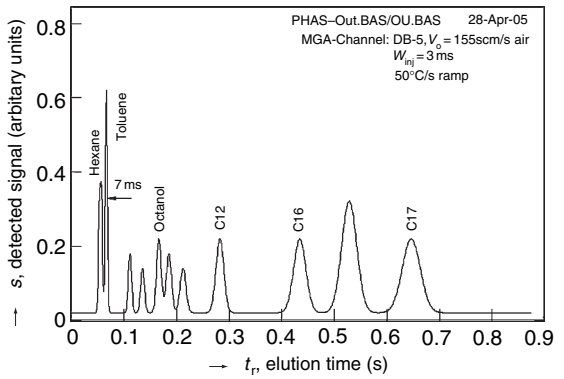


Figure 64 Simulated separation of 11 analytes up to heptadecane in an air carrier gas, eluting in less than 1 sowing to temperature ramping. (Source: Bonne U, Higashi R, Marta T, Nusseibeh F, Rezachek T (Honeywell Labs), Herring C, Kellner D, Kunze K, Castelein M (Caviton Inc.) 2006a Microgas analyzer for NeSSI and DHS: Measurements and simulations. *PittCon 2006*, Orlando, FL, USA, March 13–16, 2006, Paper #2020-6; reproduced with permission. Bonne U, Johnson K, Higashi R, Nusseibeh F, Newstrom-Peitso K, Marta T, Cabuz E, Wang T-Y, Mosher J, Iwamoto N, Lytle N (Honeywell International), Synovec R, Reid V, Prazen B, Gross G, Veltkamp D (Chemistry Department/ CPAC, University of Washington, Seattle, WA), Herring C, Kunze K, Kellner D (Caviton Inc., Champaign, IL,) 2005b PHASED*: Development of a μGC-on-a-chip with interface to NeSSI. *Sensors Expo*, Chicago, IL, USA, June 6–9, 2005, Session on 2005 DOE-ITP Sensors and Automation Annual Portfolio Review, http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0605/bonne_0605.pdf; reproduced with permission.)

2.11.5 Errors, Uncertainty, Self-Calibration, and False Positives

Everyone is familiar with the notion of margin of error or uncertainty, which may be expressed as a ±percentage or absolute peak-to-peak or RMS deviation from a sensor output. However, this deviation may consist of a number of components, such as random white noise or uncertainty, systematic or true error relative to a calibration standard, and drift versus time. Note that precision is often understood to be equivalent to repeatability, while accuracy is in the context of absolute error.

Based on our experience, we recommend that graphical representations of sensor parameter noise and drift be used whenever possible. As mentioned in relation to **Figure 10**, we found it extremely helpful, revealing, and educational to visualize noise amplitude and drift of all relevant inputs and outputs of a sensor system. The use of a $\log(y_i)$ versus time plot is preferred because relative scale for errors, noise amplitude, or uncertainty is the same all along the $\log(y_i)$ axis, as indicated by the error bar of 1% in **Figure 10**. One may use, e.g., factors of $C_i = 10^{0.1i}$ to equally space eight traces within one decade, and magnify the noise and drift, e.g., by a factor of 20× by plotting the parameters of interest, P_i, as $y_i = C_i(P_{i,n}/P_{i,1})^{20}$, where $P_{i,1}$ is the first value of each of the plotted parameters. For easy reference, the dotted line under each trace indicates the no drift level. **Figure 10** shows some drift in the test temperature, but its effect is compensated for and not causing drift in the other parameters, as intended, except maybe in c_{pv}, the volumetric specific heat.

Another helpful measure of the stability of a sensor output signal against variability in its inputs is the notion of log sensitivity (also called gauge factor), which simply indicates the relative change in output as a function of a relative change in one or all of its inputs. This is especially helpful when comparing, ranking, and selecting one correlation from among several possible ones, as is generating the correlation with one set of data (i.e., the training set) and reserving another set for testing the quality of the correlation.

To be sure that measured quantities are trustworthy, all sensors need recalibration after some period of time of operation or even of shelf time, which may be hours or months. In Section 2.11.2.4 we discussed self-normalization of flow sensors via an online, stable, and trustworthy flow–dither approach. The chemical industry cannot afford to manually and periodically recalibrate all sensors and analyzers to ensure proper control of the processes and the quality of product. It therefore adopted automatic recalibration by periodically switching from sample gas exposure of the analyzer to calibration gas and even to span gas. Automatically verifying sample, calibration, and span gas flow, and conditioning it via a standardized, building-block-type, networked, and (soon) intrinsically safe-certified set of components is the aim of the systems being developed with the processing and semiconductor industries (Bonne *et al.* 2006a, Mosher *et al.* 2003, NeSSI 2006). **Figure 65** shows a photograph and illustrates the makeup of such a NeSSI system (Mosher *et al.* 2003), some exemplary devices (Doe 2002–2006), and how it hosts microsensors and microanalyzers.

The above is easy to understand and straightforward, especially when only single or a few known measurands, outputs, or analytes are involved. Defining uncertainty is much more difficult if a sensor output is expected to quantify and to discriminate between the presence and the absence of a target measurand such as a flame in a red-hot combustor (Zook *et al.* 2000), a pathogen in a patient, a type of aircraft from a blip on a radar screen, or one of many possible toxic compounds in air calling for evacuation of a building or processing plant. To generically characterize a sensor's probability to miss a hazardous pattern or its probability to alarm when none is present (FAR) is difficult, especially because of the FAR dependence on the environmental conditions such as temperature, pressure, humidity, and other gases present, as well as the analyte concentration (to be discussed later). Some conceivable approaches to solve this problem of selectivity and discrimination are as follows:

(1) Test the sensor or analyzer under all likely test conditions and experimentally determine FAR for all those conditions.
(2) Test the sensor under a standard set of test conditions and define a FAR under such standard conditions.
(3) Establish a relation between FAR and the analyzer's orthogonal channel capacity (OCC), similar to peak capacity for GCs, except that channels may now not only be resolvable time periods or GC peaks but also resolvable wavelengths, amu, or temperature increments (with DTA) (Bonne *et al.* 2003c, 2007).

None of the above should be confused with MDL. Approaches (1) and (2) may require over 10 000 tests for each analyte, whereas approach (3) is a more rational one, more appealing, and arguably easier to check and recalibrate. Maybe this can also be a more practical approach, if we can define channel capacity in a way that is relevant, i.e., by counting only those channels that can statistically make a meaningful contribution, by, e.g., avoiding to counting wavelength or mass channels that are unlikely to ever detect a relevant photon wavelength or ion mass, other than random noise. Such an approach may then be in line with the work by NRC (Committee on Assessment of Security Technologies for Transportation 2004), in which information content or informing power P_{inf} (in the form of available bits of information) of IMS and MS analyzers was related to transportation security. In comparing OCC and P_{inf}, the latter also assigns capacity counts to each channel for the number of resolvable intensity levels of an analyzer, so that the P_{inf} values are roughly 100× higher than OCC values (Bonne *et al.* 2007).

To relate OCC to FAR (independent on analyte concentration), one might start with a few exemplary OCC values:

- 5–6 for an array of thin-film polymer or SAW sensors (Bonne *et al.* 2003),
- 14–20 for a DMS (Eice man *et al.* 2004),
- 50 for a micro-GCs (Bonne *et al.* 2003), but 200 for a desktop GC,
- 200 for a micro-MS with a 1 amu resolution and a mass range from 20 to 220 amu,
- $50 \times 200 = 10\,000$ for micro-GC–MSs, but 60 000 for desktop GC–MS, the gold standard for analyzers (Polla and Nguyen 2004),

and then consider the possible or likely scenarios in which each orthogonal channel functions as an equation that is part of a system to resolving a problem with n unknowns (analytes and interferents). At its worst, only a fraction of the OCC contributes to solving the problem. But at its best, some combination of channels less than n may uniquely identify an analyte, so that FAR may be much smaller than $1/n$. These and other possibilities involving receiver operating characteristics (ROC) curve calculations are discussed

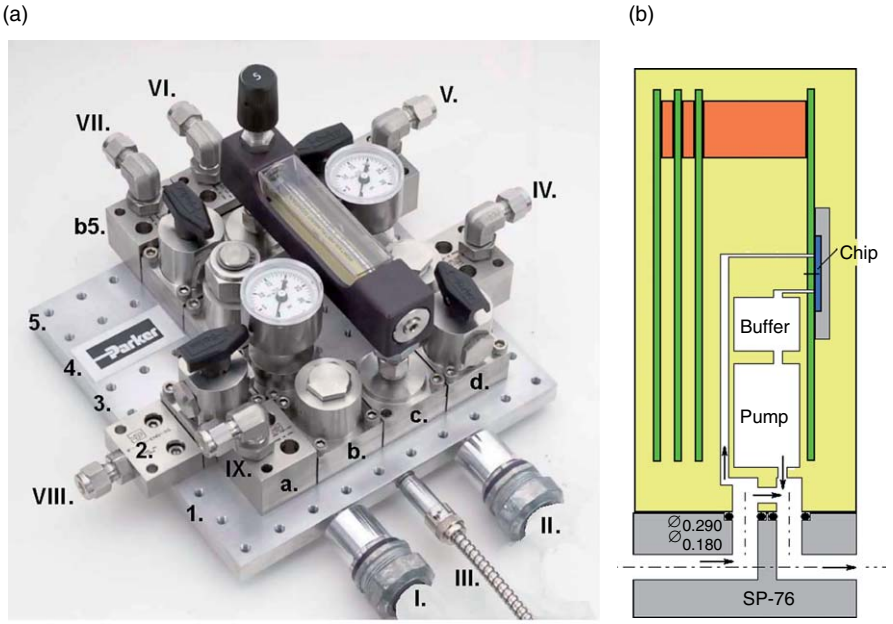


Figure 65 *Left:* Photo of an Intraflow process sampling and sensing system (NeSSi) (New Sampling and Sensing Initiative) (Source: Doe S 2002–2006 Parker Hannifin, private communications; reproduced with permission.), showing standardized devices (pressure gauge, valves, etc.) supported and interconnected on standard ANSI SP-76 substrates of 1.5″ × 1.5″ (38.1 mm × 38.1 mm) footprint. One or more such substrates can host microanalytic devices, such as the analyzer drawn at right. (Source: Mosher J, Nickels R, Bonne U 2003 Networked sampling system (NeSSI-Generation-II) development and field test. *17th Int. Forum Process Analytical Chemistry (IFPAC)*, Scottsdale, AZ, USA, January 21–24, 2003; reproduced with permission.), shown with its own flow control and PHASED MGA chip, corresponding to the block diagram of **Figure 42**. The exemplary devices on the intraflow unit are as follows. Inputs and outputs: 1, sample inlet; 2, sample bypass. Return: 3, from grab sample cylinder; 4, to grab sample cylinder; 5, sample outlet to analyzer; 6, N_2 purge inlet. Devices: a, sample inlet valve; b, check valve; c, sample pressure indicator; d, bypass filter; e, bypass rotometer; f, grab sample cylinder inlet valve; g, grab sample cylinder pressure indicator; h, grab sample cylinder return valve; I, N_2 purge inlet valve.

in Bonne *et al.* (2003, 2007). Together with other plausible considerations, the middle-of-the-road conclusion was a (nonconcentration-dependent) FAR value for the GC–MS gold standard of $\sim 2 \times 10^{-7}$. The above has been based on considering only that part of FAR that depends on the analyte signal clutter contribution, which assumes that enough concentration of analyte is available, so that no MDL effects come into play. But FAR also depends on analyte concentration and therefore on the PC function, which makes a full FAR characterization of a sensor or analyzer more involved. One approach to overcome this problem would be to only accept signals above a set S/N ratio, above the MDL.

Acknowledgments

The author is grateful to the devoted, diligent, and creative sensors and MGA research teams at Honeywell Labs (**Figure 66**), who were active in 2006, and those who participated earlier: Rudy Hegel, Bob Matthys, and Dave Kubisiak led the microbridge product design effort as did Richard Gehman, or supported our MGA effort as external collaborators: Rob Synovec (CPAC/University of Washington), Richard Sacks (University of Michigan), Mike Ramsey (University of North Carolina), Stanley Pau (University of Arizona), Max Lagally (University of Wisconsinn), Peter Griffiths (University of Idaho), Bruce Warmack (ORNL), and Olgica Bakajin and Michael Stadermann (LLNL). Special thanks go to those who directly supported this publication with inputs, support, and encouragement: Patrick Hogan, John Chapples, Keith Pratt and Jerry Evans (Honeywell Analytics), Bill Whitten (ORNL), Cy Herring (Caviton Inc.), Yogesh Gianchandani (University of Michigan), Honeywell management team members Kris Fredrick, Gregg Swenson, and Aravind Padmanabhan as well as Thom Reitz (AFRL) for their permission and clearance to publish, and especially to my family, whose

Figure 66 The core PHASED MGA Team, from left to right: J Skogan, E Cabuz, L Hilton, T Rezachek, U Bonne, J Detry, T Marta, K Newstrom-P, S Poisson, F Nusseibeh, K Nguyen, and R Higashi. Not available for the photo were N Iwamoto, T-Y Wang, R Ellman, K Johnson, S Swanson, R Hegel, B Johnson, T Rolfer, and D Kubisiak. (Source: Bonne U, Higashi R, Marta T, Nusseibeh F, Rezachek T (Honeywell Labs), Herring C, Kellner D, Kunze K, Castelein M (Caviton Inc.) 2006a Microgas analyzer for NeSSI and DHS: Measurements and simulations. *PittCon 2006*, Orlando, FL, USA, March 13–16, 2006, Paper #2020-6; reproduced with permission.)

understanding and support (my son and text editor Marc Bonne helped with final manuscript proofing) were key for the completion of the above text and the generation of its contents over recent decades.

The work described above was enabled thanks to funding by Honeywell, DOE-OIT, and DARPA-MTO, which is gratefully acknowledged.

References

Akinwande T 2004 Field ionization arrays for micro gas analyzers. DARPA-MTO, webpage at http://www.darpa. mil/mto/mga/summaries/2004_summaries/mit.html

Aliwell S R, Halsall J F, Pratt K F E, O'Sullivan J, Jones R L, Cox R A, Utembe S R, Hansford G M, Williams D E 2001 Ozone sensors based on WO_3: A model for sensor drift and a measurement correction method. *Meas. Sci. Technol*. **12**, 684–90

Altemark A et al. 1978 *Combustion-Less Determination of Heating Value of Natural Gases*. IGRC, Toronto, Canada

Application notes by City Technology Ltd (Div. of Honeywell), Walton Road, Portsmouth, Hampshire, PO6 1SZ, UK

Bamford D J, Cook D J, Allen M G 2003 *Broadly tunable mid-IR hydrocarbon sensor. DOE Sensors and Automation, Annual Review Meeting*, San Francisco, CA, USA, June 4, 2003, http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0603/Bamford_03.pdf (Physical Sciences Inc.)

Barrett J J 1976, 1997 Infrared gas analysis. *US Pat.* 3 939 348, February 17, 1976, and 4 035 643, July 12, 1977, assigned to Allied Chemical Co. (now Honeywell)

Barrettino D, Graf M, Zimmermann M, Hagleitner C, Hierlemann A, Baltes H 2004 A smart single-chip micro-hotplate-based gas sensor system in CMOS-technology. *J. Analog Integrated Circ. Signal Process*. **39/3**, 275–87 (ETH Zurich)

Bessoth F G, Naji O P, Eijkel J C T, Manz A 2002 Towards an on-chip gas chromatograph: the development of a gas injector and a dc plasma emission detector. *J. Anal. Atom. Spectrom*. **17**, 794–9 (Imperial College)

Blain M G, Riter L S, Cruz D, Austin D E, Wu G, Plass W R, Cooks R G 2004 Towards the hand-held mass spectrometer: design considerations, simulation, and fabrication of micrometer-scaled cylindrical ion traps. *Int. J. Mass Spectrom*. **236**, 91

Boger, Meier D C, Evju J, Semancik S 2005 Artificial neural networks based recognition of chemical hazards in mixed backgrounds using temperature-programmed microsensor signals. *Proc. Int. Symp. Olfaction and Electronic Nose (ISOEN)*, Barcelona, Spain

Bonne U 1990 Flowmeter fluid composition correction. *US Pat.* 4 961 348, assigned to Honeywell, October 9, 1990

Bonne U 1991 Gas composition correction for hot element flow microsensors. *Proceedings 6th IGT Symp. Gas Quality Measurement*, Chicago, IL, USA, June 10–12, 1991

Bonne U 1991, 1995 Microbridge-based combustion control. *US Pat.* 5 401 162, March 28, 1995, assigned to Honeywell; *Int. Pat.* WO 91/06809, May 16, 1991

Bonne U 1992 Fully compensated flow microsensor for electronic gas metering. *Proceedings International Gas Research Conference*, Orlando, FL, USA, November 16–19, 1992, Vol. III, p. 859

Bonne U 1993 Determination of compressibility factor and critical compression ratio with Si-based microstructure sensors. *Proceedings 7th IGT Symp. Natural Gas Quality Measurement*, Chicago, IL, USA, July 12–14, 1993a

Bonne U 1994a On-line combustionless measurement of gaseous fuels fed to gas consumption devices. *US Pat.* 5 311 447, assigned to Honeywell, May 10, 1994

Bonne U 1994b Thermal microsensors for environmental and industrial controls. *NISTWorkshop on Gas Sensors: Strategy for Future Growth*, Gaithersburg, MD, USA, September 8–9, 1993, NIST Special Publication No. 865, 1994, pp. 31–8, ISSN: 1048-776X

Bonne U 1996a Sensing fuel properties with thermal microsensors. *SPIE Smart Electronics and MEMS Conference*, San Diego, CA, USA, February 25–29, 1996, Paper No. 2722-24, Proceedings, p. 165

Bonne U 1996b Determination of fuel characteristics. *US Pat.* 5 486 107, January 23, 1996, assigned to Honeywell

Bonne U 1998 Microsensor fabrication and application. *Proceedings 11th IGT Symp. Gas Quality and Energy Measurement*, St. Petersburg, FL, USA, February 2–4, 1998

Bonne U 2007 Thermal micropump tradeoffs. *DARPA Microsystems Symposium*, San Jose, CA, USA, March 4–8, 2007, Poster Abstract, accepted for presentation

Bonne U, Higashi R E 2004 PHASED: A faster, smarter and more affordable gas analysis device. *J. Process Anal. Chem. 16th Int. Forum Process Analytical Chemistry (IFPAC)*, San Diego, CA, USA, January 22–25, 2002. J. Process Anal. Chem. August, 6

Bonne U, Kubisiak D 1994 Overpressure-proof, thermal pressure sensor for gases. *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, June 13–16, 1994, Tech. Digest, p. 76

Bonne U, Kubisiak D 1996 Microsensors for fluid properties. *Sci. Honeyweller*. **15**, 21

Bonne U, Kubisiak D 2000 Time lag approach for measuring thermal conductivity and specific heat. *US Pat.* 6 019 505, February 1, 2000, assigned to Honeywell

Bonne U, Kubisiak D 2001a Actuation-based microsensors. *J. Smart Mater. Struct*. **10**, 1185

Bonne U and Kubisiak D 2001b Self-oscillating fluid sensor. *US Pat.* 6 223 593, May 1, 2001, assigned to Honeywell

Bonne U, Kubisiak D 2001, 2003, 2004 Self-normalizing flow sensor and method for the same. *US Pat.* 6 308 553, October 30, 2001, 6 553 808, April 29, 2003, and 6 715 339, April 6, 2004

Bonne U, Kubisiak D, Matthys R, Satren E, Aoshima S, Kamiunten S, Butterworth E, Cheung S 1990 Microanemometer-based flow sensing. *Proc. 5th IGT Symp. Natural Gas Quality Measurement*, Chicago, IL, USA, July 16–18, 1990 (HTC)

Bonne U, Vesovic V, Wakeham W A 1994 Thermophysical properties of natural gases for on-line metering. *12th Symp. Thermophysics Properties*, Boulder, CO, USA, June 19–24, 1994. *Int. J. Thermophys*. **16**, 195

Bonne U, Vesovic V, Wakeham W A 1995a Sensing thermophysical and transport properties of natural gas with thermal microsensors. *Interntional. Gas Research Conference*, Cannes, France, November 6–9, 1995, Proceedings, p. 671 (Proceeding Preprints, Vol. III, p. 152)

Bonne U, Cole, B E, Schuldt S B Ule E R 1995b Process for operating a solid-state oxygen microsensor. *US Pat.* 5 389 218, February 14, 1995, assigned to GRI

Bonne U, Honda N, Ohnstein T 1996 Determination of dew point or absolute humidity. *US Pat.* 5 533 393, July 9, 1996; assigned to Honeywell

Bonne U, Kubisiak D, Matthys R, Schuldt R 2000a Method and apparatus for measuring selected properties of a fluid of interest using a single heater element. *US Pat.* 6 079 253, June 27, 2000

Bonne U, Kubisiak D, Matthys R, Schuldt R 2000b Method and apparatus for measuring selected properties of a fluid of interest using a single heater element. *US Pat.* 6 079 253, June 27, 2000, assigned to Honeywell. Doc. No.: SEN-I00-026

Bonne U, Kubisiak D, Creasy K, Francisco T 2001a Rugged microsensor for process stream composition monitoring. *15th Int. Forum on Process Analytical Chemistry (IFPAC)*, Amelia Island, FL, USA, January 21–24, 2001 *J. Process Anal. Chem.* 7: 1 (2001–2002)

Bonne U, Kubisiak D, Matthys R J, Schuldt S B 2001b Time lag approach for measuring fluid velocity. *US Pat.* 6 234 016, May 22, 2001, assigned to Honeywell

Bonne U, Detry J, Higashi R E, Rezacheck T, Swanson S 2002 New gas composition and trace contaminant sensors. *GTI Gas Technology Conference*, Orlando, FL, USA, September 30–October 2, 2002, Proceedings

Bonne U, Kubisiak D, Satren E, Weeres S R 2003a Microsensor for measuring velocity and angular direction of an incoming air stream. *US Pat.* 6 502 459, January 7, 2003, assigned to Honeywell

Bonne U, Higashi R E, Detry J, Rezachek T, Newstrom-Peitso K, Swanson S, Satren E 2003b Industrial wireless PHASED sensor. Phase 1. Feasibility demonstration. Final Report to DOE/ORNL, Contract DE-AC05-00OR22725, Honeywell Labs, Plymouth, MN, April 30, 2003, on webpage http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0603/bonne_03.pdf

Bonne U (Honeywell Labs), Eden G (University of Illinois), Frye-Mason G (Nomadics), Sacks R (University of Michigan), Synovec R (University of Washington) (2003c) Micro gas chromatography tradeoff study. Final Report, to DARPA/AFRC/UTC No.: AFRL-PR-WP-TR-2004-2060; Contract No. 03-S530-0013-01-C1, Plymouth, MN, December 1, 2003, http://www.stormingmedia.us/89/8975/A897524.html

Bonne U, Detry J, Higashi R, Newstrom-Peitso K, Pham H, Rezacheck T, Swanson S 2003d PHASED, a faster, smarter and more affordable gas analysis device – Update. *17th Int. Forum Process Analytical Chemistry (IFPAC)*, Scottsdale, AZ, USA, January 21–24, 2003

Bonne U 2004 PHASED micro gas analyzer. DARPA-MTO webpage, at http://www.darpa.mil/mto/mga/summaries/2004_summaries/honeywell.html

Bonne U, Higashi R E, Johnson K, Iwamoto N, Sacks R D, Synovec R E 2005a Stationary phase films for microanalytical measurements. *PittCon 2005*, Orlando, FL, USA, February 27–March 4, 2005, Paper No. 420-1

Bonne U, Johnson K, Higashi R, Nusseibeh F, Newstrom-Peitso K, Marta T, Cabuz E, Wang T-Y, Mosher J, Iwamoto N, Lytle N (Honeywell International) Synovec R, Reid V, Prazen B, Gross G, Veltkamp D (Chemistry Department/CPAC, University of Washington, Seattle, WA) Herring C, Kunze K, Kellner D (Caviton Inc., Champaign, IL) 2005b PHASED*: *Development of a* μ GC-on-a-chip with interface to NeSSI. *Sensors Expo*, Chicago, IL, USA, June 6–9, 2005, Session on 2005 DOE-ITP Sensors & Automation Annual Portfolio Review http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0605/bonne_0605.pdf

Bonne U, Higashi R, Marta T, Nusseibeh F, Rezachek T (Honeywell Labs) Herring C, Kellner D, Kunze K, Castelein M (Caviton Inc.) 2006a Microgas analyzer for NeSSI and DHS: Measurements and simulations. *PittCon 2006*Orlando, FL, USA, March 13–16, 2006, Paper #2020-6

Bonne U, Nusseibeh F, Higashi R, Newstrom-Peitso K, Marta T, Rezachek T, Duffy B, Stacy S, Hilton L, Iwamoto N, Evans J, (Honeywell International), Synovec R, Reid V, McBride A, (Chemistry Department/CPAC, University of Washington, Seattle, WA), Herring C, Kunze K, Kellner D, (Caviton Inc., Champaign, IL) 2006b *High-speed gas analysis with PHASED MGA Gen.1 on NeSSI. DOE-ITP Session on Sensors to Revolutionize Manufacturing, Sensors Expo*, Rosemount, IL, USA, June 5, 2006. http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0606/hw-phased_bonne_0606.pdf

Bonne U, Nguyen T-C, Polla D L 2007 Developments in ultra micro gas analyzers. Chapter 9.3. in: Koch M V, Van den Bussche K M, and Chrisman R W (eds.) *Micro-Instrumentation for High Throughput Experimentation and Process Intensification*, Wiley-VCH, Weinheim (withdrawn)

CarboCap CO_2 Transmitter, by Vaisala Oyj, PO Box 26, FI-00421 Helsinki, Finland or webpage, http://www.vaisala.com/businessareas/instruments/products/carbondioxide

Carron K T (1997) Molecular specific detector for separation science using SERS. *US Pat.* 5 693 152, December 2, 1997, assigned to the University of Wyoming

Carron K T, Kennedy B J 1995 Molecular-specific chromatographic detector using modified SERS substrates. *Anal. Chem*. **87**, 3353–6

Catalog #15, 1998 Micro Switch Division of Honeywell, September, listing AWM-Series air flow sensors based on the microbridge sensor chip. Additional info available via the Customer Response Center at (800) 537-6945

Cavicchi R E *et al*. (NIST) 2000 Micron-scale differential scanning calorimeter on a chip. *US Pat.* 6 079 873

Clifford P K, Tuma D T 1982–1983 Characteristic of semiconductor gas sensors part I: Steady state gas response. *Sens. Actuators* **3**, 233–54

Cole B, Ule E, Schuldt S, Bonne U 1986 Oxygen microsensor development. Final Report to GRI, Contract No.: 5084-234-0984, GRI-86/0190 and -90/0170, Honeywell Labs, Plymouth, MN, March 31, 1986, revised July 21, 1986

Committee on Assessment of Security Technologies for Transportation 2004 Opportunities to improve airport passenger screening with mass spectrometry. National Research Council, Report, 56 pages, http://www.nap.edu/catalog/10996.html, as part of Assessment of Technologies Deployed to Improve Aviation Security: 2nd Report, Progress Towards Objectives, NRC. Washington, DC, 2002

Dable B K, Cavicchi R E, Semancik S 2004 Calibration of microhotplate conductometric gas sensors by non-linear multivariate regressive methods. *Sens. Actuators B Chem*. **101**, 284–94

Daws C A, Exstrom C L, Sowa J R, Mann K R 1997 Vapochromic compounds as environmental sensors. 2. Synthesis and nearinfrared and infrared spectroscopy studies of $[Pt(arylisocyanide)_4][Pt(CN)_4]$ upon exposure to volatile organic compound vapors. *Chem. Mater*. **9**, 363–8

Dawson D H, Henshaw G S, Williams D E 1995 Description and characterisation of a hydrogen sulfide gas sensor based on $Cr_{2-y}Ti_yO_{3+x}$. *Sens. Actuators B Chem*. **26**, 76–80

Dieringer J A, McFarland A D, Shah N C, Stuart D A, Whitney A V, Yonzon C R, Young M A, Zhang X, Van Discovered by and named after Léon Poiseuille (1799)–1869 and Gotthilf Hagen (1797–1884)

Duyne R P 2005 *Surface enhanced Raman spectroscopy: New materials, concepts, characterization tools, and applications*. R. Soc. Chem. J. Faraday Discuss. Paper 132, www.rsc.org/faraday_d

Djorup R S (1993) Thermal anemometer transducer wind set. *US Pat.* 5 218 865, June 15, 1993 (Cossonay)

Doe S 2002–2006 Parker Hannifin, private communications

Doleman B J, Severin E J, Lewis S N 1998 Trends in odor intensity for human and electronic noses: Relative roles of odorant vapor pressure vs. molecularly specific odorant binding. *Proc. Natl. Acad. Sci. U S A*. **95**, 5442–7

Durst F, Glass J, Pickenacker K, Trimis D, Wawrzinek K 1999 Air ratio controlled combustion through measurement of the Wobbe number. *GasWaerme Int*. **48(7/8)**, 429

Eiceman G A, Karpas Z 1994 *Ion Mobility Spectrometry*. CRC, Boca Raton, FL

Eiceman E G A, Krylov E V, Krylova N S (Department of Chemistry and Biochemistry, NM State U), Nazarov E G, Miller R A (Sionex Inc.) 2004 Separation of ions from explosives in differential mobility spectrometry by vapor-modified drift gas. *Anal. Chem.* **76**, 4937

Ferran R J, Boumsellek S 1996 High pressure effects in miniature arrays of quadrupole analyzers from 10^{-9} to 10^{-2} torr. *J. Vac. Sci. Technol. A*. **14**, 1258

Frampton R, Walleshauser J, Bonne U, Kubisiak D, Hoy D, Andu I, Kelly K 1996 Gas mass flow sensor proof of concept testing for space shuttle orbiter flow measurement. *26th Int. Conf. Environment Systems*, Monterey, CA, USA, July 8–11, 1996, SAE Tech. Paper 961335

Garimella S, Fisher T, Schlitz D J (Purdue University and Thorrn Micro Technologies Inc.) 2004 Microscale ion-driven air flow cooling technology. MICRO NANO (Monthly newsletter from the Editors of R&D Magazine). **9/4**, 5, and webpages http://www.purdue.edu/dp/Nano/brochure2mb.pdf, p. 29, and http://www.chipscalereview.com/archives/0504/article.php?type=feature&article=f3

Gokhfeld Y (2006) Pump for supplying a gas to a sensor and methods therefore *US Pat.* 7 056 098, June 6, 2006, assigned to GE Co.

Grate J W, Kaganove S N, Bhethanabotla V R 1998 comparisons of polymer/gas partition coefficients calculated from responses of thickness shear mode and surface acoustic wave vapor sensors. *Anal. Chem*. **70**, 199–203 (PNNL)

Gupta M 2004 Cavity-enhanced gas analyzer for process control applications (Los Gatos Research). SBIR Phase II, DE-FG02-03ER83849 . *DOE Sensors and Automation, Review Meeting*, Arlington, VA, 2004 http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0604/presentations/09_los_gatos_gupta_cps_17139.pdf

Haynes C L, McFarland A D, Van Duyne R P 2005 Surface-enhanced Raman spectroscopy. Anal. Chem. **77**, 339A

Heaps D A and Griffiths P R (2005) Off-line direct deposition gas chromatography/surface-enhanced Raman scattering and the ramifications for on-line measurements. *Appl. Spectrosc.* **59**, 1305–9

Herring C M 2006a Caviton Inc., Private communication, June 2006

Herring C M 2006b Micro-discharge gas detector. *US Pat.* 7 100 421 B1, September 5, 2006, assigned to Caviton Inc.

Higashi R and Johnson R G 1987 Semiconductor device. *US Pat.* 4 651 564, assigned to Honeywell, March 24, 1987 (Honeywell Technology Center)

Higashi R, Johnson R G, Mathur A K, Pearman A N, Bonne U 1986 Microstructure sensors for flow, differential pressure and energy measurement. *IGT Symp. Natural Gas Energy Measurement*, Chicago, IL, USA, April 30–May 2, l986, Proceedings

Hoffheins B S, Lauf R J, Siegel M W 1986 Intelligent thick film gas sensor. *Proceedings of the Atlanta Meeting*, Atlanta, GA, USA, International Society for Hybrid Microelectronics, October 1986

Hogan P, Chapples J, Pratt K 2006 Honeywell Analytics, Private Communications, August–September

Honami S, Nagata M, Kamiunten S, Honda N, Itoh T 1999 Innovative surface micro-machined membrane flow sensor. *Fluids Engineering Division Summer Meeting*, San Francisco, CA, USA, July 18–23, 1999, Proceedings of ASME/JSME, paper FEDSM'99-7356

Honeywell Analytics, 405 Barclay Blvd. Lincolnshire, IL 60069; and City Technology Ltd, Walton Road, Portsmouth, Hampshire, PO6 1SZ, UK, or web at http://www.citytech.com/technology/pellistors.asp

Honeywell Analytics 2006 *Gas Book*, May 2006, and webpages at http://www.honeywellanalytics.com, *Product Range, Analyzers and Sensors*; Honeywell-CityTechnology at http://www.citytech.com/technology/pellistors.asp and http://www.citytech.com/technology/irsensors.asp

Huising J H, Riedijk F R, Horn G v d 1993 Developments in integrated smart sensors. *7th Int. Conf. S.S. Sensors and Actuators, Proceedings*, Hilton Head Is., Sc, USA, p. 320

Hulteen J C, Young M A, Van Duyne R P 2006 Surface-enhanced hyper-Raman scattering (SEHRS) on Ag film over nanosphere (FON) electrodes: Surface symmetry of centro-Ym-metric adsorbates. *Langmuir*, **22(25)**, 10354–64

Huseth S, Gonia P, Bonne U 2004 Wireless and sensing solutions advancing industrial efficiency, Part II: PHASED microanalyzer. DOE S&A Annual Review, June 21, 2004, and website at http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0604/presentations/04_honeywell_cps_14227.pdf

International Components Co. Audible signal devices. Stock No. 623-2012, Type BRT1209P-01; $V_{in} = 1$–2 VDC, $I_{max} = 10$ mA, pin spacing: 6.5 mm, price (for quantity of 25–99) = $0.69; and Type BRP1407P-20. Nominal eigenfrequency near 2.3 kHz

James S D, Johnson R G, Higashi R E 1988 A broad-range, absolute pressure microsensor. *IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head, SC, USA, June 6–9, 1988, Book of Abstracts, p. 107

Jerman J H, Clift D J, Mallinson S R 1990 A miniature FP interferometer with a corrugated silicon diaphragm support. *1990 Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, Proceedings, p. 140, (British Telecom)

Kenndler E 1999 Gas chromatography. University of Vienna, Austria, September 9, 1999, http://www.anc.univie.ac.at/scripts/ and the January 19, 2004 version of http://www.anc.univie.ac.at/scripts/Gas_Chromatography_in_Capillaries.pdf

Kleine-Benne E, Röck F, Gurlo A, Weimar U, Blaschke M 2005 Detection of packaging emissions using a flexible headspace sampler combined with a multi sensor system and a separation unit. Note 7/2005, Gerstel GmbH, webpage at http://www.gerstel.com/p-gc-an-2005-07.pdf

Kornienko A, Reilly P T A, Whitten W B, Ramsey J M 1999 Micro ion trap mass spectrometry. *Rapid Commun. Mass Spectrom*. **13**, 50

Kramer H, Mangieri K 2006 Gas detection. Review of sensor technologies used in portable gas monitors. ISC Article 19/5/06 12:22 pm Page 1, and at webpage http://www.envirotechpubs.com/pdf/iet/2006/05/iet200605_014.pdf

Kroot P 1996 A new oxygen sensor for cleaner, more efficient combustion. Scientific Honeywellei, 29

Kunt T, Cavichi R E, Semancik R E and McAvoy T J 2000 Method of operating a sensor to differentiate between analytes in a sample. *US Pat.* 6 095 681, August 1, 2000

Kunze K 1997 Miniaturized discharges – Prospects and limits for quantitative analysis. Ph.D. thesis, University of Dortmund, Germany

Laser D J, Santiago J G (Stanford University) 2004 A review of micropumps. *J. Micromech. Microeng*. **14**, R35–64

Lehmann U *et al*. (2005) Unique solutions for gas analysis and sensors., SLS webpage, at http://www.sls-microtechnology.de/ (SLS Micro Technology GmbH)

Lu C-J, Steinecker W H, Tian W-C, Oborny M C, Nichols J M, Agah M, Potkay J A, Chan H K L, Driscoll J, Sacks R D, Wise K D, Pangad S W, Zellers E T 2005 First-generation hybrid MEMS gas chromatograph. *Lab Chip*. **5**, 1123–31 (University of Michigan-WIMS)

Madhan R C M, Yadav M, Roth J R (University of Tennessee), Wilkinson S P (NASA Langley) 2003 Aerodynamic flow control by peristaltic acceleration of a one atmuniform glow discharge plasma. *30th IEEE Intl. Conf. Plasma Science (ICOPS)*, Jeju Island, Korea, June 2–5, 2003, Paper 5A07, or web at http://plasma.ee.utk.edu/~plasma/publi/ICOPS_2003_AERO.pdf

Manginell R P, Okandan M, Kottenstette R J, Lewis P R, Adkins D R, Bauer J M, Manley R G, Sokolowski S, Shul R J (2003) Monolithically-integrated microchemlab for gas-phase chemical analysis. *MicroTAS*, Lake Tahoe, Olympic Valley, CA, USA October 5–9, 2003, Proceedings (Sandia National Labs), or Robinson A *et al*. 2003 Microsensors for Industrial monitoring (μChemLab). *DOE Sensors and Automation, Annual Review Meeting*, San Francisco, CA, USA, June 4, 2003, website at http://www.eere.energy.gov/industry/sensors_automation/pdfs/meetings/0603/Robinson_03.pdf (Sandia National Labs)

Masel R I 2004 MEMS gas chromatograph with nanotube, nanogate and microM8 detectors. DARPA-MTO webpage, at http://www.darpa.mil/mto/mga/summaries/2004_summaries/illinois.html

Massok P, Loesch M, Bertrand D 1995 Comparison between two Figaro sensors (TGS 813 and TGS 842) for the detection of methane, in terms of selectivity and long-term stability. *Sens. Actuators B Chem.* **25(1)**, 525–8, (April 1995) or webpage of Figaro Co. sensors, at webpage. http://www.figarosensor.com/products/2442pdf.pdf

Mastrangelo C H, Muller R S191 μ-Pirani pressure gauge with digital readout. *IEEE Transducers '91, Int. Conf. Solid-State Sensors and Actuators*, San Francisco, CA, USA, June 24–27, 1991, Proceedings, p. 245

McGill J A, Abraham M H, Grate J W 1994 Choosing polymer. Coatings for chemical sensors. *ChemTech*. **24**, 27–37

Meier D C, Cavicchi R E, Boger Z, Semancik S 2004 Interference effects on chemical warfare agent detection by MEMS-based microsensor arrays. *Chem. Sens*. **20-B**, 308

MEMSIC Inc. acceleration sensors at webpage http://www.memsic.com/memsic/products/technology.html

Morris C J, Forster F K 2003 (University of Washington) 2003 Low-order modeling of resonance for fixed-valve micropumps based on first principles. *J. Microelectromech. Syst*. **12(3)**, 325–34

Mosher J, Nickels R, Bonne U 2003 Networked sampling system (NeSSI-Generation-II) development and field test. *17th Int. Forum Process Analytical Chemistry (IFPAC)*, Scottsdale, AZ, USA, January 21–24, 2003

NeSSI (New Sampling and Sensor Initiative) News and Updates 2006 CPAC webpage at . http://www.cpac.washington.edu/NeSSI/NeSSI.htm

Niemeyer D, Williams D E, Smith P, Pratt K F E, Slater B, Catlow C R A, Stoneham A M I 2002 Experimental and computational study of the gas-sensor behaviour and surface chemistry of the solid-solution $Cr_{2-x}Ti_xO_3$ ($x \leq 0.5$). *J. Mater. Chem*. **12**, 667–75

NIST WebBook webpage at http://webbook.nist.gov/chemistry/Standard Reference Database Number 69, Released June 2005

Padmanabhan A, Bonne U, Haji Sheikh MJ (2006) Robust fluid flow and property microsensor made of optimal material. *US Pat.* 7 109 842, 19 September 19, 2006, assigned to Honeywell

Pai C S, Pau S, Taylor J A, Mass spectrometers on wafer-substrates. *US Pat.* 6 967 326, 22 Nov. 2005

Patel S V, Mlsna T E, Fruhberger B, Klaassen E, Cemalovic S, Baselt D R 2002 Chemicapacitive microsensors for volatile organic compound detection. *Sens. Actuators B Chem*. **96**, 541–53

Pau S, Low Y L, Pai C S, Moxom J, Reily P T A, Whitten W B, Ramsey J M 2006 Microfabricated quadrupole ion trap for mass spectrometer applications. *Phys. Rev. Lett*. **96**, 120801

Polla D, Nguyen C 2004 Vision statement. DARPA-MTO webpage, at http://www.darpa.mil/mto/mga/vision/index.html

Poppe H 2002 Mass transfer in rectangular chromatographic channels. *J. Chromatogr. A*. **948**, 3–17 (University of Amsterdam)

Pratt K *et al*. 2003 The effect of flow rate on cap21 ozone sensitivity. Capteur Division of City Technology Ltd, Walton Road, Portsmouth, Hampshire, PO6 1SZ, UK, June 2003, website at http://www.citytech.com/PDF-Datasheets/MMOS_cap21_t2ops.pdf

Priti P, Juneja S, Chapman G H 1993 Wide-range vacuum sensor using phase-sensitive detection. *Proc. Canadian Conf. VLSI*, Banff, AB, Canada, November 4–17, 1993, Proceedings, p. 3A7

Raizer Y P 1991 In: Allen J E (ed.) *Gas Discharge Physics*., Springer-Verlag, Berlin, Heidelberg, Chap.9, p. 214

Raley N F, Ciarlo D R, Koo J C, Berriger B, Trujillo J, Yu C, Loomis G, Chow R 1992 A FP microinterferometer for visible wavelengths. *IEEE*, p. 170–3

Rich R R 2001 Removing the guesswork from furnace atmosphere control with (Raman) laser gas analysis. *21st ASM Heat Treating Society Conference*, 5–8 Nov., Indianapolis, In: reprinted in *Heat Treating Progress*, Sept. (2002) and webpage: http://www.atsmrcv.com./ and http://www.atmrcv.com/frameset.php?docsrc=templib/articles/lga Article.pdf

Roth J R 2003 Aerodynamic flow acceleration using paraelectric and peristaltic electrohydrodynamic effects of one atmosphere uniform glow discharge plasma. *Phys. Plasmas*. **10(10)**, 2117–26

Sacks R D 2004 Private communication (University of Michigan)

Sanchez J M, Sacks R D 2003 GC analysis of human breath with a series-coupled column ensemble and a multibed sorption trap. *Anal. Chem*. **75**, 2231–6 (University of Michigan)

Saridara C, Mitra S 2005 Chromatography on self-assembled carbon nanotubes. *Anal. Chem*. **77**, 7094–7 (NJIT, Newark, NJ)

Schierbaum K D, Vaihinger S, Goepel W, van den Vlekkert H H, Kloeck B, Rooij N F 1991 Prototype structure for systematic investigations of thin-fim gas sensors. *Sens. Actuators B Chem*. **3**, 171–5

Siegel M W 1989 Olfaction metal oxide semiconductor gas sensors and neural networks, July 1989 and webpage at www.cs.cmu.edu/afs/cs/project/sensor-9/ftp/papers/nato.ps (Carnegie Mellon University, Robotics Institute)

Simonson J (2004) Micro-scale gas separation analyzer. DARPA-MTO webpage, at http://www.darpa.mil/mto/mga/summaries/2004_summaries/snl.html

Sommers H, Korameier W, Vissel F 1979 Patent DE 29 28 739 B1, July 17, 1979

Stadermann M, McBrady A D, Dick B, Reid V, Noy A, Synovec R E, Bakajin O 2006 Fast gas chromatography on single-wall carbon nanotube stationary phases in microfabricated channels. *Anal. Chem*. **78**, 5639 (University of Washington and LLNL)

Staples E J 2004 Chemical profiling cargo with an ultra-high speed gas chromatograph, olfactory images, and virtual chemical sensors. *3rd Microsensors (DHS-ATF)Workshop*, Scottsdale, AZ, USA, April 19–22, 2004 on website, at http://www.estcal.com/TechPapers/CargoContainerOdors.doc and http://www.znose.com/modelsection.htm (Electronic Sensor Technology, Newbury Park, CA)

Tenax preconcentration tables from Scientific Instrument Services, (1027) Old York Rd., Ringoes, NJ 08551 phone: (908)788- webpage at www.sisweb.com and at http://www.sisweb.com/index/referenc/tenaxta.htm

Tiffany J, Cavicchi R E, Semancik S 2000 Microarray study of temperature-dependent sensitivity and selectivity or metal-oxide sensing interfaces. *Proc. Advances in Environmental and Chemical Sensing Techniques*, Boston, MA, USA, Proceedings SPIE

Visser J H, Zanini M, Rimai L, Soltis R E, Kovalchuk A, Hoffman D W, Logothetis E M, Bonne U, Brewer L T, Bynum O W, Richard M A 1994 Catalytic calorimetric gas sensors. *5th Int. Meeting on Chemical Sensors*, Rome, Italy, July 11–14, 1994, Proceedings, p. 468

Wakeham W A, Nagashima A, Sengers J V 1991 *Measurement of the Transport Properties of Fluids*. Blackwell Scientific Publishers, Oxford, UK

Whitten W B 2004 Micro ion trap mass spectrometer. DARPA-MTO, webpage at http://www.darpa.mil/mto/mga/summaries/2004_summaries/oakridge.html

Wilson M, Kubisiak D, Wood R A, Listvan M, Ridley J A 1991 An uncooled thermoelectric microthermopile camera developed using silicon microstructure sensors. *Proc. 1991 IRIS Specialty Group on Infrared Detectors*, Boulder, CO, USA

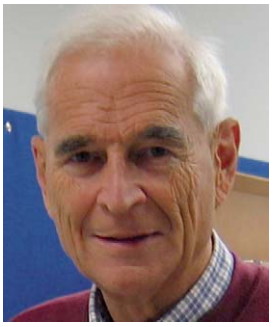
Witten W B, Ramsey M 2002–2006, Private communications

Zanini M, Visser J H, Rimai L, Soltis R E, Kovalchuk A, Hoffman D W, Logothetis E M, Bonne U, Brewer L, Bynum O W, Richard M A 1995 Fabrication and properties of a Si-based high-sensitivity microcalorimetric gas sensor. *IEEE Solid-State Sensors Workshop*, Hilton Head Island, SC, USA, June 13–16, 1994, Tech. Digest, p. 176, *Sens Actuators A Phys* **48**, 187

Zhao Y, Cai Y 2005 *Z*-axis thermal accelerometer. *USPTO Serial No.* 11125759, Filed May 10, 2005

Zook D, Bonne U, Samad T 2000 Sensors in control systems. In: Samad T, Weyrauch J (eds.) *Automation, Control and Complexity. An Integrated Approach*. John Wiley, New York, Chap. 8

Biography



Ulrich Bonne received an MS in physics in 1960, and a Ph.D. (Dr.rer.nat) in Chemical Physics in 1964 from the University of Göttingen, Germany, with a thesis titled "Optical Studies in the Reaction Zone of (Low-Pressure, Flat) Sooting Flames." Since joining Honeywell Laboratories in Minnesota, he focused on R&D and productization of sensors (of flames, O_2, and efficiency), modeling seasonal system efficiency, and then on devising an annual fuel utilization efficiency (AFUE) test method with NIST, which ASHRAE and DOE later officially adopted. Since the mid-1980s he has pioneered the development of the core micro-bridge sensor technology for environmental, medical, and aerospace applications, and most notably for an electronic, fully compensated natural gas metering system. Since 2002, he has secured funding from DOE and DARPA to model, design, and demonstrate the performance of innovative micro gas analyzers, detectors, and thermal micropumps. His work has led to over 80 US patents and 140 publications. He is a member of the International Combustion Institute; before his retirement in 2007, he served on ASHRAE technical committees DOE and NIST evaluation panels, and was a member of ASHRAE, Opt.Soc.Am., SAE, and ACS.

2.12 Chemical Sensors

Robert K. Franklin,[1] Steven M. Martin,[2] Timothy D. Strong,[3] and Richard B. Brown[4],
[1]The University of Michigan, Ann Arbor, MI, USA, [2]Intel Corporation, Colorado Springs, CO, USA, [3]Sensicore, Inc., Ann Arbor, MI, USA, [4]The University of Utah, Salt Lake City, UT, USA

Glossary

Anodic stripping voltammetry An electrochemical measurement technique that involves a plating step where the electrode is held at a negative potential for a period of time in order to concentrate the analyte by plating it onto the electrode followed by a stripping step where the electrode is switched to a positive potential and the analyte is stripped off the electrode generating a faradaic current, which is measured to determine the concentration of the analyte in the bulk solution.

Chronoamperometry An electrochemical measurement technique that involves applying a potential step to an electrode in solution and then measuring the current generated at the electrode as a function of time.

Cyclic voltammetry An electrochemical measurement technique that involves applying a periodic triangular wave potential to an electrode and then measuring the current generated at the electrode as a function of the applied potential.

Electrical double layer A model used to describe the interaction of molecules at an electrode surface. Currently, the most accurate model of this region is the Gouy–Chapman–Stern model, which includes three distinct regions: the inner Helmholtz plane, the outer Helmholtz plane, and the diffuse layer.

Ion-selective electrode An electrode that has been covered by an ion-selective membrane. The membrane causes the potentiometric response of the electrode to be selective to the concentration of a single ion in the solution.

Linear sweep voltammetry An electrochemical measurement technique that involves applying a potential ramp to an electrode and then measuring the current generated at the electrode as a function of the applied potential.

Polarizable The relative ease with which electrons transfer across the electrode–solution interface defines whether the electrode is considered polarizable or nonpolarizable.

Potentiometry A class of electrochemical techniques that measure the equilibrium potential between two electrodes in order to determine the concentration of a molecule in a solution.

Potentiostat A device that can maintain a fixed potential between a pair of nodes in a circuit by adjusting the current flowing through a second set of nodes in the same circuit. During a voltammetric experiment, a potentiostat can be used to fix the potential between the working electrode and the reference electrode by adjusting the current flowing between the working electrode and the counter electrode.

Reference electrode An electrode that develops a stable and repeatable potential when placed in a solution. Reference electrodes are used to set the relative potential of a solution in an electrochemical cell and are analogous to the ground node of an electric circuit.

Voltammetry A class of electrochemical techniques that involve applying an electric potential between two electrodes and measuring the resultant current in order to determine the concentration of a molecule in a solution.

2.12.1 Introduction

The 21st century will surely be a century of information technology. During the last quarter of the 20th century, computing technology became ubiquitous, moving from mainframe computers in large companies and institutions into the home, classroom, automobile, doctor's office, and telephone. The microprocessor dramatically increased the world's data processing capacity. The limiting factor now in expanding computer analysis and control for clinical, biotechnological, agricultural, industrial, and environmental applications is the interface between electronics and the nonelectronic world. Chemical sensors bridge the gap between chemistry and electronics by transducing chemical concentrations to electronic signals using a variety of transduction methods such as electrochemistry, surface acoustic waves (SAW), and optics. The goal of this chapter is to provide an overview of the theory behind liquid-phase electrochemical sensors, to describe two major classes of these chemical sensors – potentiometric and voltammetric sensors – and to discuss the instrumentation and packaging needed to build a complete chemical sensing microsystem.

2.12.1.1 Motivation for Building Microchemical Sensors

Applications in a number of areas demand tightly integrated microsystems that incorporate a chemical sensing subsystem. Two examples of the significant applications of this technology are implantable

sensors for monitoring biological chemicals such as glucose in the bloodstream (Beach *et al.* 1999, Guiseppi-Elie *et al.* 2005, Koudelka *et al.* 1991) and sensors for real-time monitoring of industrial effluents (Thomas *et al.* 1999). While analytical instruments and laboratory procedures are available for collecting the required data in each of these cases, the demand for faster test results and lower costs is driving research in the area of low-cost, portable instruments (Wolffenbuttel 1996). As these types of instruments become available, they will have an impact on chemical data collection, which can be comparable to the effect that microprocessors have had on data processing.

Small sensors mean smaller sample sizes and an increase in data collection density. Many microscale chemical sensors are also cheaper to mass-produce than their macroscale counterparts. This cost reduction comes in part because the sensors can be batch fabricated with several thousand sensors being fabricated simultaneously on a single substrate. Additionally, many chemical sensors utilize precious metals such as platinum, gold, and silver, and the reduction in raw materials required to manufacture the sensors is also a significant factor.

Calibration and maintenance are also important factors to consider. In virtually every precision-sensing application, periodic calibration is required to ensure accuracy. Chemical sensors are normally calibrated with a calibration solution formulated for the specific test the sensor is designed to perform. Microsystems and lab-on-a-chip solutions with integrated microfluidic handling can store the calibration solution internally and then perform the periodic calibration automatically, making the calibration function invisible to the end user, thereby reducing the complexity of using the device and reducing the cost per test.

Microfabricated chemical sensors can also be designed as small arrays of multiple sensors, where each sensor can be selective to a different chemical. These arrays can be used to measure the concentration of multiple chemicals simultaneously, and by using several sensors having different selectivities and sensitivities the concentration of a single chemical can be determined more accurately in the presence of interferants (Brown 1985). The same process, of course, could be implemented with conventional macroscale chemical sensors, but a large sample would be required and the sensors could not be placed in close proximity to each other for measurements in nonhomogeneous solutions.

There are disadvantages of small size as well. The manufacturing and handling of very small sensors is more challenging than that of macrosized sensors. Additionally, microscale sensors are more susceptible to noise and to variations in manufacturing than are similar macroscale sensors. Although techniques have been developed to mitigate both of these issues, the fixes make the designing and testing of miniature chemical sensors more complex than that of conventional chemical sensors. These disadvantages make microfabricated chemical sensors wholly unsuitable for some applications.

Despite these disadvantages, many applications exist such as the implantable glucose sensors (Koudelka *et al.* 1991) and the effluent monitoring systems (Thomas *et al.* 1999) mentioned earlier that either require miniaturization to be viable or else would be too expensive to be widely deployed without the cost savings derived from batch microfabrication.

Perhaps though, the single most compelling reason for miniaturizing chemical sensors is the promise of complete microsystems that integrate chemical sensors with electronics, microfluidics, and sensors for other physical parameters such as pressure and temperature. In the future, these lab-on-a-chip solutions will provide solutions to problems such as long-term, low-cost environmental monitoring and faster, more reliable monitoring of medical patients.

2.12.2 Electrochemical Theory

A basic knowledge of electrochemical theory is needed to understand or to design liquid chemical sensors. While an in-depth discussion of this topic is beyond the scope of this chapter, the basic concepts of electrochemistry are presented here in sufficient detail to allow the reader to grasp the principles of operation behind these sensors. For those who wish to delve deeper into the electrochemical theory, there are several excellent texts devoted to the subject, including Bard and Faulkner (2000) and Kissinger *et al.* (1996).

2.12.2.1 The Electrode–Electrolyte Interface

All electrochemical measurements rely on the behavior of an analyte in a heterogeneous environment. When an electrode is placed in a solution, a phase boundary exists between the solid electrode and the bulk solution. Additionally, a barrier to charge transfer exists between the two phases due to a difference in charge carriers (charge is carried by electrons in

solid conductors and by ions in solutions). Charge can cross this barrier only when a molecule in the solution is either reduced thereby receiving an electron from the electrode or oxidized by giving up an electron to the electrode.

A generic reduction–oxidation (redox) reaction follows the form as given below:

$$O + ne^- \rightleftharpoons R \qquad [1]$$

where O is the oxidized form of the analyte, R is the reduced form, and n is the number of electrons (e^-) exchanged in the reaction. This reaction has two components, a forward reaction and a reverse reaction, and the reaction can be either reversible or irreversible. If the reaction is reversible, then the reaction can proceed in both directions, and neither O nor R is ever present without the other species being present, although the amount of one of the two reactants may be very small. Reversible reactions eventually reach an equilibrium point where the rate of the forward reaction equals the rate of the reverse reaction. Irreversible reactions proceed in only one direction and given enough time, one of the two reactants (either O or R) will be completely consumed by the reaction. Eqn [1] represents the simplest form of a redox reaction. In practice, more complex reactions involving three or more products with intermediate reactions are often encountered. However, for the sake of simplicity, a discussion of these more complex reactions is omitted in this chapter.

2.12.2.2 The Nernst Equation

A result of the barrier to charge transfer is that almost all electrodes, when placed in solution, develop a surface charge. At equilibrium, the potential difference, E_M, that develops between the electrode and the solution depends upon the electrode material and the concentration of the analyte in the solution. This potential is known as the Nernst potential and is given by the Nernst equation

$$E_M = E^0 - \frac{RT}{nF}\ln\left(\frac{C_O}{C_R}\right) \qquad [2]$$

where E^0 is the standard potential of the reaction, R is the molar gas constant ($8.3145\,J\,K^{-1}\,mol^{-1}$), T is the temperature, F is the Faraday's constant ($9.6485 \times 10^4\,C\,mol^{-1}$), C_O is the concentration of the oxidized form of the species, and C_R is the concentration of the reduced form of the species (Bard and Faulkner 2000). In this form, the Nernst relationship applies only to dilute solutions; in more concentrated solutions, the concentrations, C_R and C_O, must be replaced by their respective chemical activities. For solid-phase reactants (such as a metal electrode placed in solution), $C = 1$. At $T = 298\,K$ the Nernst equation can be rewritten in a simpler form using the common log instead of the natural log as follows:

$$E_M = E^0 - \frac{59.2mV}{n}\log\left(\frac{C_O}{C_R}\right) \qquad [3]$$

The standard potential of the electrode, E_M, needs special consideration. Since direct measurement of the potential of a solution is impossible, all electrode potentials must be given relative to the potential of a second electrode. The electrode normally used for measuring standard values of E^0 is the normal hydrogen electrode (NHE), which has E^0 defined to be 0 V. E^0 values for a few common electrode materials are listed in **Table 1**. For a more comprehensive table of values, the reader is directed to the many extensive texts on the subject such as Bard *et al.* (1985) and Sawyer *et al.* (1995). If an electrode other than the NHE is used, the E^0 value used in calculating E_M must be adjusted to include the additional potential due to the standard potential of the measurement electrode.

The Nernst equation relates the potential between the electrode and the surrounding solution to a simple logarithmic ratio of two species in solution. For metal electrodes, one species is the metal at the surface of the electrode and the other is the oxidized form of the metal ion in solution. For example, when a copper electrode is placed in a solution containing 0.02 M Cu^{2+} the potential difference that exists between the metal and the solution phases can be calculated using the Nernst equation. The following equation describes the redox reaction between the two reactants:

$$Cu^{2+} + 2e^- \rightleftharpoons Cu(s)$$

Table 1 Standard potentials for several common electrode materials

Electrode half reaction	*E^0 for reaction (V)*
$Au^+ + e^- \rightleftharpoons Au(s)$	1.691
$Hg^{2+} + 2e^- \rightleftharpoons Hg(l)$	0.852
$Ag^+ + e^- \rightleftharpoons Ag(s)$	0.799
$Hg_2Cl_2 + 2e^- \rightleftharpoons 2Hg + 2Cl$	0.268
$AgCl(s) \rightleftharpoons Ag(s) + Cl^-$	0.222
$2H^+ + 2e^- \rightleftharpoons H_2(g)$	0.000

E^0 for the reaction is 0.339 V. Using eqn [3] with $n = 2$, $C_O = 0.02$, and $C_R = 1$ the value of E_M can be calculated to be 0.239 V.

The Nernst equation is the basis for the potentiometric and voltammetric sensors described in this chapter. Potentiometric sensors measure the value of E_M and use the measured value of E_M to determine the ratio of C_O to C_R. Voltammetric sensors control the value of E_M, drive the value of the ratio of C_O to C_R to either infinity or zero, and then measure the current resulting from the associated redox reaction.

2.12.2.3 The Capacitive Double Layer

Under equilibrium conditions, no net current flows across the solid–liquid interface and a charge builds up on both sides of the interface in much the same way as charge gathers on the two plates of a parallel plate capacitor. On the electrode side of the interface, all of the charge gathers at the surface of the electrode. However, in the solution, the charge carriers are ions and not electrons, and ions cannot pack as densely as the electrons in a conductor. In order to counterbalance the charge on the electrode and preserve charge neutrality, ions pack as densely as possible forming a dense layer of ions near the surface of the electrode known as the capacitive double layer (see **Figure 1**). The ions forming the capacitive double layer are normally grouped into three distinct regions. The first region, known as the inner Helmholtz plane (IHP), is composed primarily of polar water molecules and specifically adsorbed anions. The second layer is composed mostly of fully hydrated cations and is called the outer Helmholtz plane (OHP). The final region, the diffuse layer, extends from the OHP to the bulk of the solution and is composed of hydrated anions and cations. The capacitive double layer is very thin (normally <1 μm and often <0.1 μm) and functions similarly to a capacitor with a capacitance, C_{dl}, of approximately 10–100 μF cm^{-2} of electrode surface. However, the behavior of the double-layer capacitance varies from that of a true capacitor in that the capacitance of the layer can vary with the external potential applied to the system.

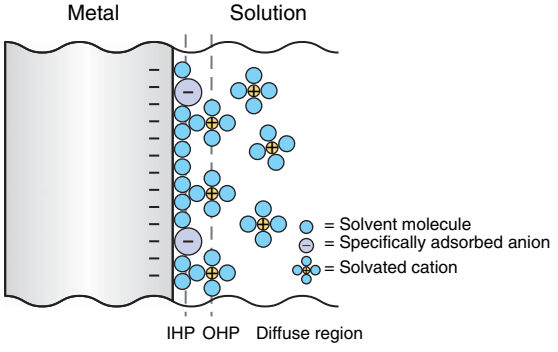


Figure 1 Depiction of the electrical double layer that builds up at the electrode–solution interface when a metal is placed in solution. In the solution there are three distinct layers. The outer Helmholtz plane (OHP), the inner Helmholtz plane (IHP), and the diffuse layer. (Source: Bard A, Faulkner L 2000 *Electrochemical Methods*. John Wiley & Sons, New York, NY.)

2.12.2.4 Polarizable and Nonpolarizable Electrodes

The relative ease with which electrons transfer across the electrode–solution interface defines whether the electrode is considered polarizable or nonpolarizable. Electrodes that utilize very little energy for electron transfer are referred to as nonpolarizable, while those for which the transfer is nearly impossible are referred to as polarizable. No electrode is purely polarizable or nonpolarizable. Real electrodes fall somewhere in between the two extremes; however, close approximations of these ideal electrodes do exist.

Polarizable electrodes pass very little current when a voltage is applied across them. Most noble metals such as platinum and gold are examples of nearly ideal polarizable electrodes. An ideal polarizable electrode behaves like an electrical capacitor with one plate acting as the electrode and the other consisting of the ions and solvent molecules collected near the electrode surface, which form the capacitive double layer.

Nonpolarizable electrodes are characterized by a very fast electrode reaction that allows the electrode's potential to remain close to the equilibrium potential even when current is flowing through the electrode. The most common nearly ideal nonpolarizable electrode is the Ag/AgCl electrode, which is formed by coating a silver electrode with silver chloride. An ideal nonpolarizable electrode behaves like an electrical resistor, with the resistance modeling the difficulty in the transfer of electrons across the interface.

2.12.2.5 The Electrochemical Cell

The apparatus used to carry out an electrochemical experiment is known as an electrochemical cell. Most electrochemical cells consist of either two or three electrodes (**Figures 2** and **3**). Two electrodes are

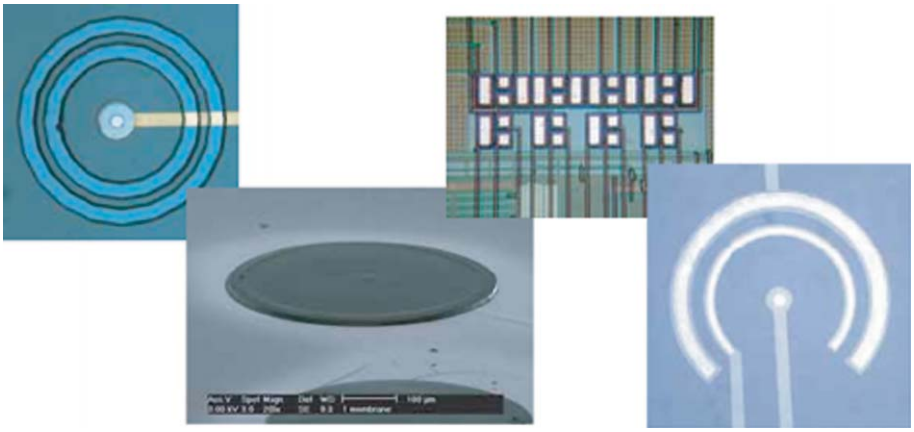


Figure 2 Several example electrodes. *From left to right*: A platinum voltammetric electrode with SU-8 well for enzyme-selective coatings; a Scanning electron micrograph (SEM) of a polymer-based ion-selective electrode (ISE) membrane using an SU-8 ring for containment during membrane dispense (Nam *et al*. 2003); an active voltammetric array with Pt working and counter electrodes and Ag/AgCl reference electrodes fabricated on top of complementary metal oxide semiconductor (CMOS) circuitry (Martin 2005); and a voltammetric sensor with Pt working (inner) and counter (outer) electrodes and integrated Ag/AgCl reference electrode (middle).

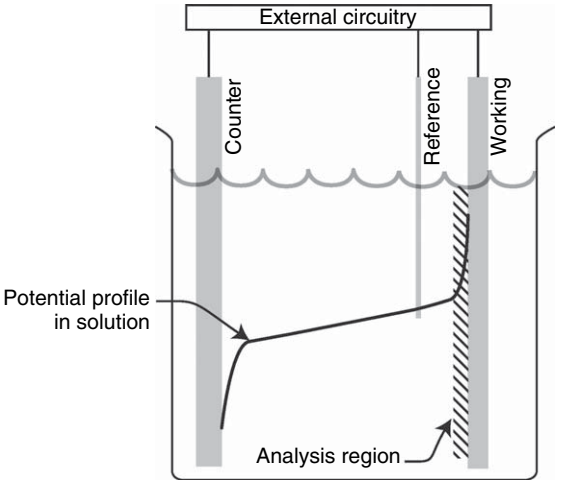


Figure 3 Electrochemical cells normally contain at least two electrodes: the working electrode and the reference electrode. In experiments that draw significant amounts of current, a third electrode, the counter electrode, may be added to the cell. The profile of an applied potential is also depicted. Most of the potential drop occurs at the surface of the electrodes. However, when current is flowing between two of the electrodes, there may also be a potential drop in the bulk of the solution due to the resistance of the solution. (Source: Strong T D 2004 Microfabricated voltammetric neuro-arrays for use *in vitro*. Ph.D. thesis, University of Michigan.)

always present in every electrochemical cell, the indicator or working electrode and the reference electrode. Normally, the reaction at the indicator/working electrode is of interest. The conditions at this electrode are carefully controlled in order to favor detection of a single analyte. If little or no current flows through the electrode as in potentiometry, then this electrode is referred to as the indicator electrode. In cases where significant current flows through the electrode as in voltammetry, the electrode is called the working electrode (McNaught and Wilkinson 1997).

The second electrode, the reference electrode, is designed to develop a well-defined and reproducible potential when placed in solution. A good reference electrode should be stable over long periods of time and should show very little hysteresis when small currents are passed through it or after temperature cycling (Sawyer *et al.* 1995). The NHE, discussed previously, is a reference electrode that is formed by bubbling hydrogen across the surface of a platinum electrode. However, due to the complexity of the NHE, self-contained reference electrodes such as the saturated calomel electrode (SCE) and the Ag/AgCl electrode are more commonly used. These self-contained reference electrodes normally consist of an inner element (either Hg/Hg_2Cl_2 (calomel) or Ag/AgCl) in an appropriate filling solution (often saturated KCl) connected to the external solution by a liquid junction such as a capillary or a porous glass frit (McNaught and Wilkinson 1997). Of these two electrodes, the Ag/AgCl electrode is used almost exclusively in microsystems, and there is an extensive body of literature describing the manufacture and characterization of miniature Ag/AgCl electrodes (Bousse *et al.* 1986, Dendo 1994, Huang and Huang 2002, Strong 2004).

The third electrode, the counter or auxiliary electrode, is normally used in methods that involve a significant current flow through the working electrode. This electrode is normally significantly larger than the working electrode and is made from a polarizable, inert material such as platinum. The only purpose of the counter electrode is to provide a complete circuit allowing the current to flow between the working electrode and the counter electrode as reference electrodes are not normally well suited to sourcing or sinking significant amounts of current.

2.12.3 Potentiometric Sensors

Potentiometric sensors generate an electric potential that is proportional to the activity of a chemical in solution (usually equivalent to the concentration). The most common of these sensors, ion-selective electrodes (ISEs), are used for measuring the concentrations of small ions. Traditionally, ISEs are formed by separating two liquids with a membrane that inhibits or modifies transport between the two phases in such a way that a potential develops across the membrane (Buck 1978). Classical ISE sensors (see ISE electrode body in **Figure 4**, *left half*) have a solution of known composition (the internal solution) on one side of the membrane and the solution being tested on the other side of the membrane (exterior solution). Because the internal solution is controlled, a stable electrical contact can be made to it using a nonpolarizable electrode. A chloridized silver wire in a saturated KCl aqueous solution is commonly used. In coated wire electrodes and some solid-state ISEs, this internal solution is replaced by a direct metallic contact to the membrane.

Transduction of the unknown chemical concentration occurs at the outer surface of the membrane. For measuring this ion concentration-dependent voltage, one must know the potential of the exterior, or test, solution. A reference electrode is normally employed to read this potential. A liquid junction electrode (see reference electrode body in **Figure 4**, *right half*) is commonly used for this application because it can be designed to have a well-defined potential between its connecting wire and the solution over a wide range of chemical compositions and concentrations. Accurate measurements of the membrane potentials can be made only when the membrane is at equilibrium, and so the electronics used in the measurement need to have a high input impedance and the potential of the test solution should be stable, so as to eliminate electric currents from flowing through the cell and driving the cell away from equilibrium.

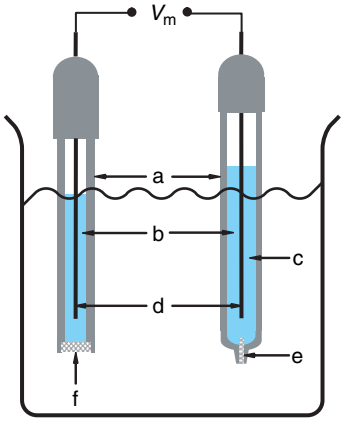


Figure 4 Schematic of a conventional ion-selective electrode (ISE) (*left*) and a reference electrode (*right*) for measuring ionic species in an aqueous solution: (a) electrode body; (b) internal filling solution; (c) reference electrolyte (e.g., 3 M KCl); (d) Ag/AgCl internal reference electrode; (e) liquid junction/porous frit; and (f) ion-selective membrane (glass pellet or polymeric membrane). (Source: Nam H, Cha G S, Strong T D, Ha J, Sim J H, Hower R W, Martin S M, Brown R B 2003 Micropotentiometric sensors. *Proc. IEEE* **91(6)**, 870–80.)

2.12.3.1 Selective Membranes

The ion-selective membranes in conventional ISEs are commonly made of one of the four materials: glass (for pH and alkali ions), pellets of insoluble salts, crystals, or polymers. A brief description of membrane operation is given based on polymeric membranes.

Polyvinyl chloride (PVC) is the most widely used matrix for permselective membranes in conventional ISEs because of its excellent electrochemical properties. PVC, however, exhibits poor mechanical adhesion to the underlying solid passivation layer (usually silicon nitride) of solid-state ISEs. It has been shown that the leading cause of failure in solid-state devices based on PVC membranes is electrolytic shunts around the membrane resulting from the poor adhesion of PVC (Brown 1985). Hence, other polymer matrices such as polyurethane (PU) (Liu *et al.* 1993) and silicone (Malinowska *et al.* 1996) have been developed for use in solid-state chemical sensors (Ryu *et al.* 1994).

Plasticizer, lipophilic salts, and other electroactive ingredients are added to the high molecular weight polymer matrix to give ions mobility in the membrane and to control the membrane's resistivity. The membrane is made selective to a specific ion by the addition of a small weight percent of an ionophore to the membrane. Neutral carrier ionophores (Bakker *et al.* 1997, Buhlmann *et al.* 1998) provide size-selective binding sites for ions of interest, imparting a selectivity to the intended analyte of typically three orders of magnitude over that of interfering ions. Membranes for conventional ISEs are typically formed by mixing the membrane components with an organic solvent, dispensing the membrane cocktail onto a solid surface, evaporating the solvent, and then peeling the polymeric membrane from the casting surface and clamping it into an electrode body. To form coated wire electrodes, the wire can be dipped into the membrane cocktail and the membrane forms as the solvent evaporates. Similarly, microfabricated ISEs are often formed by dispensing the membrane cocktail directly onto a miniature electrode, where it forms a membrane as the solvent evaporates.

Figure 5(a) is an illustration of a membrane cast directly on an electrode; as the membrane contacts the liquid under test and hydrates, water molecules and the analyte ions (in this case, potassium ions) are able to move between the solution phase and the membrane phase. The measured voltage, V_M, includes all the interfacial potentials in the circuit, including the metal–membrane potential. The analyte ions occupy the ionophore binding sites in the membrane, imparting charge to the membrane at the surface. As shown in **Figure 5(b)**, these charges are mirrored by the counter ions in the solution. As in a semiconductor p–n junction, this separation of charge generates a built-in surface potential, E_S, that opposes further diffusion of ions into the membrane. The rates of absorption and desorption of ions are brought into balance at a surface potential that is logarithmically related to the activity of the ions in the solution, as given by eqn [3]. At 298 K, for an ion that exchanges a single electron, the change in E_S in response to changes in concentration of the selected ion is approximately 59.2 mV per decade, i.e., for every 10-fold increase in ionic concentration, E_S increases by 59.2 mV; ions that exchange two electrons generate a potential difference of half this magnitude, i.e., 29.6 mV per decade.

2.12.3.2 Micropotentiometric Sensors

Conventional ISEs have traditionally been about 1.5 cm in diameter and 10 cm in length; they are liquid filled and are expensive and hence must be used for an extended time. Conventional ISEs also require periodic refilling of the internal solution and must be

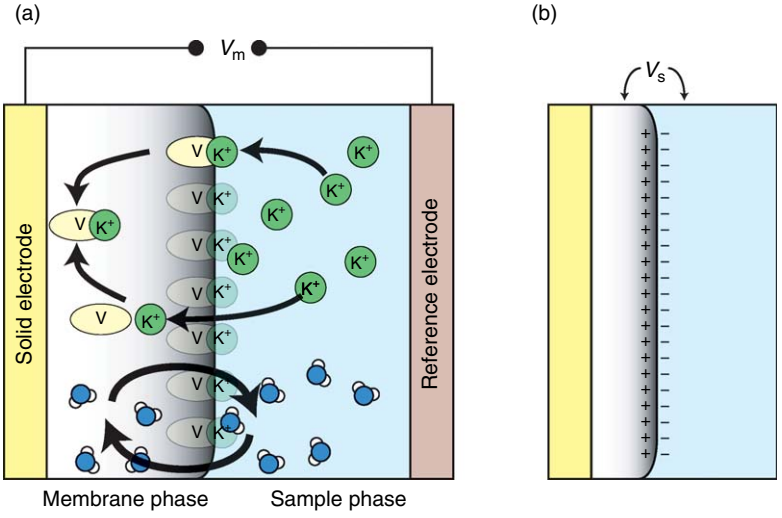


Figure 5 Representation of a potassium ion-selective micropotentiometric sensor having a polymeric membrane with valinomycin (V), a neutral carrier ionophore, in direct contact with the metal electrode. (a) The movement of water and potassium ion between the aqueous and membrane phases. (b) The charge separation that results in the surface potential, which is measured.

stored in the same solution between measurements in order to prevent the membrane from drying.

Microfabrication technologies opened the possibility of miniaturizing ISEs and mass-producing them so that they could be made inexpensively enough to be used only once, or at least used for a short period of time, so that no maintenance is required. The first report of thin-film technology being applied to chemical sensors was Bergveld's (1970) ion-sensitive field effect transistor (ISFET), which is shown in **Figure 6** (*left half*). This device merged the same solid-state electronics technology used to fabricate transistors with chemical sensors by including a field effect transistor (FET) under a chemically selective glass layer. The principles of operation for these devices were clarified over the following decade (Janata and Huber 1980). Researchers have modified the original ISFET by adding metal or polysilicon gates between the transistor channel and the chemical-selective layer (Smith 1982), by incorporating organic polymer membranes instead of silicon dioxide as the chemical selective layer, by coating the ion-selective layer with enzymes, and by using a gate material having a work function that can be modulated chemically (Jakobsen *et al.* 2002, Janata 2004, Shepherd and Toumazou 2005). This whole set of FET-based chemical sensors has been called chemical field effect transistors (CHEMFETs).

An alternative approach for microfabricating ISEs is the use of thick-film (screen printing) technology on ceramic, glass, or plastic substrates (Yoon *et al.* 2000). The dimensional control of photolithography is roughly two orders of magnitude better than that of screen printing, so the thick-film features cannot be as small or as well controlled as features on thin-film-fabricated sensors. However, thick-film manufacturing is very cost-effective, and can form sensors that are small and inexpensive enough to be used just once. The use of a ceramic, glass, or plastic substrates and screen printing to fabricate sensors precludes the option of integrating electronics with the sensors on the same substrate.

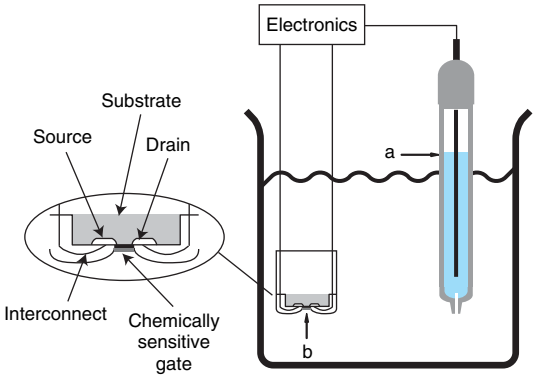


Figure 6 Schematic of an ion-selective field effect transistor (ISFET) in use: (a) reference electrode, as in **Figure 4** and (b) ISFET device. (Source: Nam H, Cha G S, Strong T D, Ha J, Sim J H, Hower R W, Martin S M, Brown R B 2003 Micropotentiometric sensors. *Proc. IEEE* **91(6)**, 870–80.)

The chemical-selective membranes used for potentiometric sensing have high impedances. Reducing the area of the sensor increases the impedance even further. These very high impedances make the sensor system extremely susceptible to noise pickup. Transistor-based sensors such as the ISFETS described above have the advantage of lower output impedance, which should improve noise performance. Unfortunately, MOSFETs (and CHEMFETs) have their own noise susceptibilities, being sensitive to light, temperature, voltage, and semiconductor parameter variations. CHEMFETs have often been used in differential circuits that employ a true MOSFET as the other input device, to compensate for the semiconductor-related noise. Unfortunately, since CHEMFETs and MOSFETs have different gate materials, they are not well matched and the compensation is far from ideal (Hammond *et al.* 2004).

The interposition of a metal or a polysilicon gate material between the chemically sensitive layer and the gate insulator is a substantial improvement to the FET-based devices, as it causes the gate potential to be uniform across the channel. A metal-gate ISFET is exactly analogous to a coated wire electrode driving a single-transistor amplifier (with a very short wire). The next step in the evolution of these active chemical sensors was the use of an operational amplifier (op-amp) instead of the single-transistor amplifier (Brown *et al.* 1985, Lauks *et al.* 1985) (**Figure 7**). With the op-amp configured as a voltage follower, the input impedance is increased by the open-loop gain of the amplifier (typically several orders of magnitude), and the output impedance is reduced by the same amount. The output signal from the op-amp-based sensor is a voltage that tracks the membrane voltage, whereas in the ISFET, the output signal is a current that is quadratically related to the change in gate voltage. The feedback of the voltage follower configuration also improves the stability (insensitivity to a change in circuit parameters) of the electronics by a factor equal to the open-loop gain of the amplifier. The input transistors of the op-amp can be small and well matched for good compensation over temperature, aging, and exposure to light.

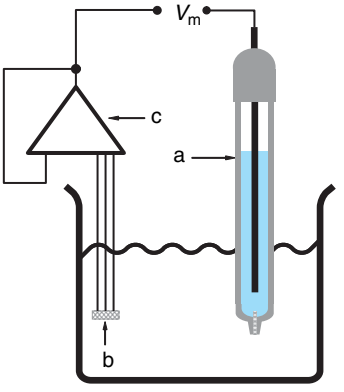


Figure 7 Solid-state ion-selective electrode (ISE) with electronic buffering: (a) reference electrode, as in Figure 4; (b) ISE; and (c) operational amplifier buffer. (Source: Nam H, Cha G S, Strong T D, Ha J, Sim J H, Hower R W, Martin S M, Brown R B 2003 Micropotentiometric sensors. *Proc. IEEE* **91(6)**, 870–80.)

The benefits of integrating op-amps with potentiometric sensors have been realized by placing the sensor directly on top of the amplifier (Brown *et al.* 1985, Milgrew *et al.* 2005) and by integrating the amplifiers on-chip, but separated from the sensors (Lauks *et al.* 1985). As described in more detail in Section 2.12.5, the advantages listed above can also be realized by locating the op-amp very close to the transducer, but not on the same substrate. As noted above, the membrane impedances of microsensors are high, so the leads between the transducer and the amplifier must be either short or well shielded to avoid ambient electromagnetic noise. When a sensor has a long expected lifetime or the application demands a very small size, integration of the electronics with the transducer may be justifiable. On the other hand, when a sensor is disposable or has a short lifetime, manufacturing the sensor and electronics on separate substrates may be more practical and cost-effective.

2.12.3.3 Micropotentiometric Examples

Examples of polymeric membrane-based potentiometric sensors made using thick- and thin-film technology, both with and without integrated electronics, are presented here. Examples of different polymer matrices are given. Some of these sensors are microfabricated, planar coated wire electrodes. Others have hydrogel buffer layers between the membrane and the metal contact to stabilize the membrane-contact potential, and some have integrated op-amps.

2.12.3.3.1 Microfabricated coated wire electrodes

Microfabricated coated wire electrodes might appear to be just miniaturizations of conventional ISEs, but as noted previously, conventional ISEs have an internal solution designed to establish a well-defined Nernstian potential between the solid phase and the liquid phase of the electrode. Coated wire electrodes and ISFETs, on the other hand, lack an analogue to the internal filling solution of conventional ISEs and instead rely on direct contact between the internal electrode and the membrane. This solid contact does not establish a good thermodynamic equilibrium between the two materials and is a source of sensor drift in these devices. Better stability at the membrane–solid interface has been shown when the membrane is adhered tightly to the solid contact. Hower *et al.* (1995) demonstrated that the rough surface of silver epoxy is an effective anchor for the membrane, and Lutze *et al.* (1999) found that the incorporation of a silver ion-containing ligand into the membrane stabilized the membrane-contact potential for some ions, but this approach is not a generalizable solution.

Good examples of micropotentiometric coated wire ISEs are the calcium and chloride sensors described by Hower (2005). These sensors were fabricated on a silicon substrate. A layer of silicon dioxide was grown on the wafer, heavily doped polysilicon was deposited and patterned as interconnect, another layer of silicon dioxide was deposited, and then a silicon nitride layer formed by low-pressure chemical vapor deposition (LPCVD) was deposited as an encapsulant. Contacts were etched down to the polysilicon at the sensing sites and at the bonding pads, and a thin layer of silver was deposited in these openings. A 50-μm-thick layer of silver epoxy was then screen printed as the solid contact at the sensing sites using a Nu-Long microscreen printer. The silver was chloridized, and chemically selective membranes were screen printed directly over the solid contacts and the surrounding silicon nitride surface. The optimized calcium ion-selective membrane was made using 2 mg ETH129 as the ionophore (1 wt.% after solvent removal); 1.5 mg potassium tetrakis(4-chlorophenyl) borate (KTpClPB) to decrease the anionic response; 126 μl 2-nitrophenyl octyl ether (NPOE) plasticizer; 52.8 mg PU and 13.2 mg PVC as the polymer matrix; and 125 μl dimethyl phthalate (DMP) solvent (which is evaporated during membrane curing). As seen in the calibration curve in **Figure 8(a)**, the resulting calcium ISE had a slope of 29.9 mV per decade over a range of

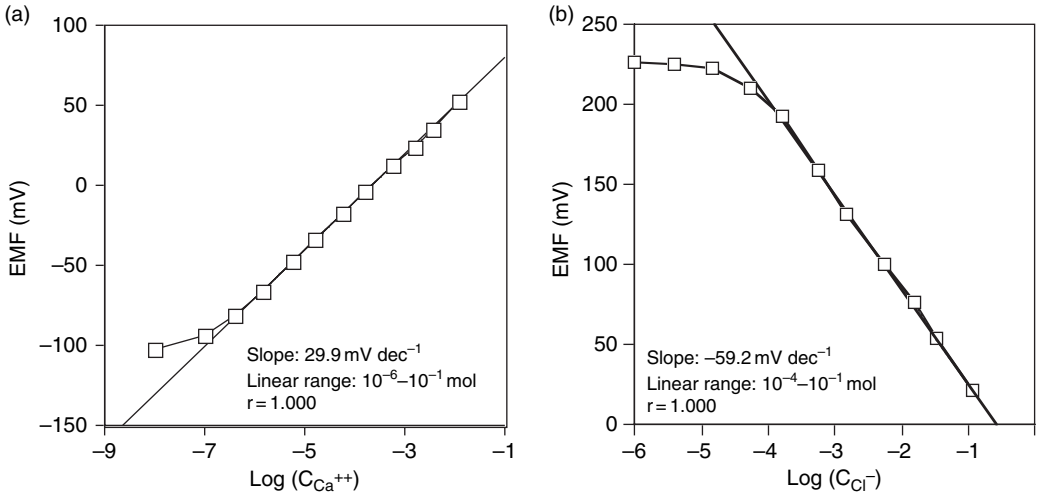


Figure 8 Calibration curves for (a) passive calcium and (b) chloride sensors made on silicon substrates with polyurethane (PU)/PVA membranes, polysilicon interconnect, and silver–epoxy contacts. The ionophores used were ETH129 for calcium and tetradodecylammonium bromide (TDDABr) for chloride. (Source: Hower RW 2005 The integration of potentiometric and optical chemical sensor arrays. Ph.D. thesis, University of Michigan.)

10^{-6} to 10^{-1} mol. The chloride-selective membrane whose calibration curve is shown in **Figure 8(b)** was made using 2 mg tetradodecylammonium bromide (TDDABr) (1 wt.% after solvent removal) as the ionophore and the same components as in the calcium membrane except that it did not include KTpClPB. The slope was −59.2 mV per decade over a range of 10^{-6} to 10^{-1} mol.

2.12.3.3.2 Micropotentiometric sensors with internal layers

As stated previously, the membrane–solid interface in coated wire-type electrodes can be a source of drift in microfabricated ISEs. This is of particular concern for pH sensors or for other membranes that have a secondary pH sensitivity because dissolved carbon dioxide gas and water can permeate the thin membrane and form an unbuffered, slightly acidic intermediate layer at the membrane–solid interface. If either the membrane or the electrode is pH-sensitive, then this can lead to a significant signal drift (Han *et al.* 2001). Hence, buffered hydrogel layers are usually formed beneath pH-selective polymeric membranes, and are often used under all polymeric ion-selective membranes to stabilize the membrane–solid interface. The one serious problem with this approach is that the buffer layer can develop a large osmotic pressure, which can lead to the detachment of the membrane unless design steps are taken to prevent this (Lesho and Sheppard 1996).

An example of sensor arrays made using internal hydrogel layers are those commercialized by Sensicore Inc., Ann Arbor, MI, USA, to monitor pH, ammonia, calcium, and a number of other parameters pertinent to potable water quality. **Figure 9** is the layout of an early version of a sensor array of this type. **Figure 10** is a dynamic calibration curve for pH and calcium taken from a sensor of this type. The Sensicore chips include

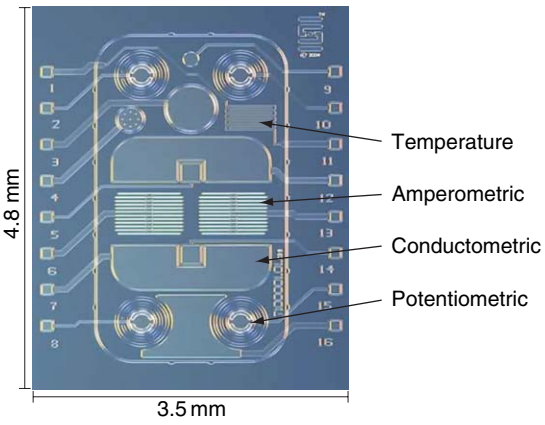


Figure 9 Design of a mixed sensor chip on silicon (potentiometric, voltammetric, conductivity, and temperature sensors) for monitoring water quality. The potentiometric sensors have internal hydrogel buffer layers. (Source: Sensicore, Inc. 2006 *Sensor Test Kits* [Brochure]. Sensicore, Inc., Ann Arbor, MI, retrieved February 8, 2007 from http://www.sensicore.com/data/sensortestkits11-06.pdf.)

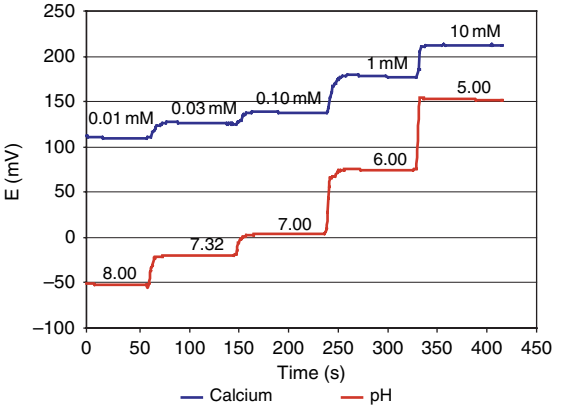


Figure 10 Dynamic calibration plots for calcium and pH ions from silicon substrate solid-state sensors having polyurethane (PU)–polyvinyl chloride matrix membranes. (Source: Sensicore, Inc., 2006 *Sensor Test Kits* [Brochure]. Sensicore, Inc., Ann Arbor, MI, retrieved February 8, 2007 from http://www.sensicore.com/data/sensortestkits11-06.pdf.)

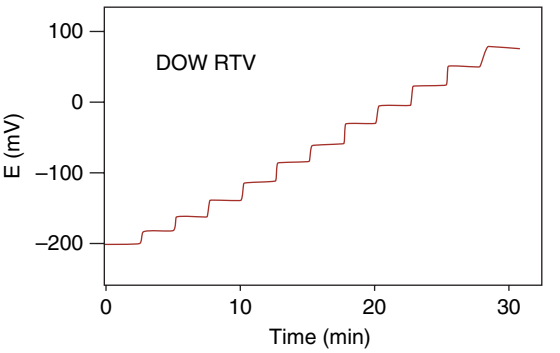


Figure 11 Dynamic calibration plot for a potassium-selective RTV 3140 membrane having 1 wt.% KTpClPB and 2 wt.% valinomycin. (Source: Brown R B 1998 Solid state liquid crystal sensors. *Proceedings of the 4th Annual Symposium Chemistry Forum '98*, Warsaw, Poland, pp. 120–6.)

passive (i.e., no on-chip electronics) potentiometric, voltammetric, conductivity, and temperature sensors (Sensicore 2006).

2.12.3.3.3 Active micropotentiometric sensors

As mentioned above and described in detail in Section 2.12.5, the integration of amplifiers and other circuits onto the same piece of silicon with chemical transducers is possible. As has also been stated, this technologically elegant approach may not always be the best solution from an economic point of view. However, there is another benefit to placing the amplifier close to the transducer, as in an integrated sensor: having the micropotentiometric sensor and the electronics tightly integrated expands the list of usable membrane materials by allowing the use of membranes with higher intrinsic impedances.

In Section 2.12.3.1, silicone was suggested as an alternative to PVC for membrane matrices because of its adhesion to the surface of solid-state sensors. Single-component silicones, such as silanol-terminated polydimethylsiloxane (Petrarch Systems, Bristol, PA, USA) and room temperature vulcanizing (RTV) silicones (Dow Corning, Midland, MI, USA), have been shown to have excellent adhesion to silicon nitride. The adhesive force holding silicone membranes onto the sensor surface was shown by Yoon *et al.* (1999) to be more than 50 times than that in a PVC membrane. **Figure 11** is a dynamic calibration plot for a potassium-selective RTV 3140 membrane in a conventional ISE body made with 1 wt.% KTpClPB and 2 wt.% valinomycin. Unfortunately, silicone has a very high resistivity, leading to large impedances, slow responses, and poor detection limits. This problem is exacerbated in microsensors because the membranes are small in diameter and impedance is inversely proportional to area.

Integration of complementary metal oxide semiconductor (CMOS) electronics onto the same substrate as in the potentiometric sensors has been shown to alleviate the disadvantages of silicone membranes. Specifically, placing a CMOS op-amp near the sensor reduces the interconnect capacitance between the membrane and the instrumentation, and provides a very high input impedance to the instrumentation. The result is a significantly reduced electrical time constant for the sensing system. **Figure 12** shows the difference in dynamic response among a conventional microsized ISE body, a passive solid-state ISE, and an active solid-state ISE with a CMOS-integrated, unity-gain buffer. Each electrode consisted of identical RTV 3140 membranes with 2 wt.% valinomycin and no lipophilic additives. In this case, the op-amp was integrated on the sensor chip, but nearly the same effect could be achieved by a hybrid sensor having the electronics very near the transducers.

2.12.3.4 Comparison of Potentiometry and Voltammetry

The potentiometric sensors described in this section are used to detect small ions such as hydrogen (H^+), potassium (K^+), and chloride (Cl^-) ions. Ions that need to be detected at very low levels or that have

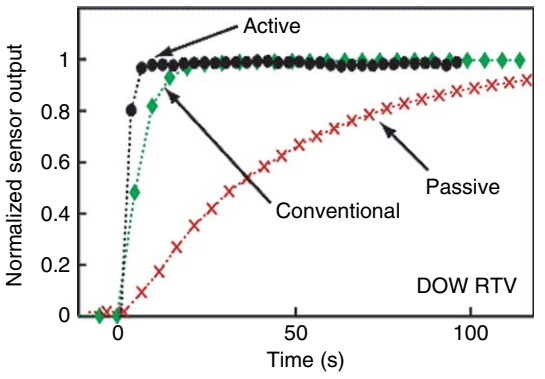


Figure 12 Dynamic response to a ten-fold change in concentration for silicone potassium-selective membranes in a conventional ion-selective electrode (ISE) body, on a passive solid-state sensor, and on an active (op-amp buffered) solid-state sensor. (Source: Martin S M, Ha J, Kim J, Strong T D, Cha G, Brown R B 2004a ISE arrays with improved dynamic response and lifetime. *Tech. Digest Solid-State Sensor, Actuator, and Microsystems Workshop*, Hilton Head Island, SC, USA, pp. 396–9.)

poor ionophores as well as uncharged species are not usually measured using potentiometric sensors. Many of these analytes such as phenols, nitrate, and the heavy metals, however, can be oxidized and reduced at characteristic potentials so that their concentrations can be monitored with voltammetric sensors, which are described in Section 2.12.4. Solid-state chemical sensors of both potentiometric and voltammetric types can be fabricated with the same photolithographic processes, leading to the integration of both types of sensors in a single device.

2.12.4 Voltammetry

Many chemical species are electroactive and will be either reduced or oxidized when an electric potential is applied between a set of electrodes immersed into a solution containing the species. These redox reactions transfer electrons, generating a current flow between the electrodes through the solution. This current, referred to as the faradaic current, i_f, is related to the concentration of the material undergoing the redox reaction by the following equation:

$$i_f = -nFAD\left(\frac{\partial C(x,\,t)}{\partial x}\right)\Bigg|_{x=0} \qquad [4]$$

where A is the area of the electrode involved in the reaction, D is the diffusion coefficient, and $C(x, t)$ is a function describing the concentration of analyte at a distance x from the electrode at time t.

2.12.4.1 Principle of Operation

To perform a voltammetric analysis, an excitation voltage is typically applied between two electrodes, and the current flowing through the cell is measured. Redox reactions must occur at both the electrodes in order to support the current flow; however, only the reaction at the working electrode is important. In most voltammetric analyses, the working electrode is a polarizable electrode, and the test solution is a strong electrolyte and thus a good ionic conductor. If the test solution is not a strong electrolyte, a background electrolyte such as NaCl may be added to increase conductivity. Under these conditions, the voltage applied to the solution primarily drops across the double layer. Since the double layer at the working electrode is extremely thin ($\ll 1\,\mu m$), large electric fields are developed in this area, providing the energy necessary to cause electron transfer and to drive the resulting redox reaction of the analyte material.

As discussed earlier, a generic redox reaction follows the form described by eqn [1] and the applied potential at which the rates for the forward and reverse reactions are equal and at which no net current flows in the cell is determined by the Nernst equation (eqn [2]). If an external voltage is applied to the cell using the counter electrode, the forward and reverse rates for the redox reaction change, resulting in an increase of either O or R at the surface of the electrodes and a net current flow. Although eqn [2] applies only to a system at equilibrium, when the cell is driven away from equilibrium, the basic relationship described by eqn [2] still holds and the approximate ratio of O to R can be found by rearranging terms to arrive at the following equation:

$$\frac{C_O}{C_R} = \exp\left(\frac{(E^0 - E)nF}{RT}\right) \qquad [5]$$

From this relation it is apparent that the input voltage required to push the reaction entirely to one side would be infinite. However, since the relationship is logarithmic, the change from having nearly all O to nearly all R at the surface is a very small change in voltage. In fact, if the material exhibits a single electron reaction, an E that is only 177 mV greater than E^0 is large enough to make the ratio of a material 1000 times greater than the other at the surface of the electrode (Kissinger *et al.* 1996).

The rate at which the reaction proceeds and the magnitude of the faradaic current that flows are controlled by the following factors: the potential applied between the working electrode and the counter

electrode; the electron transfer kinetics of the redox reaction; adsorption of ions on the surface of the electrode; and mass transfer of the reactant to the surface of the electrode and the product away from the electrode. Mass transfer occurs due to convection, diffusion, and migration. If the cell is mechanically isolated from vibration, mass transfer due to convection can be mostly ignored, and if the background ionic strength of the solution is at least two orders of magnitude larger than that of the analyte, then most of the ionic current will be carried by background ions and mass transfer due to migration can also be ignored. Assuming that mass transfer occurs mostly due to diffusion and that the diffusion of analyte to the electrode surface dominates the other rate-controlling steps, a closed-form solution of eqn [4] can normally be found that approximates the relationship between the concentration of the analyte and the observed faradaic current.

The remainder of this section describes several of the voltammetric methods most commonly used with microfabricated sensors. All these methods adhere to the principles outlined in the preceding paragraphs, but differ in the excitation waveforms applied to the cell. The simplest method, chronoamperometry, which is presented first, consists of a single voltage step. After chronoamperometry is presented, an analysis of more complex methods such as linear sweep voltammetry (LSV), cyclic voltammetry, square wave voltammetry (SWV), and square wave anodic stripping voltammetry (ASV) are presented.

2.12.4.2 Chronoamperometry

Chronoamperometry is one of the simplest voltammetric methods and yet it is one of the most frequently used voltammetric methods. Chronoamperometry, in its most basic form, consists of applying a single voltage step at time t_0 and then measuring the current that results from the applied potential. The simplicity of chronoamperometry makes it an ideal technique to analyze it as an example of a basic voltammetric operation.

Consider a simple electrochemical cell consisting of a working electrode, a counter electrode, and a beaker of solution containing only the oxidized form of the analyte, O, at a concentration of C_O^*. An initial voltage is applied between the working electrode and the counter electrode such that no electrochemical reactions occur at the working electrode surface and no net current flows in the cell. At time t_0, the applied potential is stepped to a new voltage, E_a, that is 177 mV more negative than E^0, causing almost all of the O at the surface of the electrode to be instantaneously reduced to R and causing any additional O that arrives at the surface of the electrode via diffusion to be immediately reacted so that the concentration of O at the surface of the electrode is essentially zero. After a period of time, t_S, the applied potential is stepped back to a voltage that allows the R at the electrode surface to oxidize back to O, and the experiment is concluded.

If the electrode is much larger than the molecules of O and if the volume of the solution is large compared to the exposed area of the electrode, then a 1D solution of Eqn [5] will suffice, and the concentration profile $C_O(x, t)$ in the solution during the period when the applied voltage is E_a will approximate the series of curves illustrated in **Figure 13**. During the experiment, the concentration of O at the surface of the electrode, $C_O(0, t)$, is essentially zero while the concentration of O in the bulk solution, $C_O(\infty, t)$, remains at the original concentration, C_O^*. Unreacted O diffuses down the concentration gradient to the surface of the electrode where it gains an electron. As time passes, the region depleted of O by the reaction at the electrode extends deeper into the solution. The thickness of this depletion region is proportional to the diffusion coefficient, D.

The amount of charge, Q, transferred during the reaction is related to the number, N, of moles of O reduced to R by the following relationship:

$$Q = nFNA \qquad [6]$$

Differentiating Q with respect to time, we arrive at an equation for the faradaic current flowing through the electrode at time t

$$i_f = \frac{dQ}{dt} = nFA\frac{\partial N}{\partial t}\bigg|_{x=0} \qquad [7]$$

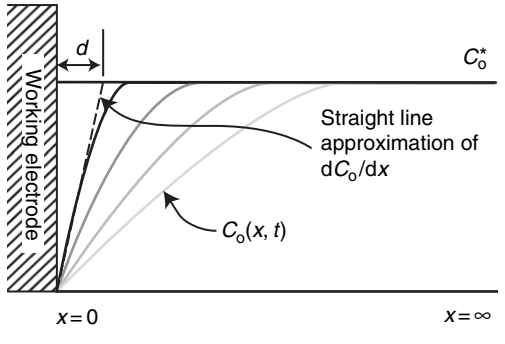


Figure 13 Concentration profile of O at various times after the application of E_S. Note that the slope of the concentration profile can be approximated by a straight line. (Source: Strong D D 2004 Microfabricated voltammetric neuro-arrays for use *in-vitro*. Ph.D. thesis, University of Michigan.)

Material diffuses only along the X-axis, so by Fick's law of diffusion in one dimension:

$$\left.\frac{\partial N}{\partial t}\right|_{x=0} = -D\left(\left.\frac{\partial C(x,\,t)}{\partial x}\right)\right|_{x=0} \quad [8]$$

Substituting eqn [8] into eqn [7] results in eqn [4], which is given in Section 2.12.4.

The important thing to stress in eqn [4] is that i_f is proportional to the slope of the concentration gradient at the surface of the working electrode and not to the applied voltage. As long as E_a is large enough to convert all of the O at the surface of the electrode to R, the current that flows will be independent of E_a. If a straight line approximation for $C_O(x, t)$, as shown in **Figure 13**, is used, then the slope at the electrode will be C_O^*/d. Substituting this into eqn [4] gives an approximate relation for the current, which is as follows:

$$i_f \approx nFAD\left(\frac{C_O^*}{d}\right) \quad [9]$$

While only an approximation, a straight line estimate of a concentration profile is often instructive as to the behavior of a voltammetric system, especially for more complex cases. In this case, however, the actual solution of eqn [4] can be found by applying the boundary conditions $C_O(0, 0) = C_O^*$, $C_O(0, t) = 0$, and $C_O(\infty, t) = C_O^*$, providing

$$i_f(t) = \frac{nFAC_O^*\sqrt{D}}{\sqrt{\pi t}} \quad [10]$$

Eqn [10] is commonly referred to as the Cottrell equation and a plot of this response is shown in **Figure 14**. Note that at time t_0, the current is infinite according to eqn [10]. In real systems, i_f is limited by the rate of electron transfer at the surface of the electrode and by the maximum current supply rate (slew rate) of the electronics applying the input voltage. After time t_0, the current rapidly decays at first and then asymptotically approaches zero.

To determine the concentration of an analyte, the researcher typically chooses a time long enough after t_0 so that the current associated with charging the electrode capacitance is negligible, but soon enough so that the current has not decayed too much to take the measurement. The current at this time will be proportional to the concentration of the analyte.

Although chronoamperometry is a very useful technique, it also has several serious drawbacks.

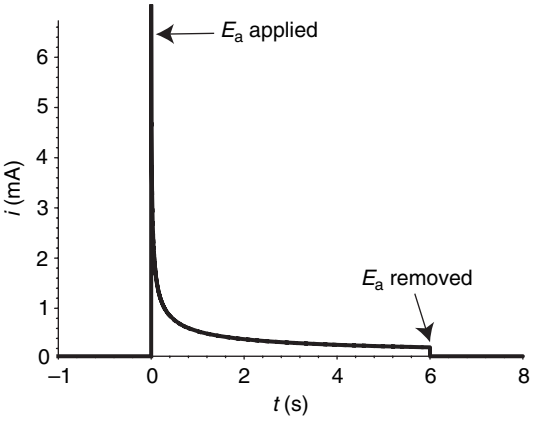


Figure 14 Ideal chronoamperometric current response to step of E_a. Plot assumes a concentration of 1 mM, an electrode area of 1 cm × 1 cm, a diffusion coefficient of 10^{-6} cm^2 s^{-1}, and $n = 1$. (Source: Strong D D 2004 Microfabricated voltammetric neuro-arrays for use *in-vitro*. Ph.D. thesis, University of Michigan.)

The two major drawbacks of chronoamperometry are the following:

(1) It is a relatively slow measurement technique with most measurements requiring perhaps 10 s or longer.
(2) It is not very selective.

The current measured during a chronoamperometric scan is the sum of the faradaic currents produced by all ions in solution that are reduced at a potential more positive than E_a. Consequently, measurement of the concentration of a single analyte in a solution containing several different molecules can be difficult.

2.12.4.3 Linear Sweep Voltammetry

The excitation waveform for LSV is a linear sweep of the applied voltage, E_a, from an initial voltage, E_i, to a final voltage, E_f. E_i is normally selected to be a voltage where no electrochemical reduction of O occurs, and E_f is selected to be a voltage at which so much reduction occurs that $C_O(x = 0) \approx 0$. The analysis from chronoamperometry applies here with the exception that $C_O(0, t)$ is no longer assumed to be zero for the entire time period. Instead, a straight-line approximation of the slope of the diffusion profile

$$\left.\frac{\partial C_O(x,\,t)}{\partial x}\right|_{x=0} \approx \frac{C_O^* - C_O(0,\,t)}{d} \quad [11]$$

can be used to approximate

$$\left.\frac{\partial C_O(x, t)}{\partial x}\right|_{x=0}$$

Combining eqns [4] and [11] gives the following approximation for i_f:

$$i_f \approx -nFAD\left(\frac{C_O^* - C_O(0, t)}{d}\right) \quad [12]$$

The quotient in this equation is the key for understanding the behavior of LSV. As time passes and E_a approaches and then exceeds E^0, the value of $C_O(0, t)$ begins to decrease, causing the numerator to increase. At the same time, the denominator increases as the depletion region extends into the solution. The result is that i_f rises exponentially to a peak near E^0. These conditions apply until the applied voltage is far beyond E^0 so that the surface concentration of O is essentially zero; after that point, the response will resemble a chronoamperometric response, tapering off inversely with $\sqrt{t}$.

The height of the peak is related to how fast both the numerator and the denominator of eqn [12] are changing, and the rate of change in the numerator and denominator is in turn dependent on the rate at which the applied potential is swept. A general linear sweep voltammogram is shown in **Figure 15**. Note that in LSV, rather than plotting the current versus time, the horizontal axis is the applied voltage.

In LSV, the two most important characteristics are E_p, the voltage at which the current peaks, and i_p, the value of the current at that point. If the system is reversible, then the Randles–Sevcik equation applies

$$i_p = 2.69(10)^5 n^{\frac{3}{2}} A C_O^* \sqrt{Dv} \quad [13]$$

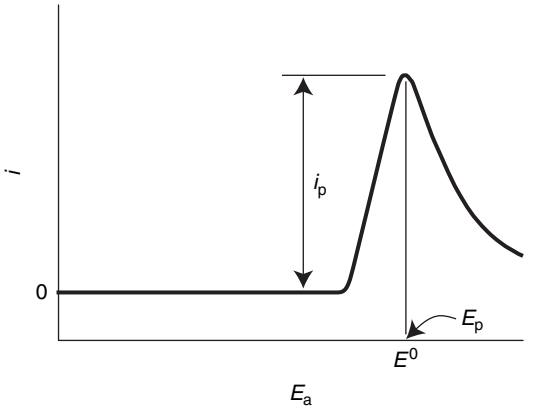


Figure 15 Linear sweep voltammogram. The potential at which the current peaks, E_p, and the current at the peak, i_p, are labeled. (Source: Strong D D 2004 Microfabricated voltammetric neuro-arrays for use *in-vitro*. Ph.D. thesis, University of Michigan.)

where v is the scan rate of the applied voltage (Bard and Faulkner 2000). If the diffusion coefficients for both O and R are approximately equal, then E_p is given by the following equation:

$$E_p = E^0 - \frac{0.029}{n} \quad [14]$$

Since the working electrode has a different E^0 for each ion in solution, each ion in the solution will generate a different current peak at E_p, which is dependent on both the electrode material and the ion. The potentials at which these peaks occur can thus be used with eqn [14] to identify the ions in the solution, and then, using eqn [13], the value of i_p can be used to determine C_O^*. Unfortunately, many chemicals have standard potentials that are close enough to each other that this kind of analysis must be done with caution.

As with chronoamperometry, the above analysis will apply equally well to an oxidation reaction if the voltage is swept in the opposite direction, and the minus sign in eqn [14] is changed to a plus sign. The current will, of course, flow in the opposite direction.

2.12.4.4 Cyclic Voltammetry

Cyclic voltammetry is an extension of LSV. Instead of sweeping the voltage in only one direction, after the linear sweep is concluded, the voltage is immediately swept back to the initial voltage at the same rate. If the analysis starts with a solution consisting of only O, at the conclusion of the first sweep, an amount of R approximately equal to the initial concentration of O will be present at the surface of the electrode, and the reverse sweep will reoxidize (if the reaction is reversible) whatever part of this material has not diffused away from the electrode and produce a current peak similar to the reduction peak, but in the reverse direction. A typical cyclic voltammogram is shown in **Figure 16**. The equations for i_p and E_p are the same as in LSV because the technique is essentially two linear sweeps run back to back.

Cyclic voltammetry has several advantages over LSV. The reverse sweep provides additional information to help identify materials. It also converts material back into its original form, which can prevent the accumulation of unwanted material. Additionally, many facts about the chemical reactions occurring at the surface of the electrode can be inferred from the observed peaks in the voltammogram. For example, if a peak is observed during the forward sweep indicating the reduction of an ion, then on the reverse sweep, a second peak should be observed indicating the reoxidation of the ion back to its original form. If the two

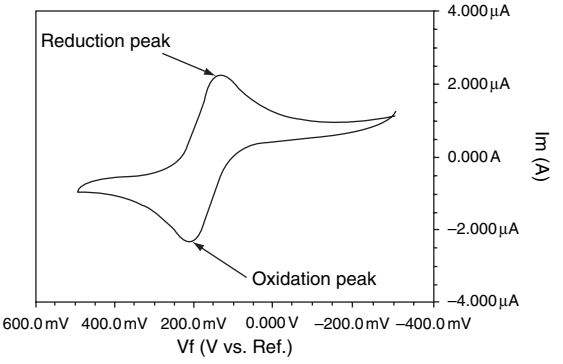


Figure 16 Voltammogram of potassium ferricyanide showing a separation of 50 mV between peaks in the forward and reverse scans. Sweep rate is 1 V s $^{-1}$. Note that the *X*-axis is reversed, the more typical orientation in chemical science texts. (Source: Strong D D 2004 Microfabricated voltammetric neuro-arrays for use *in-vitro*. Ph.D. thesis, University of Michigan.)

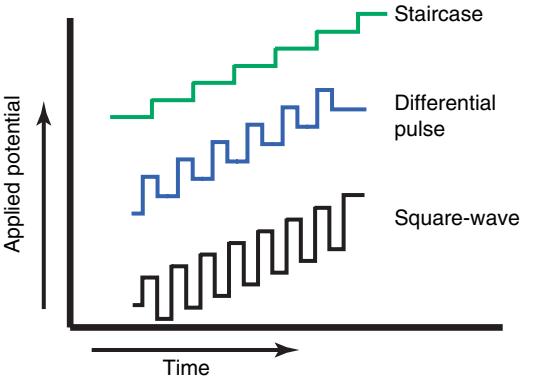


Figure 17 Small-amplitude excitation waveforms. Waveforms are shown for staircase, differential pulse, and square wave voltammetry (SWV). (Source: Martin S M 2005 CMOS-integrated liquid chemical microdetection systems. Ph.D. thesis, University of Michigan.)

peaks are separated by 59 mV then eqn [14] indicates that $n = 1$ for this particular redox reaction and that the reaction is reversible. Likewise, if $n = 2$ for the reaction, then the forward and reverse peaks will be separated by 29.6 mV. Other analyses are also possible and information regarding intermediate products of the reaction as well as transfer kinetics can be observed using cyclic voltammetry. These methods are well documented and information regarding their application can be obtained in the many texts on the subject (Bard and Faulkner 2000, Kissinger *et al.* 1996).

2.12.4.5 Square Wave Voltammetry

The techniques described prior to this point have involved the applications of relatively large potentials. SWV, on the other hand, is a small-amplitude technique that was developed to isolate the contribution from the faradaic signal and the contribution from the parasitic charging currents. Small-amplitude techniques like SWV restrict the changes in the excitation potential to small excursions around a baseline potential. Normally these excursions are typically less than $8/n$ to $12/n$ (mV) for each potential step where n is the number of electrons transferred in the redox reaction, and the baseline around which the potential steps are introduced is often a linear ramp similar to the ones used in LSV. **Figure 17** shows excitation waveforms for three typical small-amplitude voltammetric techniques (Kissinger *et al.* 1996).

In SWV, the faradaic current produced by each potential step is measured at the end of each voltage excursion, and the change in current between potential steps, Δi, is plotted versus the applied potential. The faradaic currents that flow during the positive and negative potential steps are opposite in polarity while the polarity of the charging current, which is mostly due to the ramp voltage, does not change. Thus, the component of the output current due to the faradaic signal is amplified while the component due to the charging current is reduced. Δi is directly proportional to the change in surface concentrations of the electroactive species, and the maximum current difference occurs near E^0 for each electroactive analyte.

2.12.4.6 Square Wave Anodic Stripping Voltammetry

For analytes present in only trace amounts, the peak currents generated using the techniques presented up to this point can be quite small and indistinguishable from noise. To overcome this problem, a technique known as ASV can be used to amplify these currents for heavy metals and other species, which can be electrochemically plated on the surface of an electrode. ASV is a combination of standard voltammetric techniques with a preconcentration phase (Wang 1985). The preconcentration step applies a deposition potential, E_d, on the electrode for a predefined period of time, t_d, as shown in **Figure 18**. The application of E_d reduces O in solution to R, and if R is insoluble in solution, a thin layer of R plates onto the electrode (**Figure 19**). After the plating step, a voltammetric sweep such as SWV or CV is conducted, and the film of R is stripped off the surface of the electrode as it is

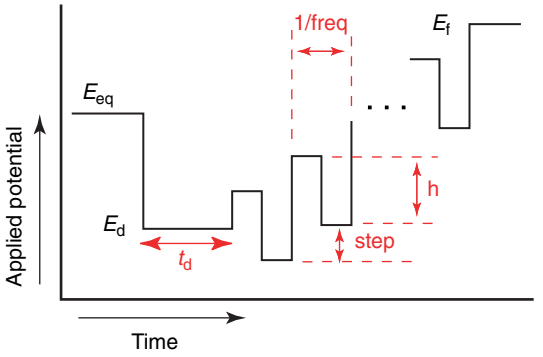


Figure 18 Typical anodic stripping voltammetry (ASV) input waveform. E_{eq} is the equilibrium potential of the solution, E_d is the voltage applied to the solution during the preconcentration step, t_d is the duration of the preconcentration, E_f is the final potential of the sweep, *step* is the step height, *h* is the pulse height, and *freq* is the frequency of the pulses. (Source: Martin S M 2005 CMOS-integrated liquid chemical microdetection systems. Ph.D. thesis, University of Michigan.)

oxidized back to O. The resulting faradaic current is larger than the current produced without plating because the effective concentration of R at the electrode surface is increased. If the voltammetric technique employed during the measurement step is SWV, then the entire process is referred to as square wave anodic stripping voltammetry (SWASV).

2.12.4.7 The Three-Electrode System

The analysis so far has described the use of two electrodes in the solution: the working electrode where the reaction of interest is occurring, and the counter electrode, which completes the electric circuit and allows the application of an external potential. In practice, a reference electrode is often added, as illustrated in **Figure 20**. In a three-electrode electrochemical cell, the applied voltage, v_a, is composed of three parts: v_w, the potential drop at the surface of the working electrode; v_S, the potential drop due to the ionic resistance of the solution; and v_c, the potential drop at the surface of the counter electrode. When performing a voltammetric experiment, only the reactions occurring near the surface of the working electrode are of interest, and in order to measure the quantity of reactants at the working electrode surface, both v_w and i_f must be known. If only a counter electrode and a working electrode are used, i_f can be measured; however, v_w cannot be set independently of v_S and v_v, and only v_a is measurable at the external cell connections.

The unknown contributions from v_S and v_c can be minimized by using a reference electrode such as an Ag/AgCl electrode as the counter electrode in order to obtain a well-defined reference potential for v_c and

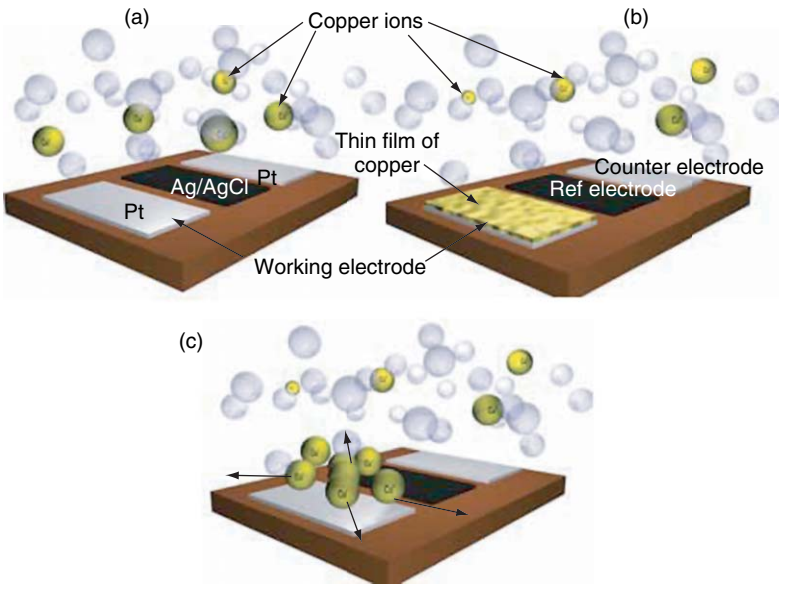


Figure 19 Diagram of anodic stripping voltammetry (ASV). (a) A three-electrode sensor and solution with no cell potential applied. (b) The system during plating with a cathodic potential applied to the cell. A thin film of copper forms on the electrode while the bulk concentration of ions in solution is unaltered. (c) The ions diffusing away from the electrode after a redox reaction has stripped them from the electrode during the voltammetric scan. (Source: Martin S M 2005 CMOS-integrated liquid chemical microdetection systems. Ph.D. thesis, University of Michigan.)

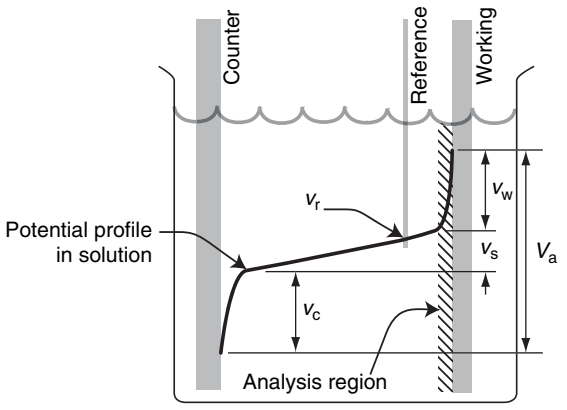


Figure 20 Figure 3 with all potential drops labeled. The analysis region is the area where the redox reaction under study occurs. Note that the interface voltage drops (v_c, v_w) occur in a region near the electrodes that is typically <1 μm thick. (Source: Strong D D 2004 Microfabricated voltammetric neuro-arrays for use *in-vitro*. Ph.D. thesis, University of Michigan.)

by using a high ionic strength solution to minimize the resistance of the solution, R_{sol}, thereby reducing v_S ($v_S = iR_{sol}$). However, in practice, these approaches may not be practical or possible.

Although reference electrodes are normally nonpolarizable, no electrode is ideal, and any current passing through a reference electrode will shift the electrode's reference potential away from its equilibrium value. Additionally, the most commonly used reference electrode in microsystems, the Ag/AgCl electrode, normally consists of only a thin layer of AgCl grown on the surface of an Ag electrode, and the faradaic reaction that occurs when current flows through an Ag/AgCl electrode consumes the layer of AgCl, greatly shortening the life of the electrode. Adding a supporting electrolyte to a low ionic strength solution is not always desirable either, as the supporting electrolyte may interfere with the measurement (such as adding NaCl to the test solution for a chloride sensor), and for many simple automated test systems (such as the remote rainwater pollution-monitoring station proposed by Martin (2005), the addition of an electrolyte would unnecessarily increase the complexity of the system.

The addition of a reference electrode in close proximity to the working electrode removes most of the measurement errors due to v_S, and completely eliminates v_c from the measurement. This removal of v_c is key in microsystems, as the small distance between the counter electrode and the working electrode in a microelectrode device has already reduced R_{sol} to a nearly negligible value. However, as the counter electrode decreases in size, v_c becomes strongly dependent on the current flowing through the cell, making the removal of its influence important.

2.12.4.8 Nonidealities

During the discussion of the more common electrochemical techniques in the previous sections, several assumptions were made that glossed over nonidealities that the reader should be aware of. Two of these nonidealities are discussed in the following sections.

2.12.4.8.1 Parasitic electrode currents

In addition to the current generated by the redox reaction of interest, parasitic currents are also drawn to charge the electrical double-layer capacitance discussed in Section 2.12.4.3, and additional current is generated by interfering redox reactions at the electrode surface by ions other than those being measured. Sometimes these additional currents can be accounted for through calibration, but often they cannot be easily separated from the signal of interest without using a more complex technique such as SWV. It is important to note that both the current drawn to charge the double-layer capacitance and the peak current measured during a voltammetric scan, i_p, are dependent on the scan rate. The parasitic current drawn to charge the double layer

$$i_c = C_{dl} v \qquad [15]$$

is directly proportional to the scan rate, while the peak current (see eqn [13]) is only proportional to the square root of the scan rate. Consequently, while faster scan rates do enhance i_p, i_c increases at a faster rate, negating at least part of the benefit of increasing the scan rate.

On the other hand, both i_c and i_p are also directly proportional to the area of the electrode. However, as the electrode gets very small (<25 μm) nonplanar diffusion effects which will be discussed in the next section increase the flux of ions to the surface of the electrode, increasing the faradaic current above what is predicted by eqn [10].

2.12.4.8.2 Nonplanar diffusion

Eqn [10] is derived by assuming that the electroactive material diffuses to the electrode in a planar fashion. This assumption holds very well in the center of the electrode as shown in **Figure 21(a)**; however, near the edges of a planar electrode,

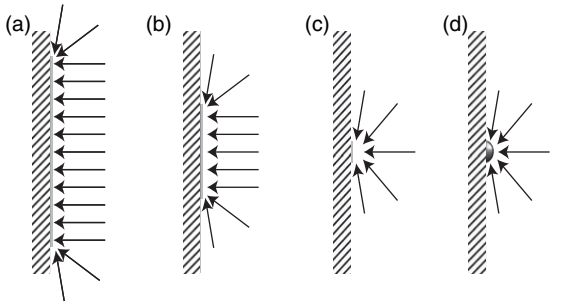


Figure 21 Planar and spherical diffusion to an electrode. Arrows represent material diffusing to the electrode. (a) A large electrode where planar diffusion dominates; (b and c) progressively smaller microelectrodes where planar diffusion no longer dominates; and (d) diffusion to a spherical electrode. Note the similarity between (c) and (d). (Source: Strong D D 2004 Microfabricated voltammetric neuro-arrays for use *in-vitro*. Ph.D. thesis, University of Michigan.)

material diffuses in a nonplanar fashion. As the electrode size is reduced, the contribution from this nonplanar diffusion begins to dominate as seen in **Figure 21(b)** and **21(c)**. Clearly, for these cases, eqn [10] will not hold. If a redox reaction at an electrode has a duration long enough that the electrode's longest dimension is much smaller than $d = \sqrt{Dt}$, the diffusion layer will have extended so far beyond the electrode geometry that the shape of the electrode starts to become less important, and material diffuses toward it in a way that is similar to diffusion toward a hemispherical electrode. For chronoamperometry, the equation for the faradaic current to a hemispherical electrode, such as that in **Figure 21(d)**, is given as follows:

$$i_f - \frac{nFAC\sqrt{D}}{\sqrt{\pi t}} + \frac{nFACD}{r} \qquad [16]$$

where r is the radius of the hemisphere. The first term is the same as in eqn [10], but the additional term is independent of time. Thus the current starts out like in the planar case, but instead of decaying to zero it decays to a steady-state value when the diffusion layer extends much farther into the solution than does r. Absent convection, planar electrodes stimulated long enough will exhibit a very similar current relationship. The case in cyclic voltammetry is similar, with the current being enhanced by the nonplanar diffusion to a higher value. In addition, rather than the current decaying beyond the peak toward zero, it will stabilize at the same steady-state value as in the chronoamperometric case. The enhanced currents of nonplanar diffusion are extremely beneficial, and as a result, in order to reduce charging currents, the electrode is normally made as small as possible without reducing the peak current below the detection limits of the interface circuitry. As the electrodes become very small and the currents generated at the electrodes get even smaller, significant benefits are realized by integrating the chemical sensors directly on top of the CMOS circuitry as described in Section 2.12.5.

2.12.5 Integrated Electronics for Chemical Sensors

Integrating readout electronics with a passive transducer for forming an active chemical sensor can provide many advantages over passive devices. These advantages include a reduction in overall system size; a reduction in system power dissipation; an improvement in sensor detection limit, speed, and linearity; and a more user-friendly interface due to the use of application-specific electronics. Additionally, certain applications where system size and power dissipation are limiting factors, such as implantable devices, might only be viable using active sensors. Integrating circuitry directly with chemical sensors, however, is not without challenges. In general, active sensors are more expensive to fabricate than are passive devices due to the additional manufacturing steps required. This additional cost can be prohibitively large for applications where the chemical sensor is a single-use, disposable device. The integrated readout electronics might also have lower resolution than a bench-top instrumentation system, since the best process for fabricating a chemical sensor might not be the best process for fabricating electronics. Accordingly, thoughtful analysis of the advantages and disadvantages of integration must be considered to decide whether passive or active sensors are a better choice for the application of interest.

Two example applications utilizing active sensors are presented below. Martin *et al.* (2004a) described a system with ISEs and CMOS electronics integrated onto a single silicon substrate. The shortening of the interconnect between the sensor and the amplifier reduced the amount of coupled noise in the system. The percent error of the measured response compared to the theoretical response of the sensors improved from 19% to 5% in moving from passive to active sensors. Schienle *et al.* (2004) demonstrated a single-chip voltammetric DNA sensor that measured faradaic currents resulting from a redox reaction used to identify specific DNA sequences. The chip

included 128 electrodes and in-pixel analog-to-digital converters (ADCs).

In this case, one major benefit of coupling the electronics and sensors was the reduction in off-chip communications achieved by transforming the current into a frequency and encoding this frequency using a digital word.

2.12.5.1 Microelectronic Instrumentation

In the design of active chemical sensors, the microelectronics must be carefully engineered to ensure that the overall system achieves performance greater than or equal to the performance of passive sensors coupled with bench-top instrumentation. This section details the development of a typical microelectronic subsystem for chemical sensing. First, a typical top-level block diagram is presented and then several of the important design considerations of the subcomponents are reviewed.

2.12.5.1.1 Complete microinstrumentation system

Figure 22 shows a block diagram of a typical active chemical sensor. The microelectronics might contain a digital control section that communicates with the data converters in the signal path and the computer or data storage device on the data processing side of the analysis. The signal processing might also be integrated in the digital control block in the form of a microcontroller or a digital signal processor (DSP). In applications where an analog control or stimulus is needed, a digital-to-analog converter (DAC) can be integrated with the system. This analog stimulus is sent to a sensor interface block, which can produce a specified condition on the transducers, provide calibration, or actuate additional components. The transducers may consist of any of the chemical sensors previously discussed. The output of the sensor connects to the sensor interface block. The signal may be amplified in the interface block or may be passed on to a dedicated preamplifier. This amplification helps improve the signal-to-noise ratio (SNR) of the transducer's signal. Further amplification or filtering may be implemented in a signal conditioning block. These circuits drive an ADC, and the digital output code is conveyed to the digital control section. Integrated sensor systems may include only some of these components, and may incorporate other electronic functionality.

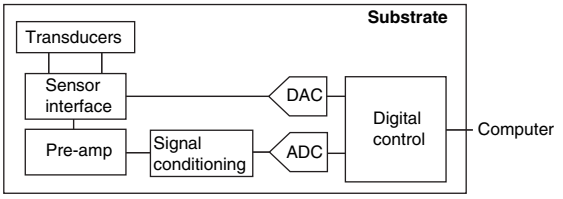


Figure 22 Block diagram of a generic active chemical sensor microsystem.

2.12.5.1.2 Digital-to-analog converter

Voltammetric sensors require a stimulus signal to induce the reduction or the oxidation of a chemical species. These signals can be generated using specialized analog circuits such as triangle wave generators or pulse generators, but a DAC is more versatile in that it can produce any arbitrary waveform such as the complex stimulus signals required by SWV and sinusoidal AC voltammetry (Kissinger *et al.* 1996). A DAC might also be used to generate a calibration or offset cancellation signal. There are many different DAC architectures and the best choice for a given design depends on the application. The output range, resolution, linearity, and power dissipation of the DAC are of primary concern. The glitch time and glitch amplitude as the DAC switches codes is an equally important specification in chemical sensor applications. It is desirable to have a continuous stimulus signal. Fortunately, the low-bandwidth nature of many chemical sensors helps filter glitches in the output of the DAC. The interested reader is directed to Razavi (1995) for a discussion of performance metrics and an analysis of design trade-offs in DACs.

2.12.5.2 Sensor Interface Circuits

The chemical sensor generally dictates which type of interface circuit is used. The following sections detail a variety of sensor interface circuits currently used in modern sensors. The following is not meant to be an exhaustive list.

2.12.5.2.1 Ion-selective electrodes

ISEs are high-output-impedance devices that require high-input-impedance circuits in order to avoid drawing current through the ISE, as even small currents will drive the ISE away from equilibrium. Consequently, the output voltage of an ISE needs to be buffered, which provides an impedance transformation transitioning the signal from a high-impedance to a low-impedance signal. This buffering is commonly accomplished with a CMOS op-amp that uses the gate of a MOSFET as the input terminal

of the amplifier and therefore has an almost infinite input impedance (Gray *et al.* 1993). Nonideal leakage currents through the gate, source, and drain connections of the input transistor, however, must be carefully analyzed to ensure minimal leakage current. Compounding the problem is the fact that gate leakage currents are increasing as process technologies scale (Henson *et al.* 2000).

Two op-amp configurations provide the large input impedance necessary for ISE applications. These two architectures, the unity-gain configuration and the noninverting gain configuration, are shown in **Figure 23(a)** and **23(b)**, respectively. The noninverting gain configuration can be used in applications requiring immediate signal amplification while the unity-gain configuration can be used as a simple buffer. The ISE's output voltage must be referenced to the potential of the solution, which can be obtained from a reference electrode. It is usually advantageous to ground the solution. This can be done through the reference electrode, or the physical connection to ground can be provided through some other solution contact and the reference electrode can be used to measure the solution potential. If, for various reasons, the solution cannot be grounded, an instrumentation amplifier (**Figure 23(c)**) can be used to reject the signal common to both the ISE and the reference electrode. One must ensure, however, that the common-mode level of both signals is within the common-mode range of the amplifier.

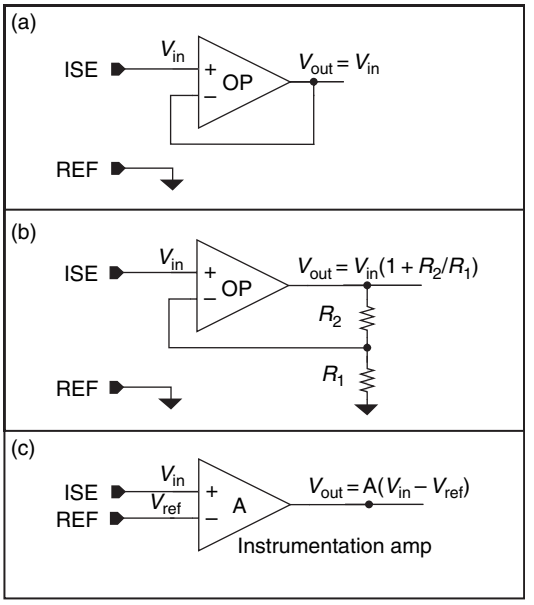


Figure 23 Ion-selective electrode (ISE) readout electronics. (a) Unity-gain buffer; (b) noninverting gain configuration; and (c) differential readout using an instrumentation amplifier. (Source: Martin S M 2005 CMOS-integrated liquid chemical microdetection systems. Ph.D. thesis, University of Michigan.)

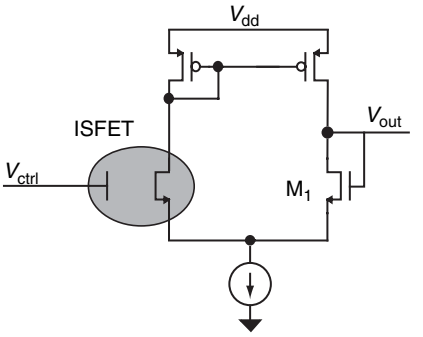


Figure 24 Simplified schematic of a readout circuit for ion-sensitive field effect transistors (ISFETs). (Source: Martin S M 2005 CMOS-integrated liquid chemical microdetection systems. Ph.D. thesis, University of Michigan.)

2.12.5.2.2 Ion-selective field effect transistors

Under a constant gate bias voltage, a change in concentration of analyte will modulate the current in an ISFET. Lauwers *et al.* (2001) provide a detailed schematic of an ISFET interface circuit. **Figure 24** shows a simplified interface circuit for an ISFET. The circuit operates by steering more or less current into the output node, depending on the ISFET's current conducting capability. V_{ctrl} can be set during operation by the DAC to reduce the offset errors presented by the difference in the nominal current-carrying capability of the ISFET and the transistor, M1. ISFETs are more susceptible to changes in operating point, temperature, and charge distribution than are the purely electronic transistors. Consequently, the offset voltage in this circuit is more difficult to eliminate than in the op-amps used with ISEs, and therefore, more sophisticated offset correcting techniques (Razavi and Wooley 1992) are needed for ISFET implementations.

2.12.5.2.3 Voltammetric sensors

Compared to the other types of electroanalytical chemical sensors, voltammetric sensors have the most complicated interface. A specialized circuit, known as a potentiostat, was developed to interface to voltammetric sensors. The standard potentiostat (Bard and Faulkner 2000) is shown in **Figure 25** and functions as follows: The potentiostat ensures that the voltage across the reference and working

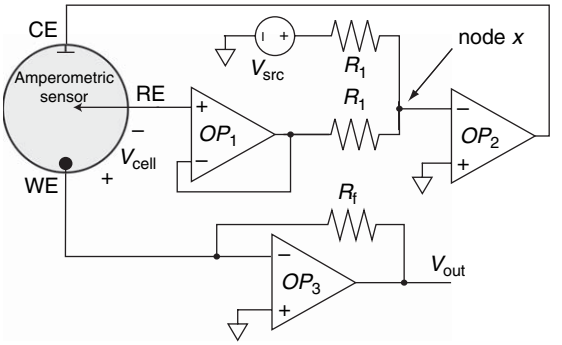


Figure 25 A potentiostat for readout and control of voltammetric sensors. (Source: Martin S M 2005 CMOS-integrated liquid chemical microdetection systems. Ph.D. thesis, University of Michigan.)

electrodes, V_{cell}, tracks an applied source voltage, V_{src}, under varying current-loading conditions. The negative feedback around op-amp OP_3 creates a virtual ground at the working electrode. Thus, V_{cell} is given by the following equation:

$$V_{cell} = -V_{RE} \quad [17]$$

The potential of the reference electrode is buffered by OP_1 to ensure that no current is drawn through the reference electrode. The negative feedback around OP_2 (which is connected via the voltammetric cell) forces node x to ground. The voltage seen at node x is one-half of the sum of V_{src} and V_{RE}. It follows that node x is equal to zero (ground) only when $V_{src} = -V_{RE}$ or by eqn [17] when $V_{src} = -V_{cell}$. Thus, the negative feedback of the potentiostat ensures that V_{cell} tracks V_{src}.

The other function of the potentiostat is to amplify the sensor's faradaic current. This is accomplished with OP_3. The current of the sensor flows through R_f (which is referenced to virtual ground) to provide an amplified current-to-voltage conversion such that:

$$V_{out} = -I_f R_f \quad [18]$$

Many variations on the standard potentiostat have been published. Wei and Lin (1992) demonstrated a chopper-stabilized potentiostat to improve DC offset. Kakerow *et al.* (1995) used switch-capacitor techniques to improve the measurement accuracy. Martin *et al.* (2005a) implemented a pseudo-differential potentiostat to enable real-time subtractive stripping voltammetry, and demonstrated a fully differential potentiostat (Martin *et al.* 2004b) that allowed potentiostats to operate in low-voltage processes. Narula and Harris (2004) demonstrated a potentiostat that enhanced the dynamic range of chemical sensors in low-voltage processes by dynamically translating the sensor signals into time-encoded values. Bandyopadhyay *et al.* (2002) implemented a distributed potentiostat system for neural applications.

Of critical importance in the design of a potentiostat is the noise performance of the signal amplification path. The SNR of a voltammetric sensor increases as the area of the electrode decreases, but the magnitude of the signal decreases as the electrode area decreases (Bard and Faulkner 2000). The electronic readout, however, also introduces noise into the signal. In fact, for microelectrodes, the electronic noise typically dominates other noise sources, as shown in **Figure 26** (Martin 2005). The graph depicts the total output noise of a voltammetric sensor and the current-to-voltage (or transimpedance) gain stage of the potentiostat for a SWV experiment. The data reveal that for typical ranges of electronic noise, the total system noise is relatively flat except when transimpedance amplifiers with gains approximately 10^5 or greater are used. This implies that the sensor noise is about five decades smaller than the electronic noise. Thus, in most scenarios, the electronic noise will dictate the sensor parameters. In an electronic noise-limited regime, an actual optimal electrode area can be found, as shown in **Figure 27** (Martin 2005). Clearly, as electronic noise is decreased, the optimal sensor area decreases and more devices can be fabricated on a single substrate. The bulk of the electronic noise originates in the op-amp, OP_3. The noise performance of this op-amp must be optimized within the context of the other design constraints. Section 2.12.5.5 describes op-amp optimization for chemical sensor applications.

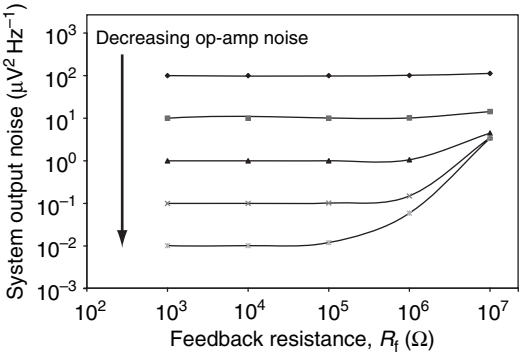


Figure 26 Total output noise of an active chemical sensor with various current-to-voltage gain settings obtained by varying the size of the feedback resistor. (Source: Martin S M 2005 CMOS-integrated liquid chemical microdetection systems. Ph.D. thesis, University of Michigan.)

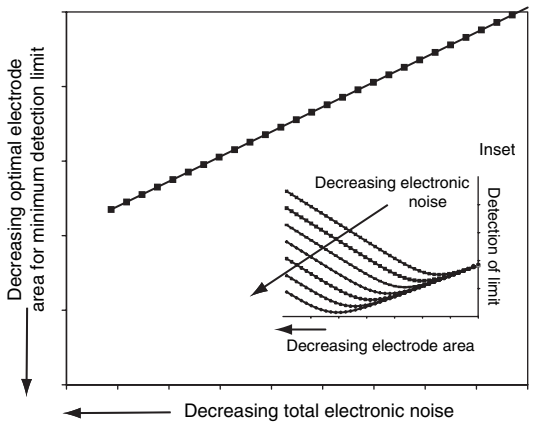


Figure 27 Optimal electrode area of a chemical sensor versus the total electronic noise in the system. (*Inset*) Detection limit of an active sensors versus electrode area and noise used to find the optimal electrode area. (Source: Martin S M 2005 CMOS-integrated liquid chemical microdetection systems. Ph.D. thesis, University of Michigan.)

2.12.5.3 Preamplifier and Signal Conditioning

If the sensor interface circuit does not provide a gain stage, then a preamplifier can be included to raise the signal amplitude. The preamplifier is often followed by a programmable gain amplifier (PGA). The PGA is advantageous in that the dynamic range of the sensor's signal might be so great that a high-gain amplifier would saturate under certain conditions, or a low-gain amplifier might not sufficiently amplify the signal to utilize the full dynamic range of the subsequent ADC. Low-pass or band-pass filters are also commonly used to limit the bandwidth of the signal and to improve its SNR. The signal must also be filtered before it is sampled with an ADC to prevent aliasing due to high-frequency components. Passive filters generally exhibit lower noise performance than do active filters, but some signal attenuation can occur and circuit loading must be carefully analyzed. A single-ended differential converter might also be included in the signal conditioning section as many modern ADC architectures use a differential input stage.

2.12.5.4 Analog-to-Digital Converter

Like DACs, ADCs come in many varieties and the proper ADC architecture often depends on the application of the sensing system. Since many analytical applications demand high precision but are low-bandwidth measurements, integrating ADCs (Breten *et al.* 2000, Reay *et al.* 1994), successive-approximation ADCs (Ryan *et al.* 1995), and oversampled ADCs (Kraver *et al.* 2001) are typical choices for chemical sensing. Important parameters to consider when choosing an ADC architecture include power, area, effective resolution, linearity, and sampling rate. The intricacies of ADCs are well beyond the scope of this work and the reader is referred to Rabii and Wooley (1999) and Razavi (1995) for more details.

2.12.5.5 Operational Amplifiers

Op-amps are the basic building blocks in most of the circuits previously described. The design of op-amps is more than adequately covered in numerous texts on the subject (Gray *et al.* 1993, Johns and Martin 1997). This section highlights some important op-amp parameters with respect to chemical sensing. Many transducers utilize op-amps in unity-gain and noninverting gain configurations to buffer signals. The common-mode range of the op-amp is an important parameter in these situations, as the inputs of the op-amp track the input signal, as opposed to being forced to a virtual ground, as in an inverting gain configuration. Rail-to-rail input op-amps are a common choice for these configurations. The charging currents that occur in voltammetric sensors have the potential to draw a significant current from the op-amp's output. Accordingly, the maximum output current and the slew rate of an op-amp become important. Fortunately, the charging currents scale with the area of the electrode, and the output current of the op-amp is less important for small electrodes. For precision applications, op-amps with high open-loop gains are necessary. The equation for the gain of a unity-gain buffer is given by the following equation:

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{A}{1+A} \qquad [19]$$

where A is the open-loop gain of the op-amp. Clearly, a finite open-loop gain will cause the buffer's gain to be <1.

When optimizing the detection limit of the sensor, the noise of the op-amp plays a critical role. At low frequencies where chemical sensors typically operate, the dominant noise source is the flicker noise. Flicker noise is inversely proportional to frequency and is often referred to as $1/f$ noise. **Figure 28** shows the measured noise spectral density of an op-amp used in the potentiostat presented in Martin *et al.* (2005a). The strong $1/f$ dependency of the noise is evident from the data. Flicker noise is inversely proportional to transistor

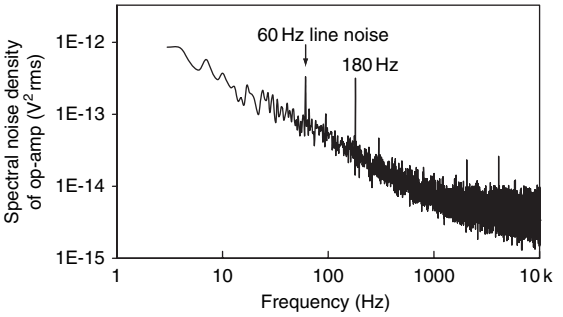


Figure 28 The noise spectral density of a complementary metal oxide semiconductor (CMOS) op-amp over the frequencies of interest in most chemical sensing applications. (Source: Martin S M 2005 CMOS-integrated liquid chemical microdetection systems. Ph.D. thesis, University of Michigan.)

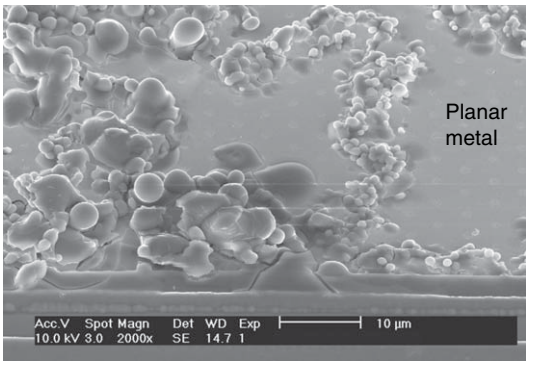


Figure 29 Scanning electron micrograph (SEM) of the surface of a platinum film over an aluminum film after heating, showing the expansion of the metals.

size and bias current. Thus, flicker noise can be reduced by using large input transistors and large bias currents, but the noise performance of an op-amp must be traded off with other design parameters such as device area, power dissipation, and input capacitance.

2.12.5.6 Active Sensor Post-CMOS Fabrication

One method of fabricating active chemical sensors is to postprocess (via thin-film fabrication) the necessary sensor layers on the same silicon substrate as the CMOS electronics. This forces the chemically sensitive layers to be postprocessed after the initial CMOS fabrication is complete and requires that the additional processing steps do not adversely affect the underlying electronics. This method generally allows for the smallest possible device and the shortest interconnect between the electronics and the sensors. Post-CMOS fabrication can be economical as hundreds of devices can be batch-fabricated on a single silicon wafer, but yield will degrade as the fabrication process grows more complex. The fabrication methodologies for many passive sensors can be amenable to post-CMOS fabrication. There are, however, certain aspects of the fabrication process that must be carefully designed to ensure a functional and long-lived device.

In commercially available CMOS processes, the top-level metallization is typically aluminum. The first desired metal in sensor fabrication can be platinum or gold or a similar metal. Unfortunately, aluminum and platinum (and similar metals) are incompatible. At standard fabrication temperatures, the metals interdiffuse and expand in volume (Hwan *et al.* 1998). The expansion (shown in **Figure 29**) causes the normally planar metal to bubble, often fracturing the passivation layer and destroying the device. While some have successfully bypassed this problem by substituting a compatible metal in the place of aluminum during electronics fabrication (Schienle *et al.* 2004), this can alter the performance of the electronics and requires a custom fabrication process for the electronics. An alternative, which continues to use foundry-standard fabrication methods of electronics, is to deposit a metal compatible to both aluminum and platinum as an intermediate step. Titanium tungsten (TiW) is one possible choice, but was found to be incompatible due to film stress by at least one researcher (Strong 2004). Titanium nitride (TiN) is another suitable diffusion barrier for these two metals (Hwan *et al.* 1998).

The passivation layer that is used to protect the electronics, the sensor interconnect layers, and the sensitive substrate from the liquid under test is also important. Many materials have been investigated for this purpose, including thermally grown silicon dioxide and silicon nitride; deposited films such as plasma-enhanced chemical vapor deposited (PECVD) silicon dioxide, silicon nitride, and parylene, and spun-on films such as polyimide and SU-8 (Martin *et al.* 2005b, Nolan and Kounaves 1998, Schmitt *et al.* 1999, Sudakov-Boreysha *et al.* 2004). The thermally grown films are typically the best choices for passivation, but the high temperatures used during their processing (1000°C) are incompatible with the metals used during the CMOS electronics fabrication. Concerns with the other films include adhesion to the substrate, solution uptake, and pinholing in the film. Additionally, most of the problems with these films are exacerbated around nonplanar surface features over metals where stress in the film may induce additional fissures

(Schmitt *et al.* 1999). Reducing surface features via etchbacks or chemical mechanical polishing (CMP) may be necessary in applications where extremely long-use life is required. Additionally, if the active sensor is to be used as a biomedical implant, the coating material must be biologically compatible.

Adhesion of the sensor layers can also be a problem. In most thin-film processing steps where a metal is deposited on a substrate, an adhesion-promoting metal (such as titanium) is first deposited to ensure good mechanical contact. Sometimes, this adhesion promoter (along with the desired electrode material) might have intimate contact with the solution under test. Standard adhesion promoters can be incompatible with the solution, and more exotic materials must be used (Strong 2004).

The openings through the passivation layer to the sensors or from one interconnect layer to another are generally performed using dry etching techniques such as reactive ion etching (RIE). In the RIE steps, charged particles are accelerated toward the surface of the chip and can charge the gates of the CMOS transistors. An excessive charge buildup can destroy the transistor. This phenomenon (along with others) can be classified as process-induced damage (PID). One solution is to use electrostatic discharge (ESD) protection circuits fabricated as part of the electronics to limit the amount of PID that occurs. Care must be taken to properly size these devices as they present a leakage path for the sensor's signals, and the noise from this leakage may be so large in some cases as to completely mask the desired signal.

As an example of a post-CMOS process, **Figure 30** shows the cross section of an active ISE (Martin *et al.* 2005b). The process methodology used to fabricate this device was based on a four-mask, passive sensor process as outlined in Poplawski *et al.* (1991) and Strong *et al.* (2000). Platinum (Pt), gold (Au), and silver (Ag) layers serve as sensor electrodes. A PECVD silicon nitride layer (Si_xN_y) serves as the passivation layer. High aspect ratio SU-8 rings serve as dams for stopping the flow of the bond wire encapsulant unto the surface of the sensor and can alternatively be used to contain the polymer membranes of specific sensors (Hower and Brown 2004). The wirebonds are sealed in a two-part process. They are first coated with MED4211 medical-grade silicone silastic (NuSil Silicon Technology, Santa Barbara, CA) and then with RTV 3140 silicone (Dow Chemical, Midland, MI). An in-depth description of the fabrication is given in Martin *et al.* (2005b). The process is capable of producing many different sensors including voltammetric, potentiometric, and temperature sensors.

2.12.5.7 Multichip Modules

A second approach to integrating electronics and chemical sensors is the multichip module (MCM) approach. In an MCM, a small silicon, ceramic, or laminate substrate is used as a host platform onto which other dies are attached via wire bonding, flip chip bonding, or wafer bonding (Wolffenbuttel 1996). In MCMs, the overall device dimensions remain small and the electronics and sensors can be fabricated in separate processes, each tailored to their specific needs. Defective devices can be sorted prior to integration onto the MCM to ensure that only good devices are packaged. Manufacturing the MCM, however, is often a serial process and can be more costly than integration using batch techniques. Sudakov-Boreysha

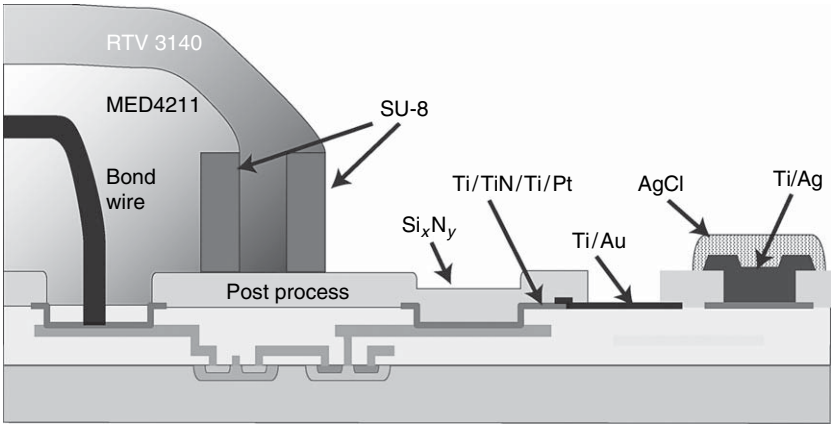


Figure 30 Cross section of a complementary metal oxide semiconductor (CMOS)-integrated active chemical sensor. (Source: Martin S M 2005 CMOS-integrated liquid chemical microdetection systems. Ph.D. thesis, University of Michigan.)

et al. (2004) developed an MCM chemical sensing system with ISFETs using flip chip bonding techniques. Implantable active sensors have also been demonstrated using MCM technology (Meyer *et al.* 2001). The interested reader is directed to Tummala (1992) for a tutorial on MCM devices and technologies. To date, few chemical sensing systems have been fabricated using MCM techniques, but it is expected that more systems will benefit from this approach in the future.

2.12.6 Conclusion

In conclusion, the technology necessary to develop and manufacture microsystems incorporating miniature liquid chemical sensors is rapidly maturing. This chapter has presented some of the theory behind the operation of electrochemical sensors as well as a basic description of the techniques used in fabricating electrochemical sensors. The field of microelectrochemical sensors has already seen great advancements over the last several decades, and it will likely see many more in the coming decades as new products appear on the market incorporating sensors based on these technologies.

References

Bakker E, Buhlmann P, Pretsch E 1997 Carrier-based ion-selective electrodes and bulk optodes. 1. General characteristics. *Chem. Rev.* **97**, 3083–132

Bandyopadhyay A, Mulliken G, Cauwenberghs G, Thakor N 2002 VLSI potentiostat array for distributed electrochemical neural recordings. *IEEE Int. Symp. Circuits and Systems*, Scottsdale, AZ, USA, Vol. 2, pp. 740–3

Bard A, Faulkner L 2000 *Electrochemical Methods*. John Wiley & Sons, New York, NY

Bard A J, Parsons R, Jordan J (eds.) 1985 *Standard Potentials in Aqueous Solution.* Marcel Dekker, New York, NY

Beach R D, Kuster F V, Moussy F 1999 Subminiature implantable potentiostat and modified commercial telemetry device for remote glucose monitoring. *IEEE Trans. Instrum. Meas.* **48**, 1239–45

Bergveld P 1970 Development of an ion-sensitive solid-state device for neurophysiological measurements. *IEEE Trans. Bio-Med. Eng.* **BME-17**, 70–1

Bousse L J, Bergveld P, Geeraedts H J M 1986 Properties of Ag/AgCl electrodes fabricated with IC-compatible technologies. *Sens. Actuators B* **9**, 179–97

Breten M, Lehmann T, Bruun E 2000 Integrating data converters for picoampere currents from electrochemical transducers. *IEEE Int. Symp. Circuits and Systems*, Geneva, Switzerland, Vol. 5, pp. 709–12

Brown R B 1985 An integrated multiple-sensor chemical transducer. Ph.D. thesis, University of Utah

Brown R B, Huber R J, Petelenz D, Janata J 1985 An integrated multiple-sensor chemical transducer. *Proc. Int. Conf. Solid-State Sensors and Actuators,* Philadelphia, PA, USA, pp. 125–7

Buck R P 1978 Theory and principles of membrane electrodes. In: Freiser H (ed.) *Ion-Selective Electrodes in Analytical Chemistry.* Plenum Press, New York, NY, Vol. 1, Chap. 1, pp. 1–141

Buhlmann P, Pretsch E, Bakker E 1998 Carrier-based ion-selective electrodes and bulk optodes. 2. Ionophores for potentiometric and optical sensors. *Chem. Rev.* **98**, 1593–687

Dendo I 1994 Precision silver/silver chloride electrodes. *Proc. 16th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, Baltimore, MD, USA, Vol. 2, pp. 810–11

Gray P R, Hurst P J, Lewis S H, Meyer R G 1993 *Analysis and Design of Analog Integrated Circuits*, 3rd edn. Wiley, New York, NY

Guiseppi-Elie A, Brahim S, Slaughter G, Ward K R 2005 Design of a subcutaneous implantable biochip for monitoring of glucose and lactate. *IEEE Sens. J.* **5(3)**, 345–55

Hammond P A, Ali D, Cumming D R S 2004 Design of a single-chip pH sensor using a conventional 0.6-μm CMOS process. *IEEE Sens. J.* **4(6)**, 706–12

Han J H, Cui G, Kim S J, Han S H, Cha G S, Nam H 2001 Effect of dissolved CO_2 on the potential stability of all-solid-state ion-selective electrodes. *Analyst* **126**, 2040–3

Henson W K, Yang N, Kubicek S, Vogel E M, Wortman J J, De Meyer K, Naem A 2000 Analysis of leakage currents and impact on off-state power consumption for CMOS technology in the 100-nm regime. *IEEE Trans. Electron Devices* **47**, 1393–1400

Hower RW 2005 The integration of potentiometric and optical chemical sensor arrays. Ph.D. thesis, University of Michigan

Hower RW, Brown RB 2004 *US Pat.* 6 764 652

Hower R W, Sin J H, Cha G S, Meruva R K, Meyerhoff M E, Brown R B 1995 New solvent system for the improved electrochemical performance of screen-printed polyurethane membrane-based solid-state sensors. *Proceedings of Transducers '95, Eurosensors IX,* Stockholm, Sweden, pp. 859–62

Huang I, Huang R 2002 Fabrication and characterization of a new planar solid-state reference electrode for ISFET sensors. *Thin Solid Films* **406**, 255–61

Hwan Y, Lee J, Lee S, Koo B, Jung D, Chun Y, Lee M, Shin D, Shin S, Lee S, Kim B, Kang N, Kim K 1998 The effect of Al/Pt interface reaction on lead-zirconate-titanate capacitor and the optimization of via contact for double metal ferroelectric RAM. *Jpn. J. Appl. Phys.* **37**, 1332–5

Jakobson C G, Dinnar U, Feinsod M, Nemirovsky Y 2002 Ion-sensitive field-effect transistors in standard CMOS fabricated by post processing. *IEEE Sens. J.* **2(4)**, 279–87

Janata J 2004 Electrochemical microsensors. *Proc. IEEE* **91(6)**, pp. 861–9

Janata J, Huber R J 1980 Chemically sensitive field effect transistors. In: Freiser H (ed.) *Ion-Selective Electrodes in Analytical Chemistry.* Plenum, New York, NY, Vol. 2, Chap. 3, pp. 107–94

Johns D, Martin K 1997 *Analog Integrated Circuit Design*. John Wiley & Sons, New York, NY

Kakerow R, Kappert H, Spiegel E, Manoli Y 1995 Low-power single-chip CMOS potentiostat. *8th Int. Conf. Solid-State Sensors and Actuators*, Stockholm, Sweden, Vol. 1, pp. 142–5

Kissinger P, Preddy C, Shoup R, Heineman W 1996 Kissinger P and Heinerman W (eds.) *Laboratory Techniques in Electroanalytical Chemistry*, 2nd edn. (revised and expanded). Marcel Dekker, Inc., New York, NY

Koudelka M, Rohner-Jeanrenaud F, Terrettaz J, Bobbioni-Harsch E, de Rooij N F, Jeanrenaud B 1991 In-vivo behaviour of hypodermically implanted microfabricated glucose sensors. *Biosens. Bioelectron.* **6(1)**, 31–6

Kraver K L, Guthaus M R, Strong T D, Bird P L, Cha G, Hold W, Brown R B 2001 A mixed-signal sensor interface microinstrument. *Sens. Actuators A Phys.* **91**, 266–77

Lauks I, Van der Spiegel J, Sansen W, Steyaert M 1985 Multispecies integrated electrochemical sensor with on-chip CMOS circuitry. *Int. Conf. Solid-State Sensors and Actuators,* Philadelphia, PA, USA, pp. 122–4

Lauwers E, Suls J, Gumbrecht W, Maes D, Gielen H, Sansen W 2001 A CMOS multiparameter biochemical microsensor with temperature control and signal interfacing. *IEEE J. Solid-State Circuits* **36(12)**, 2030–8

Lesho M J, Sheppard N F Jr. 1996 Adhesion of polymer films to oxidized silicon and its effect on performance of a conductometric pH sensor. *Sens. Actuators B* **B37(1–2)**, 61–6

Liu D, Meyerhoff M E, Goldberg H D, Brown R B 1993 Potentiometric ion- and bioselective electrodes based on asymmetric polyurethane membranes. *Anal. Chim. Acta* **274**, 37–46

Lutze O, Meruva R K, Frielich A, Ramamurthy N, Brown R B, Hower R, Meyerhoff M E 1999 Stabilized potentiometric solid-state polyion sensors using silver–calixarene complexes as additives within ion-exchanger-based polymeric films. *Fresenius' J. Anal. Chem.* **364(1/2)**, 41–7

Malinowska E, Oklejas V, Hower R W, Brown R B, Meyerhoff M E 1996 Enhanced electrochemical performance of solid-state ion sensors based on silicone rubber membranes. *Sens. Actuators B Chem.* **33**, 161–7

Martin S M 2005 CMOS-integrated liquid chemical microdetection systems. Ph.D. thesis, University of Michigan

Martin S M, Ha J, Kim J, Strong T D, Cha G, Brown R B 2004a ISE arrays with improved dynamic response and lifetime. *Tech. Digest Solid-State Sensor, Actuator, and Microsystems Workshop,* Hilton Head Island, SC, USA, pp. 396–9

Martin S M, Gebara F H, Strong T D, Brown R B 2004b A low-voltage, chemical sensor interface for systems-on-chip: The fully-differential potentiostat. *Proc. 2004 IEEE Int. Symp. Circuits and Systems*, Vancouver, BC, Canada, Vol. 4, pp. 892–5

Martin S M, Gebara F H, Larivee B J, Brown R B 2005a A CMOS-integrated microinstrument for trace detection of heavy metals. *IEEE J. Solid-State Circuits* **40**, 2777–86

Martin S M, Strong T D, Brown R B 2005b Monolithic liquid chemical sensing systems. In: Theil JA, Bhm M, Gardner DS, and Blalock T (eds.) *Materials, Integration and Technology for Monolithic Instruments.* Mater. Res. Soc. Symp. Proc. 869, Warrendale, PA, pp. 109–18

McNaught A D, Wilkinson A 1997 *Compendium of Chemical Terminology*. Blackwell Science, Malden, MA

Meyer J U, Stieglitz T, Scholz O, Haberer W, Beutel H 2001 High density interconnects and flexible hybrid assemblies for active biomedical implants. *IEEE Trans. Adv. Packaging* **24**, 366–74

Milgrew M J, Riehle M O, Cumming D R S 2005 A large transistor-based sensor array chip for direct extracellular imaging. *Sens. Actuators B* **111–112**, 347–53

Nam H, Cha G S, Strong T D, Ha J, Sim J H, Hower R W, Martin S M, Brown R B 2003 Micropotentiometric sensors. *Proc. IEEE* **91(6)**, 870–80

Narula H, Harris J 2004 VLSI potentiostat for amperometric measurements for electrolytic reactions. *Proc. 2004 IEEE Int. Symp. Circuits and Systems*, Vancouver, BC, Canada, Vol. 1, pp. 457–460

Nolan M A, Kounaves S P 1998 Failure analysis of microfabricated iridium ultramicroelectrodes in chloride media. *Sens. Actuators B Chem.* **50**, 117–24

Poplawski M, Cantor H, Midgley A, Brown R B 1991 Microfabricated amperometric biosensors. *Transducers '91: Digest of Technical Papers,* San Francisco, CA, USA, pp. 51–3

Rabii S, Wooley B A 1999 *The Design of Low-Voltage, Low-Power, Sigma-Delta Modulators*. Kluwer Academic Publishers, Boston, MA

Razavi B 1995 *Principles of Data Conversion System Design*. Wiley–IEEE Press, New York, NY

Razavi B, Wooley B A 1992 Design techniques for high-speed, high-resolution comparators. *IEEE J. Solid-State Circuits* **27**, 1916–26

Reay R, Kounaves S, Kovacs G 1994 An integrated CMOS potentiostat for miniaturized electroanalytical instrumentation. *IEEE Int. Conf. Solid-State Circuits,* San Francisco, CA, USA, pp. 162–3

Ryan J G, Barry L, Lyden C, Alderman J, Lane B, Schniffner L, Boldt J, Thieme H 1995 A CMOS chip-set for detecting 10 ppb concentrations of heavy metals. *Int. Solid-State Circuits Conf. Digest of Technical Papers,* San Francisco, CA, USA, pp. 158–9

Ryu M S, Shin J H, Cha G S, Hower R W, Brown R B 1994 Polymer membrane matrices for fabricating potentiometric ion sensors. *Technical Digest: 5th Int. Meet. Chemical Sensors*, Rome, Italy, Vol. 2, pp. 961–4

Sawyer D T, Sobkowiak A, Julian L., Roberts J L Jr. 1995 *Electrochemistry for Chemists*. Wiley, New York, NY

Schienle M, Paulus C, Frey A, Hofmann F, Holzapfl B, Schindler-Bauer P, Thewes R 2004 A fully electronic DNA sensor with 128 positions and in-pixel A/D conversion. *IEEE J. Solid-State Circuits* **39(12)**, 2438–45

Schmitt G, Schultze J W, Fabbender F, Bub G, Luth H, Schoning M J 1999 Passivation and corrosion of microelectrode arrays. *Electrochim. Acta* **44**, 3865–83

[Brochure]. Sensicore, Inc., Ann Arbor, MI, retrieved February 8, 2007 from http://www.sensicore.com/data/sensortestkits11-06.pdf.

Shepherd L M, Toumazou C 2005 A biochemical translinear principle with weak inversion ISFETs. *IEEE Trans. Circuits Syst. I Fundam. Theor. Appl.* **52(12)**, 2614–19

Smith RL 1982 Ion-selective field effect transistors with polysilicon gates. Ph.D. thesis, University of Utah

Strong T D 2004 Microfabricated voltammetric neuro-arrays for use *in vitro*. Ph.D. thesis, University of Michigan

Strong T D, Cantor H, Brown R B 2000 A microelectrode array for real-time neurochemical and neuroelectrical recording in vitro. *Tech. Dig. 2000 Solid-State Sensor and Actuator Workshop,* Hilton Head Island, SC, USA, pp. 29–32

Sudakov-Boreysha L, Morgenshtein A, Dinnar U, Nemirovsky Y 2004 ISFET CMOS compatible design and encapsulation challenges. *Proc. 2004 11th IEEE Int. Conf. Electronics, Circuits and Systems,* Tel-Aviv, Israel, pp. 535–8

Thomas G A, Frye-Mason G C, Bailey C, Warren M E, Fruetel J A, Wally K, Wu J, Kottenstette R J, Heller E J 1999 μChemLab – An integrated microanalytical system for chemical analysis using parallel gas and liquid phase microseparations. *Proc. SPIE* **3713**, 66–76

Tummala R R 1992 Multichip packaging – A tutorial. *Proc. IEEE* **80**, 1924–41

Wang J 1985 *Stripping Analysis: Principles, Instrumentation, and Applications*. VCH Publishers, Deerfield Beach, FL

Wei S, Lin H 1992 CMOS chopper amplifier for chemical sensor. *IEEE Trans. Instrum. Meas.* **41**, 77–80

Wolffenbuttel R F 1996 On-chip microsystems in silicon: Opportunities and limitations. *J. Micromech. Microeng.* **6**, 138–147

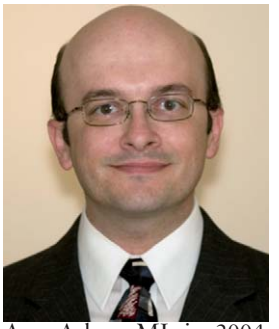
Yoon I J, Lee D K, Nam H, Cha G S, Strong T D, Brown R B 1999 Ion sensors using one-component room temperature vulcanized silicone rubber matrices. *J. Electroanal. Chem.* **464(2)**, 135–42

Yoon H J, Shin J H, Lee S D, Nam H, Cha G S, Strong T D, Brown R B 2000 Solid-state ion sensors with a liquid junction-free polymer membrane-based reference electrode for blood analysis. *Sens. Actuators B* **64**, 8–14

Biographies

Robert K. Franklin received a BS degree in Computer Engineering from Brigham Young University in 2002 and an MS degree in Electrical Engineering from the University of Michigan in 2006. During his undergraduate program, he was the recipient of the ORCA scholarship for excellence in undergraduate research at BYU and developed an automated network intrusion detection system using XILINX FPGA's. He entered the University of Michigan in September 2002 to pursue a Ph.D. in Electrical Engineering. He is currently an NSF fellow working on microelectrode probe arrays for neurochemical detection *in vivo*.

Steven M. Martin received a BS degree from the University of Florida in 1999. He received his MS and Ph.D. from the University of Michigan in 2001 and 2005, respectively. He has conducted research in the areas of high-performance interface circuits for chemical sensors, ultrasound devices, and laser detection systems. He currently works as an analog engineer for Intel Corporation designing multigigahertz I/O circuits. Dr. Martin holds two US patents and has several others pending pertaining to his work on sensors and sensor interfaces.



Timothy D. Strong received a BS degree in Electrical Engineering from Michigan Technological University, Houghton, MI, in 1992, a MS degree in 1997, and a Ph.D. in Electrical Engineering from the University of Michigan, Ann Arbor, MI, in 2004. He is currently employed in research and development at Sensicore, Inc., Ann Arbor, MI, a company developing water quality monitoring systems. His research interests include amperometric and potentiometric chemical sensing, and silicon processing and fabrication.

Richard B. Brown received BS and MS degrees in Electrical Engineering from Brigham Young University, Provo, UT, in 1976 and a Ph.D. degree in Electrical Engineering (solid-state sensors) from the University of Utah, Salt Lake City, UT, in 1985.

From 1976 to 1981, he was VP of Engineering at Holman Industries, Oakdale, CA, and Manager of Computer Development at Cardinal Industries, Webb City, MO. In 1985, he joined the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI, where he served as associate chair and interim chair. In June 2004, he was appointed 11th Dean of the College of Engineering, University of Utah. He is Professor in the School of Computing and Department of Electrical and Computer Engineering. He has conducted major research projects in the areas of solid-state sensors, mixed-signal circuits, GaAs and silicon-on-insulator circuits, and high-performance and low-power microprocessors.

Professor Brown is Chairman of the NSF MOSIS Advisory Council for Education. He was Chair of the 1997 Conference on Advanced Research in VLSI and the 2001 Microelectronic System Education Conference. He has served as Guest Editor of the *IEEE Journal of Solid-State Circuits* and *Proceedings of the IEEE* and as Associate Editor of *IEEE Transactions on VLSI Systems*.

2.13 Microfluidic for Lab-on-a-Chip

Stefan Haeberle[1,2] **and Roland Zengerle**[1,2], [1] Institute for Micromachining and Information Technology of the Hahn-Schickard-Society (HSG-IMIT), Villingen-Schwenningen, Germany, [2] Department of Microsystems Engineering-IMTEK, University of Freiburg, Freiburg, Germany

Glossary

CMOS Complementary Metal Oxide Semiconductor
DWP Dispensing Well Plate
EDC/NHS Carbodiimid/*N*-hydroxysuccinimide
EOF Electroosmotic Flow
EP Electrophoresis
EWOD Electrowetting
FID Free Interface Diffusion
GC Gas Chromatograph
HPLC High Pressure Liquid Chromatograph
HTS High Throughput Screening
IDT Interdigital Transducer
LSI (microfluidic) Large Scale Integration
MSL Multilayer Soft Lithography
PCR Polymerase Chain Reaction
PDMS Polydimethylsiloxane
SAW Surface Acoustic Wave
TAS Total Chemical-analysis System
TIR Total Internal Reflection
μTAS micro Total Analysis System

2.13.1 Introduction

2.13.1.1 Microfluidics

The history of microfluidics dates back to the early 1950s when efforts to dispense small amounts of liquids in the nano- and subnanoliter range were made for providing the basics of today's ink-jet technology (Le 1998). In terms of fluid propulsion within microchannels of submillimeters cross-section, the year 1979 set a milestone when a miniaturized gas chromatograph (GC) was realized on a silicon wafer (Terry *et al.* 1979). The first high-pressure liquid chromatography (HPLC) column device, fabricated using Si-Pyrex technology, was published by Manz *et al.* (1990b). By the end of the 1980s the first microvalves (Shoji *et al.* 1988) and micropumps (Van Lintel *et al.* 1988) based on silicon micromachining were also presented. All these examples represent microfluidic systems since they enable the precise control of the decreasing fluid volumes on the one hand and the miniaturization of the size of a fluid handling system on the other hand.

Another important aspect of microfluidics is the exploitation of effects and phenomena that can be utilized only in microdimensions. Smaller channel dimensions drastically increase the surface-to-volume ratio and thus surface-related phenomena like laminar flow, capillarity, fast thermal response, and electrokinetics gain influence. This can be used in microfluidic systems to enhance the performance of analytical procedures. Manz *et al.* (1990a) also proposed the concept of miniaturized total chemical analysis systems (TAS) based on the unique conditions in the microdomain. Today this approach is also known as micro total analysis systems (μTAS) or laboratories on a chip (lab-on-a-chip) as proposed by Harrison *et al.* (1992).

Following this μTAS or lab-on-a-chip approach, the first applications that emerged in the field of analytical chemistry were based on the electro-osmotic flow (EOF) to pump liquids into small microcapillaries and on electrophoretic separation (EP) to distinguish sample components (Effenhauser *et al.* 1993, Harrison *et al.* 1992, 1993, Manz *et al.* 1992). These developments in the early 1990s drastically increased the academic and commercial interest in microfluidic technologies. This trend continues to the present day, as described in a recent comment on the proliferation of microfluidics in literature and intellectual property, which claims that 581 of 770 microelectromechanical systems (MEMS)-related papers published in 2003 dealt with microfluidics (Kamholz 2004).

So down to the present day manifold lab-on-a-chip systems have been developed for diverse applications, e.g., for DNA analysis (Burns *et al.* 1998),

DNA amplification (polymerase chain reaction, PCR) (deMello 2003, Kopp *et al.* 1998), proteomics (Lion *et al.* 2003, Marko-Varga *et al.* 2003), sample pretreatment (de Mello and Beard 2003). Recent general reviews on the whole field of lab-on-a-chip systems can be found in Auroux *et al.* (2002), Reyes *et al.* (2002), and Vilkner *et al.* (2004). The enormous impact of microfluidic lab-on-a-chip technologies also becomes obvious by the large number of recently published books. They either cover the whole field of microfluidics (Li 2006, Nguyen and Wereley 2002) or focus on the engineering (Geschke *et al.* 2004) or applications of microfluidic systems (Andersson and van den Berg 2004, Oosterbroek and van den Berg 2003, Tay 2002, Urban 2006). Also journals exclusively dedicated to the field of micro- and nanofluidics (*Microfluidics and Nanofluidics*, Springer) or lab-on-a-chip systems (*Lab on a Chip*, Royal Society of Chemistry) have been published recently.

2.13.1.2 Microfluidics – An Enabling Technology

From a more business-related point of view, it can be stated that nowadays most MEMS-related technology roadmaps and market studies point out the significant technological and scientific impact that microfluidics will have on various industries, especially the life sciences. A dedicated microfluidics roadmap was prepared by Ducrée and Zengerle (2004). In this study the economic development related to microfluidics technologies for the life sciences has been estimated and important market drivers and road blocks have been pinpointed.

Regarding the economic impact of microfluidics, the study anticipates an overall growth rate of more than 30% per annum for microfluidic technologies and products in the life sciences. Drug discovery, medical diagnostics, and therapeutic devices represent the most promising fields. The overall global market of microfluidics in the life sciences has been estimated to be worth approximately 500 million euros in 2002, increasing with an assumed annual growth rate of 19% to 1.4 billion euros in 2008.

Besides the economic impact of microfluidics, technological trends have also been clearly identified in the roadmap. Amongst those the most relevant is the need for microfluidic platforms equipped with a basic set of validated fluidic base operations to arrive rapidly at application-specific microfluidic systems. This system-oriented platform concept contrasts the frequent approaches in which components such as pumps or valves have been optimized at an individual level and assembled afterward.

2.13.1.3 The Need for Microfluidic Platforms

Why are microfluidic platforms needed? As described in Section 2.13.1.1 the impact of microfluidic technologies in the academic world has dramatically increased during the last few years. This is quite amazing since microfluidics is no independent product consumers want to buy. Microfluidics should be merely considered as a toolbox, which is needed to develop innovative new products in such disparate fields as medical, pharmaceutical, and analytical applications. As a consequence, the most important customer for microfluidic know-how and technologies is the research community itself, which develops new products and solutions in the different application areas such as the biotechnology, diagnostics, medical, or pharmaceutical industry.

During the last two decades, thousands of researchers have spent a lot of time developing new microfluidic components or exploring the basic microfluidic operations such as fluid transport, fluid metering, fluid mixing, valving or concentration, and separation of molecules within miniaturized quantities of fluids. Today hundreds of different types of micropumps have been fabricated (Laser and Santiago 2004), hundreds of different types of mixers and hundreds of different types of microvalves are known, and almost no standards are defined in terms of interconnections. It seems to be the right time to raise the question of whether we really need more of those components? On the basis of our experience in the lab-on-a-chip field, a component-based microfluidic approach is much too slow and the R&D effort is much too expensive to explore the huge potential of different applications. In addition, the best performance you can get out of such a component-oriented solution will be far behind what you can get in an integrated system approach, or in other words in a microfluidic platform approach. Therefore we think that the described practice of assembling discrete components such as valves and pumps, at least in the field of lab-on-a-chip applications, belongs to the past and we do not expect it to continue in the future. In our view the research community really needs validated and easy-to-operate microfluidic platforms. They have to offer an adequate number of microfluidic operations that can be easily

combined to build application-specific microfluidic systems. In addition these systems should be fabricated using a standardized cost-efficient technology.

2.13.1.4 Discrete Solutions for Unique Technical Challenges

Before pointing out the advantages of the microfluidic platform concept, we describe the opposite: two examples of application-specific integrated systems, which represent unique engineering solutions for unique technical problems. The electronic fountain pen (Waibel *et al.* 2003) as depicted in **Figure 1** is a good example of such a discrete microfluidic solution. It can be regarded as the first fully functional, highly integrated, miniaturized, and self-sustaining microdosage system of its kind operating under real-world conditions. The main components are a liquid level sensor, a microvalve, and a bubble- and particle-tolerant fluidic system. The pen is optimized with respect to minimum energy consumption. It contains a programmable ASIC and is powered by two standard watch batteries ensuring operation over a period of 2 years under standard conditions.

The oral drug delivery system (**Figure 2**), currently under development at HSG-IMIT, is a similar example (Goettsche and Wolff 2006). It has the size of two buccal teeth and will be integrated into the human denture. It consists of a drug reservoir capable of incorporating a solid pill, an osmotic pump, and a flow sensor. The refillable system is loaded by a solid pill and is designed to deliver liquid drug over a period of 2 weeks at rather harsh ambient conditions

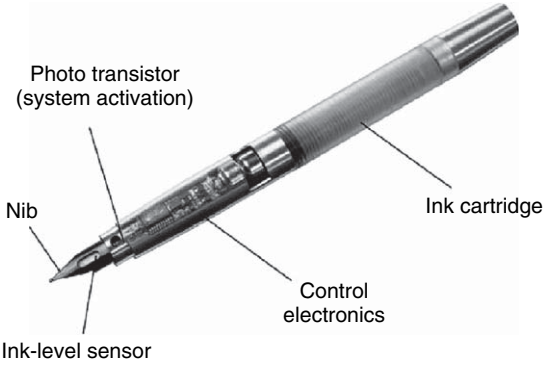


Figure 1 Photo of the electronic fountain pen (145 mm × 12 mm). (Reprinted with permission from Waibel G, Kohnle J, Cernosa R, Storz M, Schmitt M, Ernst H, Sandmaier H, Zengerle R, Strobelt T 2003 Highly integrated autonomous microdosage system. *Sens. Actuators A Phys.* **103**, 225–30. Copyright 2003 Elsevier.)

Figure 2 Photo of a drug delivery system that can be implanted into a human denture. (Source: Goettsche T, Wolff, A 2006 IntelliDrug – An integrated intelligent oral drug delivery system. *mst-news* **06**, 36–7.)

(mechanical loads, wide temperature range, and contact with any kind of food) inside the human mouth.

Both microfluidic systems perfectly fulfill the requirements for their specific applications. However, for any other application in the field of microdosage or more general in the field of microfluidics, the specific know-how from developing such a system is only of very limited value and every development of this kind always starts from the scratch again. This causes significant costs and time at a high economic risk. Although we expect this kind of development to make sense for a few selected applications in diverse fields of applications in the future also, it is quite clear that this approach will not succeed for lab-on-a-chip systems or the diagnostic applications that are dealt with in this chapter.

2.13.1.5 What Is a Microfluidic Platform?

Very similar to the ASIC industry in microelectronics, which provides validated elements and processes to make electronic circuits, a dedicated microfluidic platform comprises a reduced set of validated microfluidic elements. These elements are capable of performing the basic fluidic unit operations required in a given application area. Such basic fluidic unit operations are, for example, fluid transport, fluid metering, fluid mixing, valving, and separation or concentration of molecules or particles. The collection of fluidic unit operations needed for diagnostic applications can have only little overlap with the collection needed for pharmaceutical applications or for applications in microreaction technology. In some cases detection methods also belong to the basic set of microfluidic

Table 1 Common features of microfluidic platforms

Microfluidic operations	*Fabrication technology*
Validated elements for basic microfluidic unit operations such as • Fluid transport • Fluid metering • Fluid valving • Fluid mixing • Separation • Concentration • Detection • ...	Validated manufacturing technology for the whole set of fluidic elements (prototyping and mass fabrication) Seamless integration of different elements • Ideally in a monolithic way • Or by a well-defined easy packaging technique

operations, and in other cases this is not the case (**Table 1**). Nevertheless, in all cases the user of a platform should be capable of readily combining the elements within a given platform in order to implement an assay for diagnostic applications or to screen for new compounds in pharmaceutical applications. Often, an efficient development is intimately linked to the availability of (standard) test setups and simulation tools.

More important than providing a totally complete set of fluidic unit operations in a platform is the fact that all elements have to be amenable to a well-established fabrication technology. Furthermore all elements or modules of a platform have to be connectible, ideally in a monolithically integrated way or at least by a well-defined, ready-to-use interconnection and packaging process. If a platform allows a seamless and simple integration of different fluidic elements in a monolithic way, for example, without sophisticated additional packaging techniques, this provides a significant advantage compared to other platforms. Thus speaking about microfluidic platforms also involves at least one validated fabrication technology to realize complete systems out of the elements. This results in a definition of a platform as follows.

A microfluidic platform allows to perform a set of fluidic unit operations that are enabled by a set of fluidic elements, which are designed for easy combination with a well-defined (and low-cost) fabrication technology. The platform allows to implement and fabricate different application-specific solutions in an easy and flexible way.

2.13.1.6 Overview of the Following Chapters

When comparing microfluidic platforms with the microelectronics industry, the question of whether we really need a diverse set of different microfluidic platforms arises. The coexistence of different platforms and their specific fabrication technologies could be regarded as a drawback since R&D efforts go in many directions and cannot be focused exclusively on one technology (as is complementary metal oxide semiconductor (CMOS) in microelectronics). On the other hand, the diversity of approaches and technologies can be considered as an advantage for their successful adoption in different application fields. Owing to the diversity, specific advantages of certain platforms are likely to succeed in different application areas. It is, however, essential that a specific platform provides all of the characteristic features given in **Table 1**.

This chapter is intended to give an overview of microfluidic platforms that have been developed until today. We will thereby focus only on platforms for lab-on-a-chip application, being aware that there are also other possible fields of applications for microfluidic platforms such as microprocess engineering or microdosage systems. However, in the field of lab-on-a-chip systems also, we cannot cover all microfluidic platforms, which are known from literature. Prominent examples of platforms that will not be discussed here are certainly the electrokinetic platform (EOF, EP), the microarray platform, and the microwell plate technology platform. These platforms, however, are already well described in many scientific papers and books. It is, furthermore, not intended to assess the different platforms by their value to the industry or to the research community.

As is evident in the introduction, the scope of this chapter is not to describe single microfluidic components. For detailed reviews on single components such as micropumps (Gravesen *et al.* 1993, Laser and Santiago 2004, Shoji and Esashi 1994, Woias 2005), valves (Oh and Ahn 2006), mixers (Hessel *et al.* 2005, Nguyen and Wu 2005), microfluidic technologies in general (Squires and Quake 2005, Stone *et al.* 2004), or

simulation techniques for the design of lab-on-a-chip systems (Erickson 2005) we also refer to the published work.

In the following, each microfluidic platform is described in the same way. First, we motivate the platform approach and give a short introduction on the basic setup and functional principle. Then, the main unit operations realized on the platform and an application example based on their selection are presented. Finally, each section closes with a discussion on the strength and the challenges of the platform as we can see them today.

2.13.2 Capillary-driven Test Strips (Lateral Flow Assays)

2.13.2.1 Introduction

The capillary-driven test strip platform is the state of the art in point-of-care diagnostics, with billions of units that are produced in an extremely cheap manner. It is amazing that within the lab-on-a-chip or microfluidics community, not very much is known about this easiest way to perform assays on such an easy to handle widely used platform (e.g., diabetes testing, pregnancy testing).

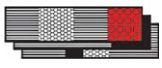
The first test strip is certainly the indicator paper for pH measurement, which is also a representative for the most simple assay on the capillary-driven test strip platform. It consists only of one single fleece with an integrated colorimetric reagent (**Table 2**). The sample liquid is transported into the fleece of the test strip by capillary forces and a color change, depending on the pH value of the sample liquid, occurs. This color change is initiated by a reaction between the sample and the reactant. Starting from that most simple point of assay format, more complex configurations with multiple reagents or even several fleeces enable the implementation of more complex assays like immunoassays.

The so-called lateral flow assays are well known in the diagnostic field since the 1960s. Although this can be regarded as the most successful microfluidic platform for lab-on-a-chip applications in terms of applications and commercialized products, hardly any publication from a microfluidic point of view exists.

The possibility of performing an automated on-site measurement, using a cheap and small disposable test strip, combined with the simple actuation principle that does not need any energy supply, gives the platform a huge potential for point-of-care and patient self-testing applications. Today, millions of diabetics all over the world use these diagnostic devices to measure their glucose concentration several times per day and therewith adjust their medication.

The basic principle of the platform is the passive liquid transport via capillary forces within the capillaries of a fleece or microstructure layer. The physical background of these actuation principle is described in Section 2.13.2.1.1.

Table 2 Degrees of complexity and application examples on the capillary-driven test strip platform

	Capillary configuration		*Application examples*
Complexity ↓	Single fleece with colorimetric reactant		Indicator paper (pH measurement)
	Single fleece with multiple reactants and colorization		BSE test, TNT test
	Multiple fleeces with different zones of reactants, often cased		Pregnancy test, cardiac markers, drug test
	MEMS-based capillary channels, multiple complex reactions		Detection of up to 100 different proteins (immunoassays)

2.13.2.1.1 About capillary forces

One effect that becomes more and more important when scaling down fluidic channels is the capillary force. A capillary pressure difference

$$\Delta p_\theta = \frac{2\sigma}{r}\cos\theta \qquad [1]$$

appears across the liquid–gas interface with surface tension σ in a capillary of radius r and contact angle θ. For hydrophilic contact angles <90° (partially wetting) the pressure in the liquid phase exceeds the pressure in the gas phase, leading to a further wetting of the capillary. For hydrophobic contact angles >90° (partially nonwetting) the meniscus withdraws respectively.

An important effect when dealing with capillary-driven liquid flows is contact line pinning. It causes the sudden stop of the proceeding meniscus at edges that represent a geometrical singularity. This has to be taken into account when designing microfluidic structures on the capillary test strip platform since no additional pressure is available to overcome this stop. On the other hand, however, this mechanism can be used to control the course of capillary priming on the platform.

2.13.2.2 Unit Operations Controlled by Capillary Forces

2.13.2.2.1 Sample loading

Several methods for sample loading exist. For bigger volumes, the sample is filled into a start reservoir from where it penetrates the underlying capillaries. If the sample volumes are not sufficient, the addition of a dilution buffer is sometimes required to enhance or allow the capillary transport in the first place. The second method, especially used in patient self-testing applications, is the direct capillary filling of the strip from the sampling point. For blood diagnostic assays, for example, the test strip is directly contacted with the blood spilled out of the finger tip that has been locally pricked with a lancet before.

2.13.2.2.2 Separation

Separation steps are required for analyzing human whole blood, for example, since the red blood cells with a volume fraction >40% would interfere with the assay. As the blood passes through a separation fleece, cells are filtered from the blood (Clark *et al.* 2002). Also microfilters or membranes with a more defined pore size can be used. An exemplary immunoassay test strip is depicted in **Figure 3**. The separation fleece is placed directly underneath the start reservoir into which the blood sample is added.

2.13.2.2.3 Transport

Depending on the complexity of the test strip, liquid propulsion is accomplished via capillary forces in fleeces and/or highly precise micromachined structures (**Table 2**). Though the system gains control when using microstructures, its fabrication becomes more expensive on the other hand. Some of the essential features of both types of capillary pumping structures are given in **Table 3**. Due to these considerations, the requirements of the assay decide which propulsion method should be used and certainly the combination of these two can increase the controllability of the assay protocol while still keeping the costs low.

An important issue for liquid transport with merely capillary-driven systems is the transport balance. Since the maximum volume of labeled sample should pass the detection zone in order to have an optimum sensitivity, certain arrangements have to be made. The volume of the transport zones, for example, should be kept small, so that only a small fraction

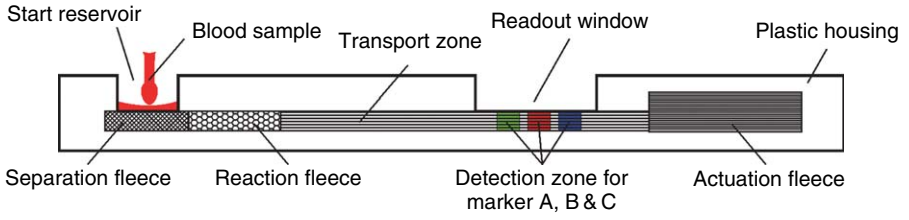


Figure 3 Simplified cross-section of a typical capillary-driven test strip set-up for an immunoassay. The blood sample is applied at the start reservoir and passes through the separation fleece where the red cells are separated from the plasma. The plasma enters a reaction fleece containing labeling reagents comprising fluorescent antibodies that bind to the proteins in the blood plasma. The sample is then transported through the detection zone with immobilized complementary antibodies to the markers under investigation (A, B, and C). The actuation fleece at the end of the strip ensures that the whole liquid volume is sucked through by the capillary liquid propulsion through all fleeces. All fleeces are typically packed into a low-cost plastic housing.

Table 3 Comparison of fleeces and microstructures for capillary propulsion in test strips

Properties	*Fleeces*	*Microstructures*
Cost	Cheap	Expensive
Availability	Good (various types)	Bad (must be created)
Realization of filtering	Easy	Elaborate
Flow guiding	3D structures	2D structures
Control of flow direction	Bad (isotropic, toward unsaturated fleece)	Good (follows structure)
Flow prediction	Not exact	Good

of the initial sample volume remains within that dead volume. Another measure for a good volume balance is to ensure that the capillarity of the input zone (separation and labeling fleece) is lower than the capillarity of the actuation fleece, leading to a complete drainage of the sample into the actuation fleece before the liquid propulsion terminates.

A very fundamental requirement for a successful assay is that the initially added sample volume exceeds a critical minimum. Otherwise, the complete wetting of all the essential test strip zones cannot be assured and the assay fails.

2.13.2.2.4 Incubation

During an assay protocol on the capillary-driven test strip platform, the sample has to get in contact and react with several predeposited reagents. They are integrated into the fleece or microstructure of the strip during fabrication. This is mostly done by fleeces that are saturated with the reagent and dried afterward. If needed, the reactant molecules can also be covalently bound to the surface using standard immobilizing techniques like the carbodiimid (EDC)/*N*-hydroxysuccinimide (NHS) chemistry.

The setting of an appropriate incubation time for reagent dissolution is managed by different zones within the test strip, exhibiting different wetting properties. The basic principle behind this approach is depicted in Figure 4. First, the liquid primes the reaction chamber, where a dry reagent is predeposited in a pillar microstructure. This is achieved very fast due to a low contact angle of 10° only. The propagation of the liquid meniscus is then slowed down within a so-called time gate with an increased contact angle of $\theta = 80°$ and consequently a reduced capillary force. The time for the dissolution of the dry reagent is set by the length of the time gate and ends as soon as the liquid reaches the detection zone with an increased capillary pumping force ($\theta = 10°$) speeding up the flow again.

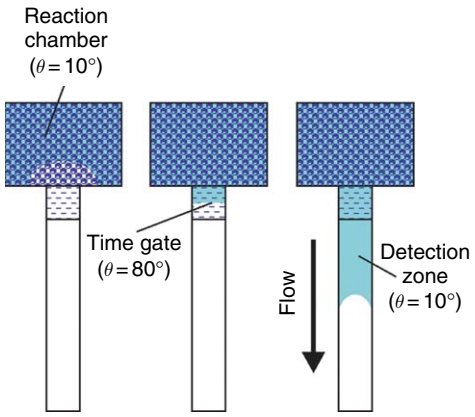


Figure 4 Schematic realization of a certain incubation time in the capillary-driven test strip platform. The liquid flow is throttled in the time gate of reduced wettability ($\theta = 80°$) thus leading to an extended period of time for reaction within the reaction chamber. Consequently, the dried reagents can be dissolved completely before the liquid proceeds along the detection zone.

2.13.2.2.5 Mixing

The dissolution of the preloaded dried reagent situated in the reaction fleece, described as an incubation step before, can also be regarded as a kind of mixing operation. A characteristic feature of the capillary test strip platform is the fact that mixing generally is purely passive and is based only on diffusion. It cannot be accelerated externally and is also one of the major drawbacks of this kind of platform. In order to keep the mixing time short, a high surface-to-volume ratio within the reaction chamber is preferred. However, the dissolution time typically still lies in the range of several minutes and thus the capillary flow through the reaction chamber has to be throttled as described before.

2.13.2.2.6 Metering

The metering of liquids is an important unit operation for quantitative assays. Within a test strip, metering is done by the defined volumes of the fleeces and

microstructures. The liquid flow stops automatically as soon as the actuation fleece (**Figure 3**) is fully wetted with liquid. This way the amount of liquid that has passed the detection zone is well defined. The only thing that has to be ensured is that the start reservoir is filled with enough liquid at the beginning of the assay, i.e., with the volume of the complete test strip (all fleeces and microstructures) and a certain reliability excess to ensure proper functioning.

2.13.2.2.7 Amplification

The results from a test strip assay are mostly read out by optical markers such as fluorescent molecules. Since the concentration of these molecules within the sample liquid is potentially small, they have to be accumulated within the detection zone. The sample volume passes through the detection zone with an adequate flow rate, ensuring the non-diffusion-limited binding of the marked sample molecules to the immobilized capture molecules in the detection zone. A remarkable fluorescent signal is obtained after a multiple of the detection zone volume has passed the immobilized molecules. This can be regarded as a microfluidic amplification unit operation.

2.13.2.2.8 Readout

The fluorescent emission of the detection zone is excited using a laser diode of corresponding wavelength. The phase-shifted (longer wavelength) fluorescent response is detected using a photodiode and an appropriate optical filter within common test strip readers. If regions of different capture molecules are immobilized within the detection zone, the test strip is moved with a stepper motor underneath the detection unit (laser diode and photodiode).

Some assays are also read out using electrochemical mechanisms. The glucose concentration of a blood sample is determined by measuring the electrical charge generated during the enzymatic oxidation of glucose to gluconic acid, for example. The test strip reader applies an external electric potential and measures the current, which is a function of the number of electrons generated.

Besides the above-described device-based readout methods, the reading of assay results with the naked eye is also possible. This is of interest for those applications where a cheap and fast readout is required. A manual readable signal is produced by binding small gold or latex particles to the detection molecule, which accumulate at the detection zone and color it. However, only clear and binary signal-generating assays such as pregnancy tests are capable of manual readout.

2.13.2.3 Application Example: The Cardiac® Reader

Cardiac markers are very important in emergency medicine. They allow a fast decision if a patient with chest pain is suffering from an acute cardiac infarction and has to be treated accordingly. Only the combination of several markers enables a clear diagnostic conclusion on whether the patient had a heart attack or not. Since this diagnostic information and the need for adequate medication should be gained as fast as possible, a lab-on-a-chip system for the point-of-care measurement is demanded.

The Cardiac Reader® from Roche (2006) is a commercial point-of-care system for the detection of four different cardiac markers (myoglobin, troponin T, D-dimer, and NT-probBNP). The test strip reader and the corresponding test strips are depicted in **Figure 5** and allow the determination of these four markers within minutes from a single blood sample. The system is used in doctor's offices or emergency rooms.

The complete test strip consists of different fleeces for liquid transport, blood separation, and readout. First, 150 µl of heparinized venous whole blood is added onto the inlet region above two

Figure 5 The Cardiac® Reader and the four different test strips (Source: Roche 2006 Basel, CH, www.roche.com, accessed 2006.)

reagents containing fleeces. Within these fleeces, two antibodies specifically bind to the antigen (e.g., troponin T protein) of the blood sample. One antibody is labeled with biotin, the other with a gold nanoparticle. After the binding step, the blood sample is transported through a separation fleece, where the cellular constituents are removed.

In the last assay step, the antigen–antibody complex is transported along a detection zone, featuring immobilized strepdavidin proteins, which bind to the biotin of the complex. Thus the protein complex is captured in the detection region, and the assay result is read out optically via a dark line generated by the accumulating gold nanoparticles. To control the assay performance, antigens are immobilized in a control region subsequently passed by the sample liquid. A successful assay run with a negative result, that is, no Troponin T within the blood sample, can be determined by a dark line arising at this control position only (Raschke 2005). The complete assay, starting from the application of the blood sample until the readout of the result, takes 8–12 min.

2.13.2.4 Strengths and Challenges of the Platform

The major strength of the capillary-driven test strip platform is certainly the robust microfluidic flow propulsion principle, relying on capillary forces only. No external energy is required, which opens up a wide field of applications especially for simple color-changing assays such as pH measurement or pregnancy tests. However, complex immunoassay protocols have also been implemented during the last few years. Therefore this special microfluidic platform is setting a benchmark in terms of costs and integrated, automated assay implementation for all microfluidic platforms discussed in this chapter.

Drawbacks of the platform certainly arise from its simplicity. Assay protocols in the capillary-driven systems follow a fixed process scheme, imprinted in the microfluidic channel design. Passive liquid propulsion by capillary forces only cannot be influenced actively once the process is started. As a consequence the exact timing of assay steps depends on variations of viscosity and surface tension of the sample. Therefore the precision of the assay result, for example, is on the order of 10%, which is not always sufficient for several future challenges in the implementation of diagnostic assays. More complex diagnostic assays also cause a larger number of process steps such as reactant dissolution. As previously described, this mixing operation cannot be accelerated in the capillary test strip platform, which leads to long assay times.

Crucial unit operations for the assay precision are metering and incubation steps whose accuracy is limited to the merely capillary-driven system. A further critical point is the long-term stability of the wetting properties inside the fleeces or the microstructures. Usually, the materials are plasma treated or coated by an additional layer to ensure the desired contact angle and thus wetting behaviors. These surface activations or coatings have to be stable at different temperatures and over a long period of time as they define the test strip lifetime.

2.13.3 Pressure-driven Systems

2.13.3.1 One Step Forward in Controllability and Complexity

Assay protocols in capillary-driven systems follow a fixed process scheme, imprinted in the microfluidic channel design. Passive liquid propulsion by capillary forces only cannot be influenced actively once the process is started. In addition the exact timing depends on variations of viscosity and surface tension of the sample, and mixing is rather slow. This certainly offers other platform technologies the chance to show up and get into the market. As an example, pressure-driven platforms enable the control of the running processes. Liquid flows can be stopped, reversed, and their flow rates and ratios can be adjusted online in order to change the course of the assay. A platform-integrated pressure source is required to make this step forward in controllability possible. So the gain of control is accompanied by an increased complexity of the system.

2.13.3.2 Unit Operations in Pressure-driven Microfluidics

2.13.3.2.1 Fluid transport (Laminar flow)

Independent from the liquid propulsion method, the pressure-driven liquid flow within microchannels has some common properties. Strictly laminar flow due to the small Reynolds numbers within the channels of small cross section in combination with the no-slip boundary condition at the channel walls lead to a parabolic velocity profile. Thus, the flow velocity depends on the position within the cross section of

the channel, i.e., long residence time near the walls and short residence time in the center of the channel.

The laminar volumetric flow rate I_v through microchannels

$$I_v = \frac{\pi}{8\eta}\frac{r^4}{l}\Delta p \qquad [2]$$

depends on the applied pressure difference Δp, length l, and radius r of the channel as well as the liquid viscosity η (law of Hagen–Poiseuille).

Different possibilities for pressure generation are described below and are also categorized as shown in **Figure 6**. A constant pressure source can be activated only once and operates autonomously afterward. This is similar to capillary systems, whereas the potential pressures and thus flow rates can be considerably higher compared with merely capillary-driven flows. In addition, liquid propulsion sustains after the whole structure is filled with liquid since no capillary pressure-generating air–liquid interface is required.

The easiest way to accomplish a pressure-driven flow by a constant pressure source is certainly using an external pressure cartridge, connected to the microfluidic chip. However, the macro- to microfluidic interfacing problem has to be solved for these off-chip solutions, which is a common problem when dealing with microfluidics (Fredrickson and Fan 2004). This fluidic interface has to be pressure tight and should not lead to a lot of additional dead volume.

Constant pressure can also be generated on the chip by integrating, e.g., a spring-loaded liquid reservoir. Again, the interfacing problem remains if the reservoir is loaded while already integrated in the chip. More simple techniques for pressure generation on the chip use, for example, thermal phase transition effects and can be switched on via a heater.

If the pressure can be switched on and off, respectively, and if its amplitude can be adjusted within a certain working range, the system gains in controllability. The actuation mechanisms used for pressure generation can be basically divided into hydraulic (liquid) and pneumatic (gas) principles. Liquid propulsion is accomplished by either applying an overpressure at the inlet (pushing of fluid) or applying an underpressure at the outlet (sucking of fluid) of a microchannel structure. Depending on the microfluidic structures as well as the application, both methods can be advantageous in terms of bubble-free priming and leak tightness.

All off-chip macropump solutions require a macro-to-micro interfacing technology, as described in the previous section, leading to a certain dead volume required for interconnection tubing. Although, the dead volume can be decreased by placing a micropump onto the microfluidic chip, a micro-to-micro interface is still needed for interconnection. This is not consistent with the ideal microfluidic platform that allows to realize all fluid unit operations in a monolithically integrated manner using the same fabrication technology. Only this way expensive and error-prone liquid interfaces can be avoided. A large number of micropumps have been presented during the last two decades (Laser and Santiago 2004, Shoji and Esashi 1994, Woias 2005). Many of these pumps are based on the combination of a displacement chamber with two active or passive microvalves. These devices are discussed in more detail in Chapter 2.09 and are not considered here.

More simple and robust pressure-generating principles, like the displacement of a membrane via a finger push or a manually operated syringe, have also been demonstrated as interesting alternatives.

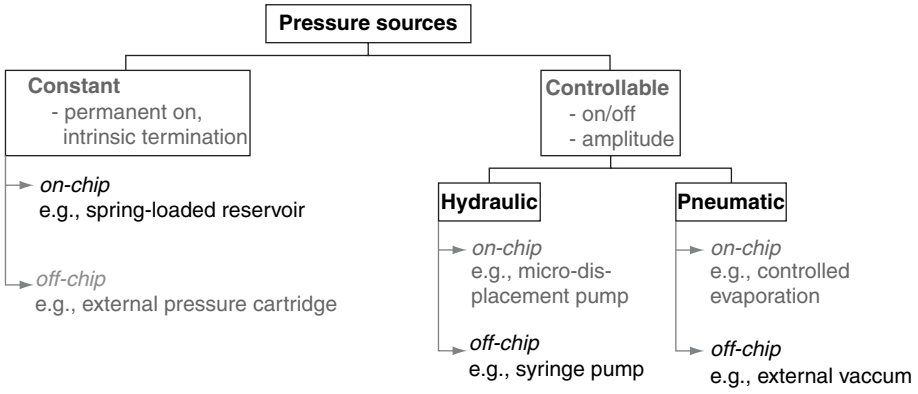


Figure 6 Different pressure sources for microfluidic platforms.

These methods are used in point-of-care testing applications, where a simple and low-energy consuming pumping principle is required. In the following, three examples of novel on-chip pumping mechanisms based on thermal effects are described.

Two nonmechanical implementations of micropumps that allow monolithic integration and that are based on the phase change of a working fluid are depicted in **Figure 7**. Heating elements, integrated into the microfluidic channels, increase the temperature of the fluid locally when actuated. This leads to a phase change of the liquid volume at the position of the heater (evaporation) if the temperature exceeds the boiling point. A virtual movement of the bubble by actuating several discrete and serial aligned heaters subsequently has been proposed by Song and Zhao (2001) (see **Figure 7**, *left*). Using this actuation principle, deionized water has been pumped through a 37.2 cm-long Pyrex glass tube (1.0 mm inner diameter) using 12 serial-aligned heating elements (heating power, 8–12 W). A maximum pressure head of 57 mm H_2O and a maximum volumetric flow rate of 300 µl min^{-1} are reported. Another pumping principle, also based on the thermal phase change is depicted in **Figure 7** (*right*). This device consists of two triangular-shaped heat sources only (Yokoyama *et al.* 2004). The heat flux inside the heater has its maximum at the apex of the triangle, defining the seed position of the bubble during evaporation. A flow rate of 12.5 mm s^{-1} at a driving frequency of 1 Hz has been achieved within a microchannel of 220 mm length and 600 µm × 120 µm cross section using two of the depicted micropumps (overall: four heaters).

Another pumping method that is also based on thermal actuation is depicted in **Figure 8** (Zimmermann *et al.* 2005). The device consists of three independent microchannels, arranged perpendicular to three linear Peltier elements. The channels have an open filling zone, a sealed microchannel in which the assay can be performed, and a likewise open pumping zone. The device is initially primed with liquids by capillary forces via the filling zone. Afterward, the Peltier-1 cools the filling zone to prevent evaporation, i.e., loss of sample or reagent. The ability to keep 90% of 0.6 µL solution in an open filling zone for 60 min is reported.

For liquid propulsion, the Peltier-3 additionally heats the pumping zone leading to a controlled evaporation of liquid at the open interface. The evaporation of liquid sets the flow rate in the microchannel. The functional principle can be considered as a capillary pump maintaining the capillary pressure on both sides of the channel. Flow rates ranging from $\sim 1.2\,nl\,s^{-1}$ to $\sim 30\,pl\,s^{-1}$ have been demonstrated using this principle. The flow rate perfectly matches with the binding kinetics of the molecules inside the liquid to the channel surface, which is diffusion limited, and does exclude faster flow rates.

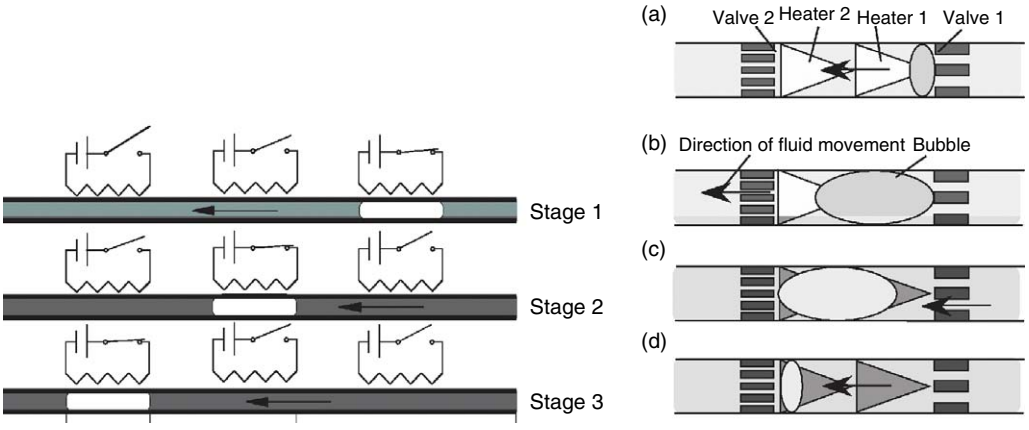


Figure 7 *Left*: Illustration of the phase-change pumping mechanism proposed by Song and Zhao (2001). Three serial-aligned and sequentially actuated heating elements initiate a moving vapor bubble and thus a directed liquid propulsion (stages 1–3). (Source: Song Y J, Zhao T S 2001 Modelling and test of a thermally-driven phase-change nonmechanical micropump. *J. Micromech. Microeng.* **11**, 713–19, IOP Publishing Limited.) *Right*: Basic principle of the thermal micropump described in (a) start of heating (heater 1 on, heater 2 off); (b) bubble growing (heaters 1 and 2 on); (c) start of condensing (heater 1 off, heater 2 on); (d) bubble collapsing (heaters 1 and 2 off). (Reprinted with permission from Yokoyama Y, Takeda M, Umemoto T, Ogushi T 2004 Thermal micro pumps for a loop-type micro channel. *Sens. Actuators A Phys.* **111**, 123–8. Copyright 2004 Elsevier.)

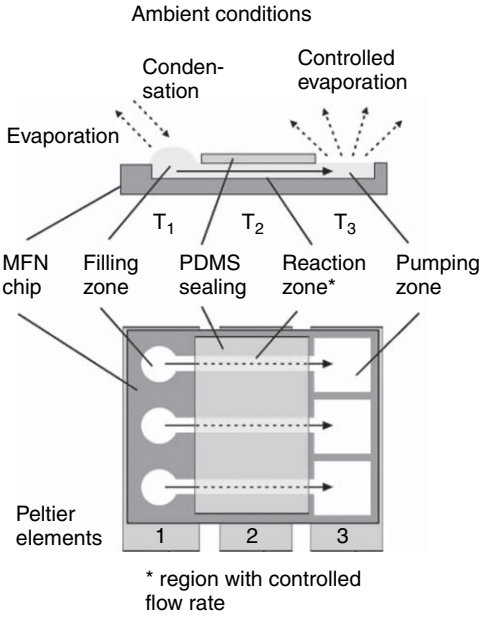


Figure 8 Controlled evaporation method for liquid propulsion. Three microchannels are aligned perpendicular to a set of three Peltier elements allowing the temperature control at the filling, reaction, and pumping zone. Cooling of the filling zone and heating of the pumping zone initiate a controlled evaporation of liquid at the pumping zone. Thus, a liquid flow through the channel is initiated, while preventing evaporation at the inlet zone. (Source: Zimmermann M, Bentley S, Schmid H, Hunziker P, Delamarche E 2005 Continuous flow in open microfluidics using controlled evaporation. *Lab Chip* **5**, 1355–9, Reproduced by permission of The Royal Society of Chemistry.)

2.13.3.2.2 Valving

A large variety of different active microvalves has been reported during the last two decades (Oh and Ahn 2006) associated to the development of displacement micropumps. However, especially in the field of lab-on-a-chip applications, the required power supply and the high costs of these active valves seem to limit their practical use in pressure-driven systems (**Table 4**).

Therefore, passive microvalves based on capillary stops, which are defined by a change in the channel cross section or the wetting properties of the channel, are more suitable for disposable systems. These valves work similar to passive valving structures in merely capillary-driven systems with the only difference being that the valve can be actively opened at any time by applying an additional pressure load onto the valve.

2.13.3.2.3 Mixing

Since no turbulences can be generated within the laminar flow regime, mixing is related to diffusion processes only in microfluidic channels. Hundreds of micromixing principles and structures have been presented during the last few years to overcome this limitation, see Hessel *et al.* (2005) and Nguyen and Wu (2005) for a detailed and up to date overview.

The basic idea of all micromixers is to increase the interfacial area between the fluids that should be mixed. Well-established methods for continuous mixing are multilamination (Ehrfeld *et al.* 1999, Hessel *et al.* 2003) and split-and-recombine (Chen and Meiners 2004, Cheng *et al.* 2001, Schonfeld *et al.* 2004) flow schemes. A large number of alternating thin lamellae of the liquids A and B are generated within these mixers (A–B–A–B–A···). This increases the interfacial area between the two liquids, which fastens up the homogenization of the two liquid phases via diffusion.

In a multilamination mixer, the initial liquid flows (phases A and B) are split into several substreams, guided to a junction area and then merged into an alternating phase pattern (A–B–A–B–A–B···). The split-and-recombine principle depends on the repetitive splitting of a phase pattern and combining the substreams again. This way the number of phase interfaces is doubled by each split-and-recombine stage (e.g., A–B leads to A–B+A–B).

Another method to increase the interface between two liquid flows is transversal advection. Centrifugal

Table 4 Comparison of active and passive microvalves

Properties	*Active valves*	*Passive valves*
Power supply	Required	Not required
Moving parts	Yes	No
Capable of complex fluid processing	Yes	Limited
Cost-efficient	No	Yes
Works well with fluids containing higher solvent or surfactant concentration	Yes	No
Design	Multilayer structures	Simple
Structure size	Large	Small

forces within bended channels, for example, are utilized to generate an additional flow component perpendicular to the initial direction of flow (Jiang *et al.* 2004, Sudarsan and Ugaz 2006). Due to this transversal advection, the contact interface increases along the channel and thus mixing is supported. Advection can also be induced in simple straight channels at low Reynolds numbers (<100) by manufacturing a (staggered herring bone-shaped) surface texture on one channel wall (Johnson *et al.* 2002, Stroock *et al.* 2002).

All the principles described so far are intended to mix different liquid flows and therefore called continuous micromixers. However, the mixing of liquids, located within a microchamber, can also be of interest. Active micromixers are mainly used for this purpose, e.g., by moving magnetic particles through the chamber.

2.13.3.2.4 Metering

Metering on pressure-driven platforms is typically realized by controlling liquid flows into or through reaction chambers possibly supported by flow rate sensors. With these methods, however, metering, isolation, and further processing of a discrete liquid volume are not possible. Therefore additional active or passive valving structures are required that can release the liquid out of a metering chamber after it has been metered to a defined volume via an overflow channel. Such a structure utilizing a passive, hydrophobic valve for stopping the liquid and metering it to a defined volume has been presented by Yamada and Seki (2004). Metered liquid volumes of 3.5 and 20 nl have been demonstrated in these structures.

2.13.3.2.5 Switching

Flow switching can be realized in a straightforward approach by integrating several (at least two) active microvalves, e.g., on the two continuative channels of a T-shaped structure. Depending on the desired flow path, one of both valves is opened while the other remains closed. However, for complex liquid processing, which is required for the implementation of assay protocols on the platform, a large number of active valves have to be integrated.

Another possibility to switch continuous liquid flows between several outlet channels is based on hydrodynamic focusing (Lee *et al.* 2001a, b). Generally, a sample flow that is symmetrically enframed by two sheath flows at the junction of a Y-shaped channel is focused to a thin stream line under laminar conditions (**Figure 9**) due to the so-called hydrodynamic focusing effect (Knight *et al.* 1998). Depending on the ratio of the flow velocities $\overline{v}_1$, $\overline{v}_2$, and $\overline{v}_3$, the sample liquid stream ($\overline{v}_2$) is directed toward one of the seven outlet ports (A–G). This switching method, however, is based on continuous flows and consequently a certain sample volume is required for liquid routing. This limits the applicability of the method to applications with large sample volumes only.

2.13.3.2.6 Separation

Separation processes are basic techniques to decrease the complexity of material mixtures and are often required in analytical chemistry. Many different physical and chemical separation methods are used in classic analytical assays, e.g., sieving, centrifugation, chromatography, crystallization, adsorption, distillation. Several of these methods have been used in microfluidic systems during the past few

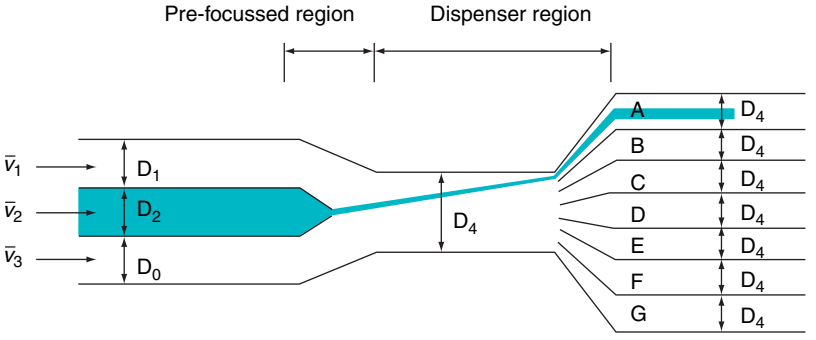


Figure 9 Principle of the continuous flow switch based on hydrodynamic focusing. The liquid stream $\overline{v}_2$ is focused by two lateral flows $\overline{v}_1$ and $\overline{v}_3$ within the prefocused region and then deflected into one of seven outlets (A–G) within the dispenser region. (Source: Lee G B, Hung C I, Ke B J, Huang G R, Hwei B H 2001a Micromachined pre-focused 1×*N* flow switches for continuous sample injection. *J. Micromech. Microeng.* **11**, 567–73, IOP Publishing Limited.)

years, and novel separation principles utilizing the fundamental different hydrodynamics in microchannels have also been proposed.

A separation method that depends on the different diffusion constants of particles of different sizes is the so-called H-Filter, developed and distributed by the company Micronics Inc. (2006). The functional principle of the structure is depicted in **Figure 10**. A sample liquid containing particles of different sizes (Inlet 2) and an extraction liquid without particles (Inlet 1) are combined at one side of a H-shaped channel network. Subsequently, they are guided through the horizontal connecting channel and are completely split into two fractions of equal flow rate at the outlet of the channel.

The flow within the connecting channels stays undisturbed due to the laminar flow conditions. Therefore, the particles present in the sample liquid leave the structure via outlet 2 as far as they do not enter the upper flow portion in the horizontal channel by diffusion. By setting the flow rate, the residence time within the connecting channel and thus the timeframe for the diffusion of the particles can be adjusted. Smaller particles can diffuse a larger distance within a certain timeframe compared with bigger ones. Thus, a fraction of small particles are found within the extraction liquid at outlet 1 and the bigger particles stay within the sample liquid at outlet 2 (Weigl *et al.* 2003). Using this method, the extraction of the antibiotic cephradine from human blood with a recovery of ~32% has been demonstrated (Jandik *et al.* 2002).

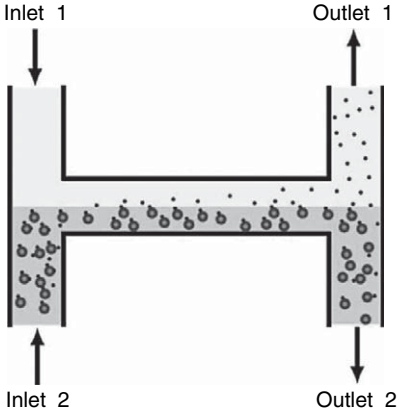


Figure 10 Functional principle of the H-filter. Depending on their diffusion coefficient, the particles start to diffuse across the fluid interface within the horizontal channel. Under optimized conditions, 50% of the population of small particles can be removed from a sample containing large particles also. (Reprinted with permission from Weigl B H, Bardell R L, Cabrera C R 2003 Lab-on-a-chip for drug development. *Adv. Drug Deliv. Rev.* **55**, 349–77. Copyright 2003 Elsevier.)

Another method that is based on particle diffusion utilizes the asymmetric bifurcation of laminar flow around obstacles. The particles follow different migration paths, depending on their size. Microspheres of 0.8, 0.9, and 1.0 µm size were successfully sorted within 40 s in such a device (Huang *et al.* 2004).

A hydrodynamic particle separation method, the so-called pinched flow fractionation (Yamada *et al.* 2004), is based on the parabolic flow profile and on the fact that the center of the particle cannot get close to the channel wall than the radius. The particles of different sizes within the sample flow are aligned to the channel wall by a second pinching liquid flow after a Y-shaped junction. The channel then expands and the flow spreads while retaining laminar conditions. Thus, each particle follows the streamline of its center. This leads to a separation of the particles perpendicularly to the flow direction according to their size. Using the pinched flow fractionation principle, the successful separation of 15- and 30-µm-diameter particles has been demonstrated. Even a mixture of 1- to 5-µm particles can be separated using an improved structure based on this principle (Takagi *et al.* 2005).

Another hydrodynamic method, called hydrodynamic filtration (Yamada and Seki 2005), is based on the topology of the particles and the laminar flow profile (**Figure 11**). Here, no additional liquid flow is needed for the alignment of the particles along the channel wall as is the case for the pinched flow fractionation. In the first part of the structure (concentration and alignment), only a small fraction of the liquid is withdrawn through the side channels (**Figure 11(a)**). Thus, neither of the particles enter the side channel since their center still lies beyond the flow fraction directed to the side channel (dashed line). Due to the permanent volume extraction on both sides of the channel, however, the particles are aligned onto both channel walls after some concentration steps.

In a subsequent side channel (first selection), the relative flow rates are adjusted to extract small particles from the suspension within the main channel (**Figure 11(b)**). At the second selection branch, the flow rate into the side channel is increased once again to extract the bigger particles. The concentration of polymer microspheres with diameters of 1–3 µm was increased 20- to 50-fold, and they were collected independently according to the size using such a

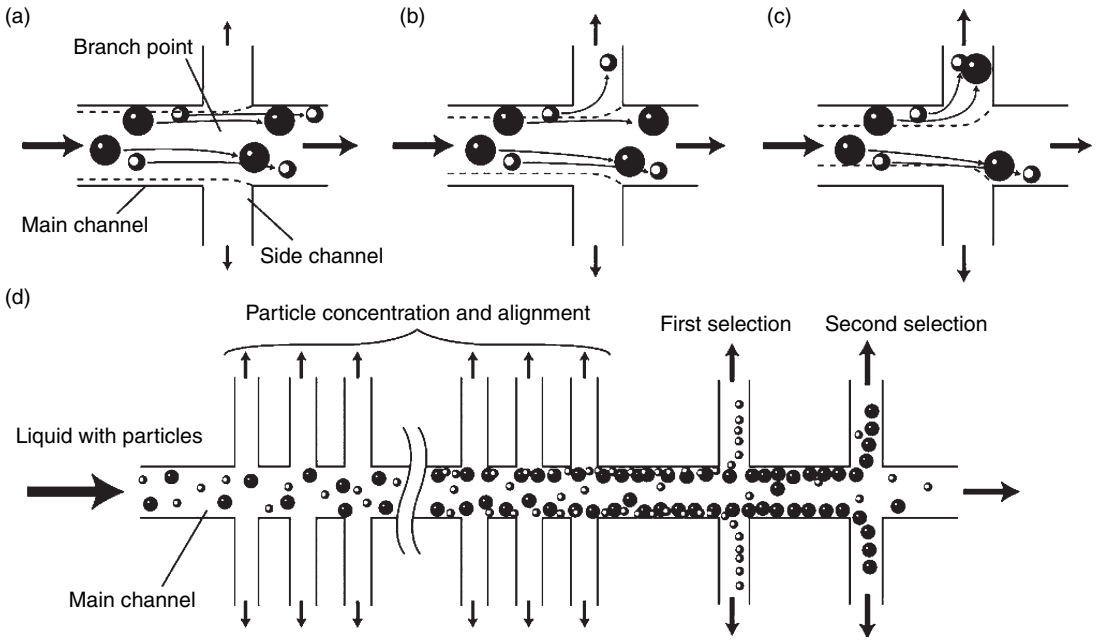


Figure 11 The hydrodynamic filtration method. The situation at a branch point for different flow ratios is depicted in (a)–(c). The relative flow rate into the side channel is (a) low, (b) medium, or (c) high; (d) The complete structure consists of a concentration and alignment as well as a selection part. (Source: Yamada M, Seki M 2005 Hydrodynamic filtration for on-chip particle concentration and classification utilizing microfluidics. *Lab Chip* **5**, 1233–9, Reproduced by permission of The Royal Society of Chemistry.)

structure. In addition, selective enrichment of leukocytes from blood was successfully demonstrated (Yamada and Seki 2005). Using an improved device for continuous particle concentration and size-dependent separation based on the same principle but featuring an additional splitting and recombining of the fluid flow, the concentration of 2.1- to 3.0-µm particles can be increased 60- to 80-fold (Yamada and Seki 2006).

2.13.3.3 Application Examples

2.13.3.3.1 Portable handheld analyzer: i-STAT®

The i-STAT® analyzer from Abbott Point of Care (2006) is a portable diagnostic platform. Several disposable cartridges for the determination of a multitude of blood parameters (blood gases, electrolytes, coagulation, cardiac markers, and hematology) can be read out using the same handheld analyzer (**Figure 12**, *left*). Only the low-cost polymer cartridge is contaminated with the blood sample and can be disposed after the diagnostic assays while the readout and assay processing analyzer is reused. The system

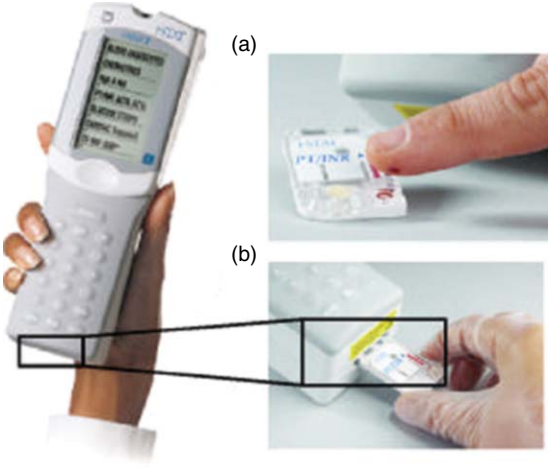


Figure 12 *Left*: The portable i-STAT® analyzer for clinical blood tests. *Right*: Depending on the blood parameters to be measured, (a) a certain disposable cartridge is filled with blood by capillary forces from the fingertip and (b) later loaded into the analyzer for assay processing and readout. (Source: Abbott P-C 2006 East Windsor, NJ, USA, www.abbottpointofcare.com, accessed 2006.)

reduces the time-to-result for the determination of important blood parameters down to several minutes, enabling faster diagnostic decisions.

The reagent solution for sensor calibration is contained in the cartridge within a foil pouch. As depicted in **Figure 12**, the blood sample is filled into the cartridge by capillary forces (**Figure 12(a)**) and placed into the analyzer (**Figure 12(b)**). During the subsequent assay cycle, the analyzer presses the front of the cartridge, causing a barb to puncture the pouch. This releases the calibrant solution and additionally generates a pressure-driven flow over the silicon sensor array for measurement. Subsequently, the analyzer presses an air bladder on the cartridge, which pushes the calibrant into the waste reservoir and the blood sample over the sensor array. Thereby, the blood parameters are determined depending on the type of cartridge and presented at the display of the handheld analyzer.

2.13.3.3.2 Diffusion-based assays: T-Sensor®

Micronics Inc. (2006) has developed the microFlow™ platform for the implementation of automated diffusion-based assays. The system consists of a disposable plastic card containing the microfluidic structures and a microfluidic work station for external pressure generation (syringe pumps) and detection. Microfluidic unit operations such as separation in the ActiveH™ card with an integrated H-filter or controlled mixing via diffusion through the ActiveT™ card, can be combined to conduct diffusion-based assays.

The microFlow platform as well as the ActiveT card containing the microfluidic T-Sensor® is depicted in **Figure 13**. A stable sheath flow is generated under laminar conditions at the junction of three inlet channels. No chaotic mixing occurs in the subsequent detection stream due to the strictly laminar flow conditions in the microchannel. The only transport of material from one stream to another is by diffusion. Consequently two diffusion zones on both sides of the central stream evolve. The left zone between the reference solution and the detection liquid represents a system-integrated real-time calibration. The other zone evolves at the right side of the detection stream being in contact with the sample liquid. The faster diffusion of smaller molecules also enables the detection of smaller components of interest present in a complex sample without previous separation. The reaction between the reference and the detection as well as between the sample and the detection stream is measured via a single fluorescent readout across the complete channel.

2.13.3.4 Strength and Challenges of the Platform

Pressure-driven platforms are certainly the most investigated ones in terms of the number of developed microfluidic components like micropumps, micromixers, microvalves, etc. However, only a small number of platform applications have been commercialized so far. The main reason for this is that many of the proposed components are based on different fabrication technologies. Thus, their combination on a pressure-driven platform causes liquid interfacing problems. However, the advantages of the pressure-driven approach are obvious: an increased level of controllability as compared with pure capillary-based test strips, easy integration of almost all

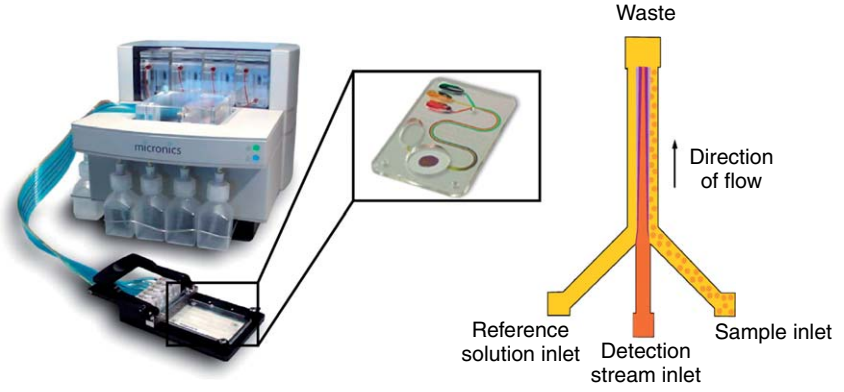


Figure 13 *Left*: The microFlow™ platform with an ActiveT™ card in the insert. *Right*: The functional principle of the T-Sensor® structure. The detection stream meets at the junction part with a reference liquid on the left and the sample liquid on the right side, respectively. (Source: Micronics Inc. 2006 Redmond, WA, USA, www.micronics.net, accessed 2006.)

biochemical sensor principles, and the compatibility to standard lab equipment.

The real potential of a pressure-driven platform becomes evident in Section 2.13.4, where the microfluidic large-scale integration (LSI) platform – a special type of pressure-driven platform – is described, which is probably the most sophisticated platform concept presented in the field so far.

2.13.4 Microfluidic Large-scale Integration

2.13.4.1 Introduction

Before the beginning of the 1990s, the screening for novel active agents in the pharmaceutical industry was done by laboratory staff in manually pipetted assays. A biological target molecule is exposed to many compounds in order to find a positive, i.e., matching compound (hit). The assay result is mainly read out optically (change in color or fluorescent signal) after the reaction. Since the texture of the biological target molecule is unknown, all possible combinations out of a compound library are tested. This merely statistical approach is called combinatorial chemistry.

Over the years, the number of screening experiments rapidly grew just as the compound libraries increased, making manual processing impossible. With the advent of highly automated liquid handling instrumentations such as pipetting robots, the number of experiments per day can be increased enormously. Over 100 000 experiments can be performed within a single day using these so-called high throughput screening (HTS) technologies. A current trend is the further reduction of the liquid volumes per screening experiment, e.g., by novel dispensing systems for liquid volumes in the nanoliter and picoliter range (see Section 2.13.7). Besides cost issues the finite amount of (biological) target molecules is the main driving force for this progression.

However, when shrinking reaction volumes why not also miniaturize the liquid handling devices to do thousands of chemical experiments on the footprint of a stamp? In the early 1990s, the realization of such an integrated microfluidic platform seemed unrealistic in terms of simply shrinking microvalves, micropumps, and mixers developed so far. Comparable to the invention of the transistor in 1947, a pivotal innovation in microfluidics was needed to overcome the existing barrier. This innovation arose with a novel fabrication technology for microfluidic channels, called soft lithography. Using this technology, the monolithic fabrication of all necessary fluidic components from one single elastomer material (polydimethylsiloxane, PDMS) became possible, similar to the silicon-based technology in microelectronics.

2.13.4.1.1 PDMS – Soft lithography

PDMS is an inexpensive but still powerful material offering several advantages compared to silicon or glass. It is a cheap, rubber-like elastomer with good optical transparency and biocompatibility. It can be structured using the soft lithography technique based on the replication molding on micromachined molds. It has first been used by George Whitesides's group for the fabrication of optical devices (Xia *et al.* 1996) and stamps for chemical patterning (Xia and Whitesides 1998a, b). A general and detailed up-to-date view of the use of PDMS in different fields of applications can be found in Sia and Whitesides (2003).

Thereafter, microfluidic devices have also been manufactured using PDMS technology (Delamarche *et al.* 1997, Duffy *et al.* 1998, Effenhauser *et al.* 1997, Fu *et al.* 1999, Hosokawa *et al.* 1999). So far, however, PDMS has been used merely as a passive material for the realization of microfluidic channels. The strength of the technology, however, really became obvious, when Stephen Quake's group expended the technology toward the multilayer soft lithography (MSL) process (Quake and Scherer 2000, Unger *et al.* 2000). With this technology, several layers of PDMS can be hermetically bonded on top of each other resulting in a monolithic, multilayer PDMS structure. The single layers are formed from a master structure either by casting or by using spin coating (**Figure 14**). One of the two layers contains a sparse cross-linking agent, while the other has an excess concentration. Both layers are separately prehardened, removed from the masters, and aligned to each other. Afterward, the PDMS stack is again baked, initiating the reaction of the cross-linking agent on the layer interface. A monolithic, multilayer elastomer structure results, which is finally bonded to a glass substrate for easier handling. Today, this technology is pushed forward by the company Fluidigm Corporation in the United States (Fluidigm Corporation 2006).

2.13.4.2 Microfluidic Large-scale Integration: Unit Operations

2.13.4.2.1 Valving

The resulting microfluidic platform mainly depends on the elastomer properties of PDMS with a low Young's modulus value of ~750 kPa (Lotters *et al.*

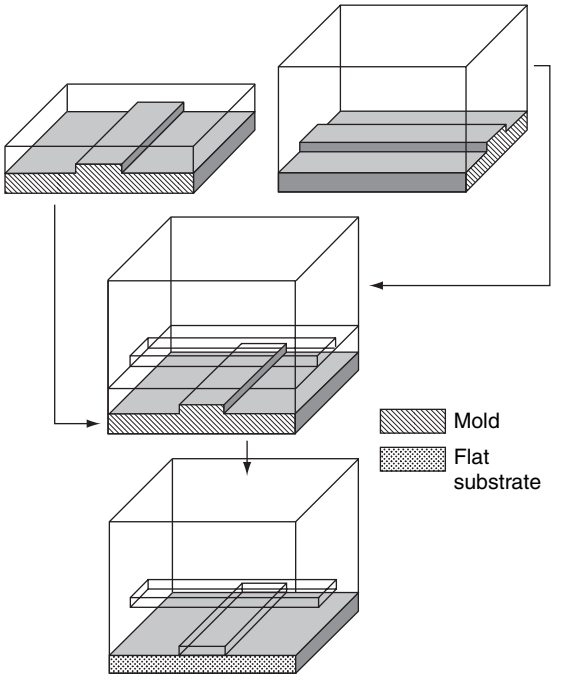


Figure 14 The multilayer soft lithography (MSL) fabrication principle. Two fluidic channel layers are casted from microfabricated molds. In a second step both layers are combined and compounded chemically. At last the compound is placed on a glass substrate for robust handling. (Reprinted with permission from Unger M A, Chou H P, Thorsen T, Scherer A, Quake S R 2000 Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113–16. Copyright 2000 AAAS.)

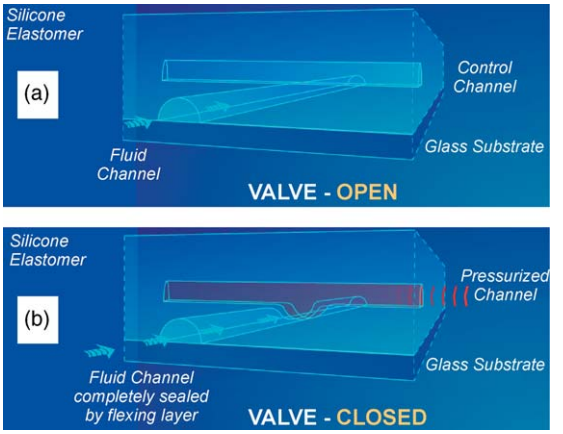


Figure 15 Schematic sketch of the microfluidic NanoFlex™ valve realized in the multilayer soft lithography (MSL) technology. (a) Liquid can pass through the fluid channel as long as no pressure is applied to the control channel. (b) An increased pressure within the control channel causes the thin elastic polydimethylsiloxane (PDMS) membrane between the two channel layers to deflect into the fluid channel and blocks the liquid flow therein, i.e., the valve is closed. (Source: Fluidigm Corporation 2006 San Francisco, USA, www.fluidigm.com, accessed 2006.)

1997). Based on this elasticity, the basic microfluidic unit operation is a valve, which is made of a planar glass substrate and two layers of PDMS on top of each other. The lower elastomer layer contains the fluidic ducts and the upper elastomer layer features pneumatic control channels. To realize a microfluidic valve, a pneumatic control channel crosses a fluidic duct as depicted in **Figure 15**. A pressure applied to the control channel squeezes the elastomer into the lower layer, where it blocks the liquid flow. Because of the small size of this valve on the order of $100 \times 100\,\mu m$, a single integrated fluidic circuit can accommodate thousands of valves. Compared to the development in microelectronics, this approach is called microfluidic large-scale integration (LSI) (Thorsen *et al.* 2002).

The valve technology called NanoFlex™ is the core technology of the complete platform. Placing two of such valves at the two arms of a T-shaped channel, for example, realizes a fluidic switch for the routing of liquid flows between several adjacent channels. Other unit operations based on the MSL technology are described in the following sections.

2.13.4.2.2 Pumps and mixers

The platform offers two possibilities to transport liquids. On the one hand, external pumps can be used to generate liquid flows within the fluid channel layer. With that approach, the PDMS multilayer device works merely passively, controlling the externally driven liquid flows with the integrated valves. On the other hand, an integrated pumping mechanism can be achieved by combining several microvalves and actuating them in a peristaltic sequence (**Figure 16(b)**).

Also a micromixer for the accelerated homogenization of liquids can be realized using the above-described pumping mechanism (**Figure 16(c)**). The liquids to be mixed are segmentally introduced into the fluidic loop through the left inlet, while the right outlet valve is still closed. Afterward, the inlet and the outlet valve are closed and the three control channels on the orbit of the mixing loop are displaced with a peristaltic actuation scheme leading to the circulation of the mixture within the loop (Quake and Scherer 2000). Thereby the liquids are mixed and afterward flushed out of the mixer by a washing

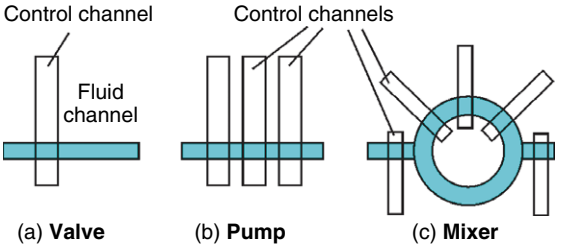


Figure 16 Realization of the main unit operations on the multilayer polydimethylsiloxane (PDMS)-based platform. Based on the NanoFlex™ (a) a valve, (b) an integrated peristaltic micropump, (c) and a micromixer can be designed.

liquid (inlet and outlet valves are opened again). Using this mixing scheme, an increase in the reaction kinetics of surface binding assays by nearly two orders of magnitude has been demonstrated (Chou *et al.* 2001).

Another mixing concept in the microfluidic LSI platform involves simply opening the valve between two different chambers containing different liquids. In this mixing scheme, the liquids mix only by diffusion. Although a merely diffusional mixing is very slow, it is also very controllable in return. A controllable and slow mixing process in particular is required for protein crystallization, which is described in more detail in Section 2.13.3.3.1.

2.13.4.2.3 Multiplexing

The key principle to tap the full potential of the LSI approach is the multiplexing technology, which allows the control of N fluid channels with $2 \log_2 N$ control channels as depicted in **Figure 17** (Thorsen *et al.* 2002). This, for instance, means that 256 fluid channels can be controlled individually by 33 pneumatic control channels.

Thorsen *et al.* demonstrated a microfluidic storage device with 1000 independent compartments of approximately 250 pl volume and 3574 microvalves. Therefore, two multiplexers, one for the row decoding and the other for the column decoding, have been used. The complete microfluidic chip has a surface area of 25 mm × 25 mm.

2.13.4.2.4 Metering

Liquid volumes can also be metered using crossed fluid channels and a set of microvalves. Addressed by a multiplexer, the liquid is loaded into a certain fluid channel in the first step of the metering protocol (**Figure 18**). Afterward the valves between the different chambers are closed and thus the continuous

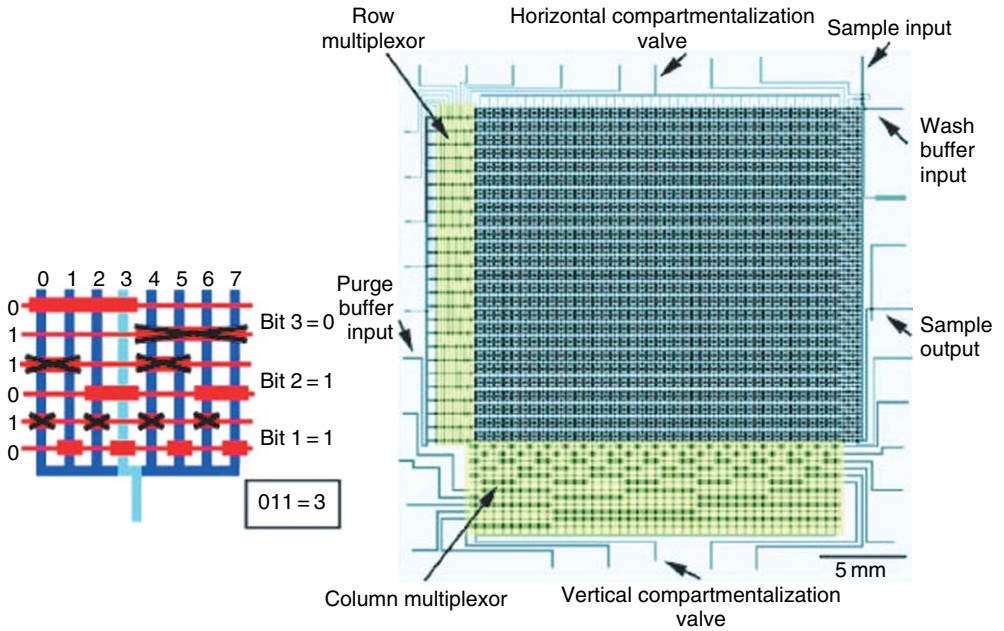


Figure 17 *Left*: Operational diagram of the microfluidic multiplexer: eight vertical fluid channels can be individually controlled by six horizontal control channels. The pressurized control channels block the fluid channels only at positions where the width of the control channel is widened. *Right*: Mask design for the microfluidic memory storage device. (Reprinted with permission from Thorsen T, Maerkl S J, Quake S R 2002 Microfluidic large-scale integration. *Science* **298**, 580–4. Copyright 2002 AAAS.)

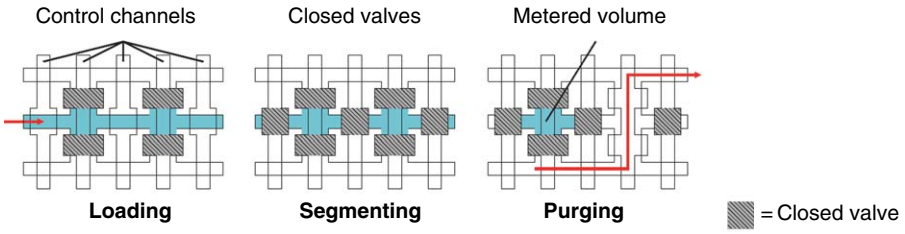


Figure 18 Using crossed fluid channels, liquid metering is also possible on the microfluidic large-scale integration (LSI) platform. In the first loading step, a certain fluid channel is addressed using the multiplexer structure and filled with liquid. In the second step, the liquid within that channel is segmented by closing the valves between the chambers. In the last step, fluid that should not be kept is purged via two additional fluid channels on both sides of the metering channel.

liquid plug within the fluid channel is segmented into several equal-sized (if the chambers all have the same dimensions) liquid compartments. During the last step, a washing liquid is pumped through the system, addressed to the fluid channel right next to the metering fluid line. All chambers that should not contain a metered volume of the initial liquid after the metering process are now purged vertically since the vertical valves to the two neighboring channels are now open. As a result a metered liquid volume is retained in designated metering chambers only.

The described unit operations should demonstrate the enormous possibilities of the platform. Depending on the design of the fluid channel and the control channel layer, many different processes for the conduction of certain assay protocols can be realized. To give an impression on these possibilities, two application examples, protein crystallization and nucleic acid processing, are described in Section 2.13.4.3, both realized on the microfluidic LSI platform.

2.13.4.3 Application Examples

2.13.4.3.1 Protein crystallization

How does this protein that causes a certain disease look like? Answering this question in an acceptable period of time would cause a revolution in the merely statistically driven pharmaceutical research based on the HTS of large compound libraries. The trail-and-error approach would be replaced with a more rational procedure; the systematic tailoring of a complementary active compound to the known structure of the target protein would then be feasible (structural biology).

In order to investigate the structure of a protein, first of all it has to be transferred into a crystalline state suitable for X-ray crystallography. As there is currently no way to predict crystallization conditions *a priori*, a large number of experiments are required to uncover successful crystallization conditions. It is obvious that a microfluidic-based solution would have many advantages, since it requires only a small amount of protein, and precise volume definition as well as controllable mixing processes by diffusion can be implemented.

The crystallization method that is used in the microfluidic LSI platform is called free interface diffusion (FID) (Salemme 1972). It is based on the counterdiffusion of the protein and the precipitant solution through the liquid–liquid interface between these two phases. During the diffusion process, the concentration profile changes and crystal growth is initiated as soon as the appropriate conditions are met. In conventional FID setups a stable and undisturbed liquid–liquid interface between the protein and the precipitant solution has to be assured by proper liquid handling to avoid unwanted mixing. Because of buoyancy-driven convection caused by density differences this can hardly be achieved in macroscopic capillaries.

Within a microfluidic crystallization structure, however, a stable interface with diffusion-based mixing only can be implemented very easily (negligible gravitational effects due to high surface-to-volume ratio). The successful growth of protein crystal has been demonstrated in a microfluidic chip based on the LSI technology (Hansen and Quake 2003, Hansen *et al.* 2002). Overall, 144 crystallization experiments can be performed in parallel on the presented microfluidic chip. One such microfluidic unit cell is depicted in **Figure 19(a)** and comprises microchambers of 5, 12.5, and 20 nl. They are arranged and coupled to define mixing ratios of 1:4, 1:1, and 4:1 with a fixed total volume of 25 nl.

The crystallization assay starts with the filling of the chambers while the valve between the chambers (horizontal yellow bar in the middle) is still closed (**Figure 19(b)**). The air is vented via diffusion

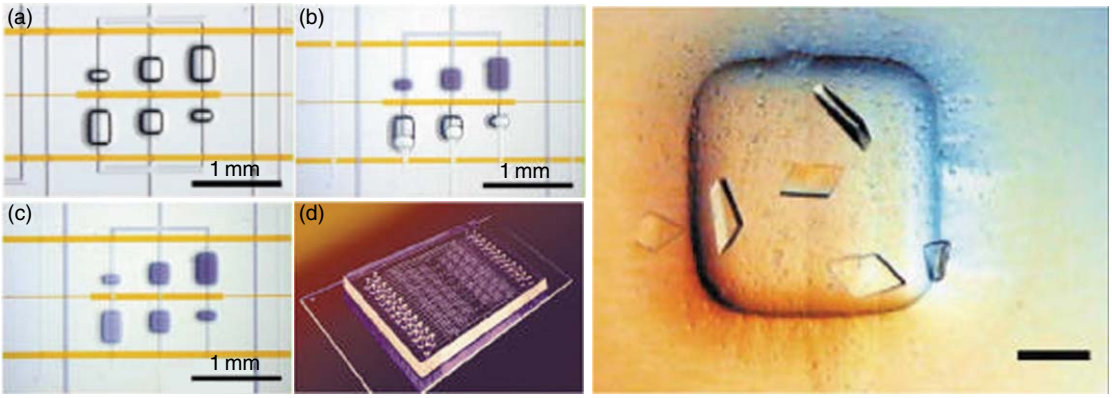


Figure 19 *Left*: Microfluidic crystallization structures. (a) Three crystallization assays with different mixing ratios (1:4, 1:1, 4:1) and a fixed total volume of 25 nl. (b) The central interface valve is closed and the chamber is filled with solution while the air diffuses through the bulk elastomer. (c) The intermediate valve is opened, initiating the diffusional mixing process. (d) The complete chip features 48 unit cells performing 144 simultaneous crystallization experiments while using 3 μl of protein sample only. *Right*: Example of protein crystals grown in the described microfluidic structure (scale bar, 100 μm). They have been grown within 12 h only, while the conventional process takes over a week. (Reprinted with permission from Hansen C, Quake S R 2003 Microfluidics in structural biology: Smaller, faster··· better. *Curr. Opin. Struct. Biol.* **13**, 538–44. Copyright 2003 Eksevier.)

through the gas-permeable bulk elastomer during the filling of the dead-end chambers. The filling valves are closed in the next step and the counterdiffusion is initiated as soon as the middle valve is opened (**Figure 19(c)**). The concentration within both chambers now changes with time due to the diffusion of species through the connecting channel. A crystal growth initiates in chambers with proper transient concentrations distributions only.

The channel between the two chambers has a small cross-section (10 μm × 100 μm) compared to that of the two chambers (300 μm × 100 μm). Thus, it acts like a diffusion impedance, slowing down the diffusion process and implying almost homogeneous concentrations within the chambers. As a concentration gradient is apparent only along the length of the channel, the rate of equilibrium can be influenced by the geometry of the structure (namely the length of the channel and the ratio of the chamber volume to the channel cross-sectional area). The end point itself, however, is predefined by the volume of the two precisely fabricated chambers. The accurate control over the kinetics of equilibration without changing the chemical evolution has important implications for crystal optimization where it is often desirable to approach crystallization conditions slowly while conserving the successful thermodynamic variables (Hansen and Quake 2003).

The parallel implementation of 48 unit cells in a single crystallization chip facilitates 144 different simultaneous protein crystallization reactions while consuming only 3.0 μl of protein solution – only 20 nl per assay (**Figure 19(d)**). **Figure 19** (*right*) shows a crystal that has been grown within 12 h using the microfluidic LSI chip. In a conventional system, this process lasts over 1 week. The protein crystallization technology on the LSI platform has been commercialized by the company Fluidigm (Topaz® technology).

2.13.4.3.2 Nucleic acid isolation

A second application example of the microfluidic LSI platform is the purification of nucleic acids from a small amount of cells. The isolation of DNA or mRNA from microbial or mammalian cells, including their lysis on-chip, has been showed by Stephen Quake's group at Stanford (Hong and Quake 2003, Hong *et al.* 2004). For the extraction of DNA from a cell suspension, the cell membrane has to be destroyed first (lysis of the cell). Then, the DNA has to be specifically separated from the residual cell constituents within the solution. This extraction protocol is completely implemented on the microfluidic platform using the basic unit operation structures for valving, metering, mixing, and switching of fluids.

The channel network for performing the DNA isolation assay is depicted in **Figure 20(f)**. Three extractions can be performed in parallel with three rotary reactors aligned vertically (the different fluid channel lines are colored with food dye for visualization). The chip is operated in two different modes, loading and processing. In the loading mode, the

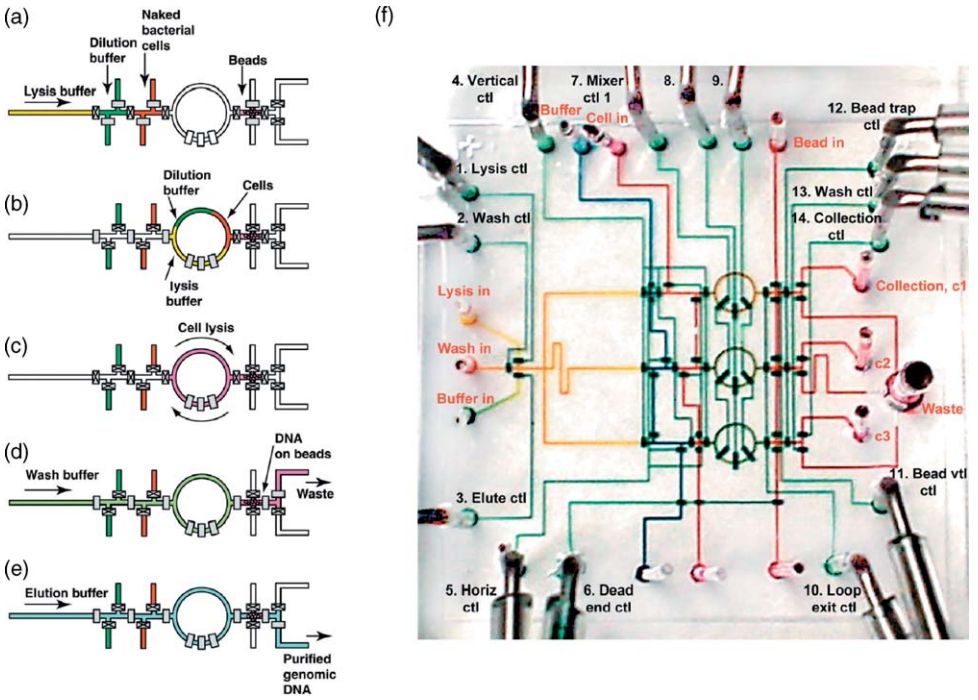


Figure 20 Microfluidic large-scale integration (LSI) DNA purification chip (f) and the course of an isolation protocol. (a) The bacterial cell culture, the dilution, and the lysis buffer are introduced into the chip via fluid channels (open valve, rectangle; closed valve, x in rectangle). (b) The vertical valves are opened, and the three liquids are externally pumped into the circular mixing channel. (c) Mixing is initiated by circular peristaltic pumping. (d) The mixture containing the lysed cells is flushed by a washing buffer over a column of microbeads. The DNA molecules bind to the bead surface and the buffers and the residual cell constituents are directed toward the waste outlet. (e) The elution buffer is flushed over beads, the DNA molecules are transferred into the liquid phase again and leave the structure through the lower outlet. (Reprinted with permission from Macmillan Publisher Ltd: Hong J W, Studer V, Hang G, Anderson W F, Quake S R 2004 A nanoliter-scale nucleic acid processor with parallel architecture. *Nat. Biotechnol.* **22**, 435–9. Copyright 2004.)

valves of the control layer are activated, so that the liquids flow in the north/south direction. During this step, the horizontal metering chambers are loaded with the diluent buffer (green), the bacterial cell solution (red), and the bead suspension (white) as depicted in **Figure 20(a)**. Then, the control valves on the north/south channels are closed leading to segments of the respective liquids. Also a certain volume of bead suspension is metered within the microchannel to define the number of beads that lead to a defined binding capacity for the later solid-phase extraction (100 pg of DNA on the beads of diameter 2.8 µm).

To exploit the high integration possibilities of the platform, the three isolation structures differ with regard to the volume of the bacterial cell solution. The volumes used in the top, middle, and bottom structure are 1.6, 1.0, and 0.4 nl, respectively. Since the total volume of each rotary reactor is the same (5 nl), 0.4-, 1.0-, and 1.6-nl dilution buffer are metered on the platform and added to the cell solution. Subsequently, the chip is switched to the processing mode: valves are actuated such that flow is allowed only in an east/west direction and all the three batch reactors are operated in parallel. Lysis buffer (3.0 nl) is injected from the left to flush the cell solution and the dilution buffer into the rotary reactor (**Figure 20(b)**).

Within the rotary reactor, the three different valves are actuated sequentially to generate a circulating flow. This reduces the mixing time from several hours for merely diffusional mixing to several minutes in the micromixing scheme (**Figure 20(c)**). As a result, the bacterial cells within the reactor are completely lysed after some minutes and can be forwarded to the next process step.

Now, the valve at the outlet of the rotary reactor is opened, and the solution containing the lysed cells is flushed out of the ring over the aggregated beads and leaves the structure via the upper waste outlet (**Figure 20(d)**). The beads are retained within the channel by a partially closed valve, where the beads larger than the valve opening cannot pass through. The DNA molecules within the lysed solution bind

to the bead surface. In a subsequent washing step, all unbound components are washed away from the bead column and likewise drained through the waste outlet.

Purified DNA is recovered from the chip in the last process step, by introducing elution buffer from the left side of the chip, which passes over the aggregated beads. The DNA is transferred back into the liquid phase and leaves the system through the lower DNA outlet. Two valves at the end of the structure act as a fluidic switch, guiding the liquid flow either toward the waste or toward the purified DNA outlet. The DNA gained from the on-chip isolation is now readily available for further analysis or manipulation. **Figure 20(f)** shows the complete microfluidic chip for four parallel DNA isolation assays; it has 26 access holes, 1 waste hole, and 54 integrated valves within the overall 20 mm× 20mm chip.

Purified genomic DNA from less than 28 bacterial cells (*Escherichia coli* culture) could be successfully isolated on the platform. Additionally, seven chips were used as negative controls in which pure water was introduced instead of bacteria, and no signal was detected within the off-chip detection (PCR amplification). Thus, it could be shown that it is possible to reduce the required amount of bacterial cells for the isolation of the needed DNA using the microfluidic LSI platform. This corresponds to an increase in the sensitivity of this process by three to four orders of magnitude compared with that of conventional methods (Hong *et al.* 2004).

Based on this technology, a complete nucleic acid processor is currently being developed in Stephen Quake's group at Stanford University. Starting from single cells, the isolation of nucleic acid, with a subsequent synthesis of cDNA (Marcus *et al.* 2006a), PCR (Liu *et al.* 2003), or real-time polymerase chain reaction (RT-PCR) (Marcus *et al.* 2006b) completes the platform for complete single-cell gene expression analysis.

2.13.4.4 Strength and Challenges of the Platform

The microfluidic LSI platform has certainly the potential to become one of the foremost microfluidic platforms for highly integrated applications. It is a flexible and configurable technology, which stands out owing to its suitability for LSI. The PDMS fabrication technology is comparably cheap and robust, and it can be used to fabricate disposables. Reconfigured layouts can be assembled from a small set of validated elements, and design iteration periods for new chips are on the order of days. Some of the system functions are hardware defined by the fluidic circuitry but others, such as process sequences, can be programmed from outside easily. The approach of building up an entire processor from a very limited set of elementary units resembles the microelectronics industry in which functional steps are configured by a circuit of transistors and capacitors.

Limitations of the platform are related to the material properties of PDMS: for example, chemicals that are not inert to the elastomer cannot be processed, or elevated temperatures, such as in microreaction technology, are not feasible. Also the implementation of applications in the field of point-of-care diagnostics, where often a handheld device is required, does not seem to be beneficial using the LSI platform. The external pressure sources and valves have to be shrunk to a smaller footprint, which is technically feasible of course, but the costs would be higher in comparison to other platform concepts.

2.13.5 Centrifugal Microfluidics

The approach of using centrifugal forces to process samples and reagents dates back to the end of the 1960s (Anderson 1969, Burtis *et al.* 1972). At that time, centrifugal analyzers were used to transfer and mix a series of samples and reagents in the volume range from 1 up to 110 µl into several cuvettes followed by spectrometric monitoring of reactions and real-time data processing. In the beginning of the 1990s, the company Abaxis Inc. (2006) developed the portable clinical chemistry analyzer (Schembri *et al.* 1992, 1995). The system consists of a plastic disposable rotor for processing the specimen, dried reagents preloaded to the cartridge, and an analyzer instrument for actuation and readout (**Figure 21**).

The next generation of centrifugal devices emerged from the technical capabilities offered by microfabrication and microfluidic technologies (Duffy *et al.* 1999, Ekstrand *et al.* 2000, Madou and Kellogg 1998, Madou *et al.* 2001). Length scales of the fluidic structures in the range of a few hundred micrometers allow parallel processing of up to a hundred units assembled on the disk. This enables a high throughput of many tests by highly parallel and automated liquid handling. In addition, the new opportunities arising from the miniaturization of the centrifugal fluidics cut down the assay volumes to <1 µl. In particular, fields such as drug screening, where precious samples are analyzed, benefit from the low assay volumes.

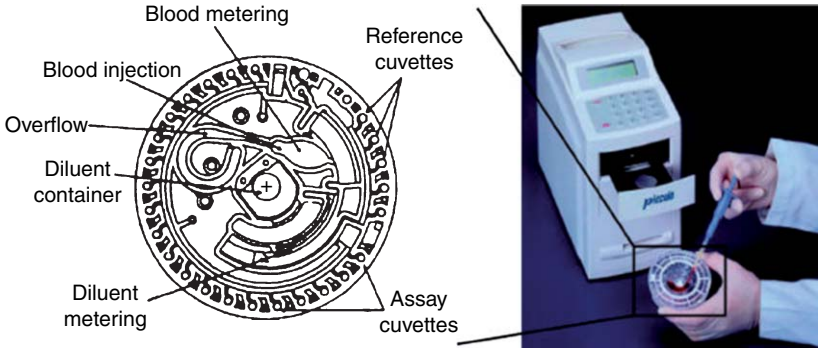


Figure 21 *Left*: Schematic of the rotor comprising the fluidic channels, reservoirs, and the cuvettes for optical detection. (Source: Schembri C T, Ostoich V, Lingane P J, Burd T L, Buhl S N 1992 Portable simultaneous multiple analyte whole-blood analyzer for point-of-care testing. *Clin. Chem*. **38**, 1665–70. Reproduced by permission of American Association for Clinical chemistry.) *Right*: Image of the portable analyzer. (Source: Abaxis Inc. 2006 Union City, CA, USA, www.abaxis.com, accessed 2006.)

Furthermore, the miniaturized assays can also offer shorter time-to-result and thus either higher throughput or shorter therapeutic turnaround times for critical care applications. The centrifugal microfluidic approach is also referred to as 'lab-on-a-disk'.

2.13.5.1 Introduction to the Lab-on-a-Disk Approach

The basic concept of centrifugal microfludics is the transport of fluids within a rotating channel by means of the centrifugal force. For a module rotating with an angular frequency $\omega = 2\pi v$, the centrifugal force density

$$f_\omega = \rho\omega(\omega \times \mathrm{r}) \qquad [3]$$

acts on a fluid of mass density ρ at a radial position r. This pumping force is directed in a radial outward direction, thus driving liquid plugs toward outer diameters as long as they possess a net radial plug length. Besides the liquid properties, only the radial position and the rotational speed define the pumping force and thus the liquid handling actuation. This eases the parallelization of several identical microfluidic structures being placed on one rotating substrate at different angular, but the same radial position. Although these structures experience the same centrifugal force, they still operate independently of each other, thus making this scheme of parallelization very robust.

The substitution of pressure-generating pumping devices by the pulse-free centrifugally initiated fluid flow enables complex liquid handling processes within a completely passive microstructure. Without any moving part, the rotating modules called disks or CDs can be fabricated using cost-efficient microfabrication technologies for plastics like injection molding. The processes are readily available from the audio compact disk industry, where microstructure polymer discs have been injection molded for several decades. Low-cost disposable disks combined with a nondisposable rotary drive and detecting unit build the modular centrifugal microfluidic platform.

Besides the centrifugal force, the Coriolis force can also be used to manipulate liquid flows on the centrifugal platform as depicted in **Figure 22**. It acts on every moving liquid portion within the rotating frame of reference and scales with the flow velocity as well as the frequency of rotation. The Coriolis pseudo force is always directed perpendicular to the direction of flow and even prevails over the centrifugal force over a critical frequency of rotation (Ducrée *et al.* 2005).

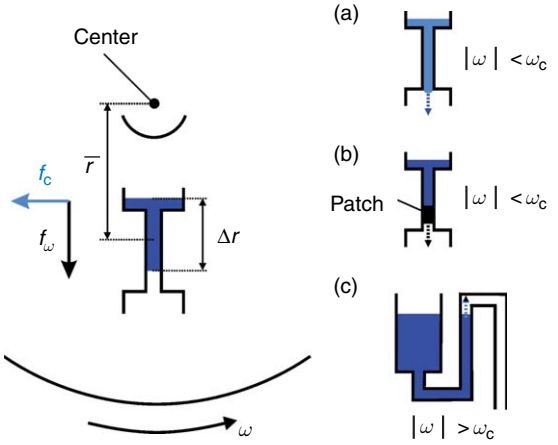


Figure 22 Principal centrifugal approach and schematic sketch of the three valving techniques on the centrifugal platform. (a) Geometric capillary valve, (b) hydrophobic valve, and (c) hydrophilic siphon valve.

2.13.5.2 Microfluidic Unit Operations on the Centrifugal Platform

2.13.5.2.1 Fluid transport

On the centrifugal platform, two pressure-generating mechanisms are employed to accomplish complex liquid processing protocols. The centrifugal force f_ω (see eqn [3]) enables the liquid transport in radial outward direction. This actuation principle can be scaled over a wide range by the frequency of rotation. Small flow rates in the order of nanoliter per second can be accomplished to extract molecules from a continuous flow while passing immobilized capture molecules. For this purpose the residence time for the streaming molecules have to be larger than the time for diffusion of the immobilized molecules. In return also high throughput up to $1\,\mathrm{ml\,s^{-1}}$ has been reported in a 3 cm-long microchannel of $260\,\mu\mathrm{m} \times 195\,\mu\mathrm{m}$ cross section for microprocess engineering applications (Haeberle *et al.* 2005). So scaling of flow rates over six orders of magnitude has been demonstrated. Many different (bio-) fluids can be processed on the platform, independent of their chemical composition, ionic strength, conductivity, or pH value opening a wide field of possible applications.

The second microfluidic effect that plays an important role for liquid propulsion is capillarity. As soon as the disk is at rest or rotates with a small frequency of rotation, the always coexistent capillary pressure gains impact. If there is at least one liquid–gas transition, a pressure drop appears at the interface. Liquid plugs can be transported in any direction, especially toward the center of the disk also in a hydrophilic channel using the capillary pumping mechanism. This is due to the fact that the capillary pressure is independent of the angular and radial position of the channel. It, for example, allows the complete initial priming of microchannels if the proceeding meniscus experiences a higher capillarity than the capillary pressure at the other end of the liquid plug. Most of the microfluidic structures for the different unit operations described in this section are based upon the interplay of centrifugal and capillary transport mechanisms.

2.13.5.2.2 Valving

Stopping of liquid flows, that is, the realization of a liquid valve, can be realized with different microfluidic structures on the centrifugal platform. A very simple valve arises at the sudden expansion of a microfluidic channel, e.g., into a bigger reservoir. The valving mechanism of this capillary valve is based on the surface tension that develops at this expansion and prevents further proceeding to the liquid–gas interface (**Figure 22(a)**). Under rotation, however, a certain pressure load

$$p_\omega = \rho\omega^2\overline{r}\Delta r \qquad [4]$$

due to the artificial gravity is applied on the liquid–gas interface. It depends on the liquid density ρ, the frequency of rotation ω, the mean radius $\overline{r}$, and the radial length Δr of the liquid plug (Duffy *et al.* 1999). As soon as this pressure exceeds the surface tensional counter pressure, the valve breaks. For a given liquid plug position and length, i.e., for a given set of geometric parameters, the valve can be influenced by the frequency of rotation only and a critical burst frequency ω_c can be attributed to every valve structure.

Another possibility to stop the liquid flow within a channel is the local hydrophobic coating of the channel walls. The hydrophobic patch defines the stop position for the liquid–gas interface within a hydrophilic channel of constant channel cross section (**Figure 22(b)**). Also this valve is opened as soon as the rotational frequency exceeds the critical burst frequency ω_c and the pressure load due to the liquid plug above the hydrophobic valve overcomes capillarity. The net capillary pressure

$$\Delta p_\theta \propto 2\sigma\left(\frac{\cos(\theta_1)}{r_1} - \frac{\cos(\theta_2)}{r_2}\right) \qquad [5]$$

of a liquid plug with surface tension σ, located in a channel of radius r_1 at the first and r_2 at the second liquid–gas interface can be used besides the radial plug length and position of the plug to engineer the critical burst frequency ω_c.

A third method utilizes the interplay of inertial and capillary forces in a different way. At high frequencies of rotation, the centrifugal force tends to leverage of the two liquid–gas interfaces within a U-shaped hydrophilic siphon channel of different cross sections (**Figure 22(c)**). Only a small difference in the radial position caused by the unequal channel cross sections and thus capillary pressures can be observed, which is negligible for high frequencies of rotation. Below a critical frequency ω_c, however, the meniscus proceeds toward the smaller duct due to the higher capillary force therein. As soon as the meniscus has passed the bend and proceeds in a radial outward direction again there comes a moment when a net radial length $\Delta r > 0$ exists. From that moment on, the centrifugal force supports the moving meniscus and the whole liquid plug escapes the siphon.

All the valves described are passive structures that do not need any movable parts within the chip. They are opened or closed by the frequency of rotation only. Therefore they can be considered as preprogrammed microfluidic networks with imprinted geometric parameters once fabricated. An alternative approach for the control of liquid flows on the centrifugal platform is followed by the company Spin-X Technologies (2006), Switzerland. A laser beam individually opens fluidic interconnects between different channel layers on a plastic substrate (virtual laser valve, VLV). This enables an online control of the liquid handling process on the rotating module almost independently of the geometry of the structures. Consequently, metered volumes and incubation times can be manipulated within a wide range. Hence, the Spin-X platform works with a standardized fluidic cartridge that is not custom-made for each specific application, but can be programmed right before the measurement or even online during a running process.

2.13.5.2.3 Metering and aliquotting

The valving principles, described in Section 2.13.5.2.2, present the basic fluidic function for many unit operations on the centrifugal platform. Using hydrophobic valves, for example, the exact metering of liquid volumes can be accomplished, which is one of the most important unit operations needed for all quantitative assay protocols. Especially in point-of-care applications, only an undefined sample volume, for example, a droplet of blood derived from the fingertip, is available. However, for quantitative analysis a defined sample volume is required, which has to be defined within a designated metering structure. The high precision of microfabrication technologies facilitates small coefficients of variations (CVs = standard deviation/mean value) down to metered volumes in the lower nanoliter range (Steinert *et al.* 2007).

Typical metering structures exhibit a hydrophobic valve at the outlet and an overflow channel connected to a metering chamber of designated volume (Steigert *et al.* 2005). During the metering process, the liquid initially fills the metering chamber at a frequency below the burst frequency ω_c. The liquid stops at the hydrophobic valve (**Figure 23**). Venting of the chamber is done via the outlet or a separate venting channel. After the metering chamber is completely filled with liquid, the excessive fluid volume is removed via an overflow channel and is cut off at a defined position at the inlet of the structure. Thereby, the volume is defined to the geometrically imprinted volume of the metering chamber while the valve still holds back the liquid since the centrifugal pressure falls short of the critical pressure of the valve.

In the last step, the frequency is increased and the hydrophobic valve breaks to release the metered liquid portion through an outlet channel toward further downstream processes. Using this technology, metered liquid volumes from the microliter to the nanoliter range with an extremely high reproducibility, e.g., a CV < 5% for a volume of 300 nl, have been demonstrated (Steigert *et al.* 2006).

By arranging several metering structures of, for example, different volumes interconnected via an appropriate distribution channel, simple aliquotting structures can be realized (Zoval and Madou 2004).

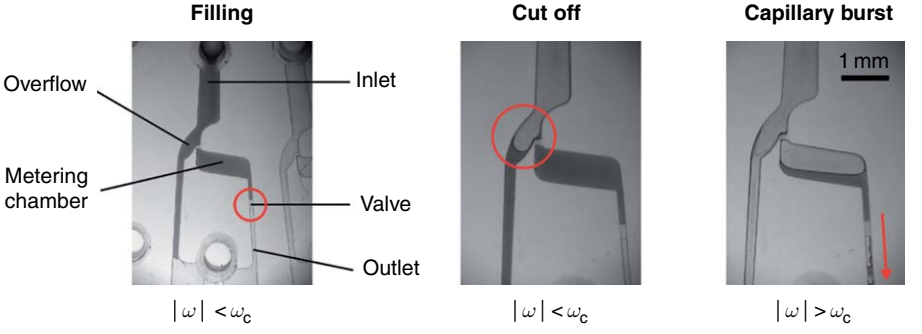


Figure 23 Metering structure based on a hydrophobic valve. In the initial filling stage, the fluid network is primed with an undefined sample volume. After the metering chamber is filled, the residual fluid is cut off at a frequency still beyond the critical burst frequency of the valve ω_c. In the last step the frequency is increased over the burst frequency of the valve leading to a radial drainage of the metered volume through the structures outlet. (Source: Steigert J, Grumann M, Brenner T, Riegger L, Harter J, Zengerle R, Ducrée J 2006 Fully integrated whole blood testing by real-time absorption measurement on a centrifugal platform. *Lab Chip* **6**, 1040–4. Reproduced by permission of The Royal Society of Chemistry.)

These structures split a sample into several defined volumes enabling the parallel conduction of several assays from the same sample. Multiparameter diagnostics and pharmaceutical screening are possible applications where such an aliquotting structure is required.

2.13.5.2.4 Mixing

Mixing of miscible fluids within the laminar regime on the centrifugal platform can be in the simplest case accomplished by joining two microfluidic channels to a single meander-shaped channel. By maintaining the feeding flows, a stoichiometric correct mixing ratio can be achieved. The required length of the mixing channel depends on the flow rate, i.e., the retention time within the channel, as well as the diffusional velocity in the transversal direction. In general, mixing processes can be speeded up by increasing the interfacial area between the phases to be mixed. However, if two liquid streams are simply contacted within the laminar regime, the contact interface equals the channel height and does not increase in the downstream direction. An increasing interfacial area along the direction of flow is possible if the liquid moves not only along the channel but also perpendicular to it (transversal advection).

Considering mixing of continuous liquid flows within a radial-directed rotating channel, the transversal-directed Coriolis force f_C automatically generates a transversal liquid flow $u(f_C)$ as depicted in **Figure 24** (Ducrée *et al.* 2005, 2006, Haeberle *et al.* 2005). The Coriolis force scales with the radial flow velocity $u(f_\omega)$, which has its maximum in the center of the channel due to the parabolic radial velocity profile (no slip boundary condition at the channel walls). Therefore, at the center of the channel, the flow experiences the highest Coriolis force f_C directed laterally. This leads to a transversal liquid movement $u(f_C)$ from the center to the left and backward on the upper and lower channel parts where the smallest Coriolis counter force is present. This transversal advection, also called Coriolis stirring, constantly increases the interfacial area between the two concurrent liquid flows and thus accelerates mixing.

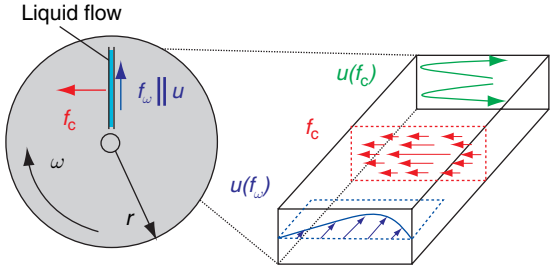


Figure 24 The Coriolis stirring effect within a radial, rotating microchannel. The inhomogeneous transversal Coriolis force field f_C initiates a transversal advection $u(f_C)$ thus increasing the interface between two liquid streams. (Source: Haeberle S, Brenner T, Schlosser H P, Zengerle R, Ducrée J 2005 Centrifugal micromixer. *Chem. Eng. Tech.* **28**, 613–16. Reproduced by permission of Wiley VCH.)

A continuous centrifugal micromixer, utilizing the Coriolis stirring effect, showed an increasing mixing quality toward very high volume throughputs of up to 1 ml s^{-1} per channel. This is enabled by the increasing impact of the transversal Coriolis force toward higher flow rates (Haeberle *et al.* 2005). It opens an interesting and so far barely explored microprocess engineering field for centrifugal microfluidic technologies, which however is not discussed in detail here.

Besides the mixing of continuous liquid flows, the homogenization of discrete and small liquid volumes inside small chambers is also of importance especially when analyzing small sample volumes (batch-mode mixing). One possibility to enhance the mixing of two small liquid portions is to guide them into a common chamber. The streams violently splash against a common chamber wall causing their efficient mixing (Zoval and Madou 2004).

Another possibility is the active agitation of the liquid within a mixing chamber either by using periodically actuated magnetic microparticles under rotation (Grumann *et al.* 2005) or by means of an inertial effect induced by a fast change of the sense of rotation (shake-mode mixing). The course of mixing within a microchamber under the impact of the shake mode is depicted in **Figure 25** (*left*). The course of mixing strongly depends on the applied mixing scheme as depicted in **Figure 25** (*right*). The shake-mode mixing under an alternate spinning protocol drastically enhances the mixing efficiency. Characteristic mixing times on the order of several seconds compared to several minutes for pure diffusion-based mixing can be achieved. The shake-mode mixing has been integrated into a complete colorimetric assay protocol as described in Section 2.13.5.3.4.

2.13.5.2.5 Switching

Switching liquid flow from one common channel to one of at least two continuative channels is very important, e.g., for extraction protocols. A promising switching method on the centrifugal platform should

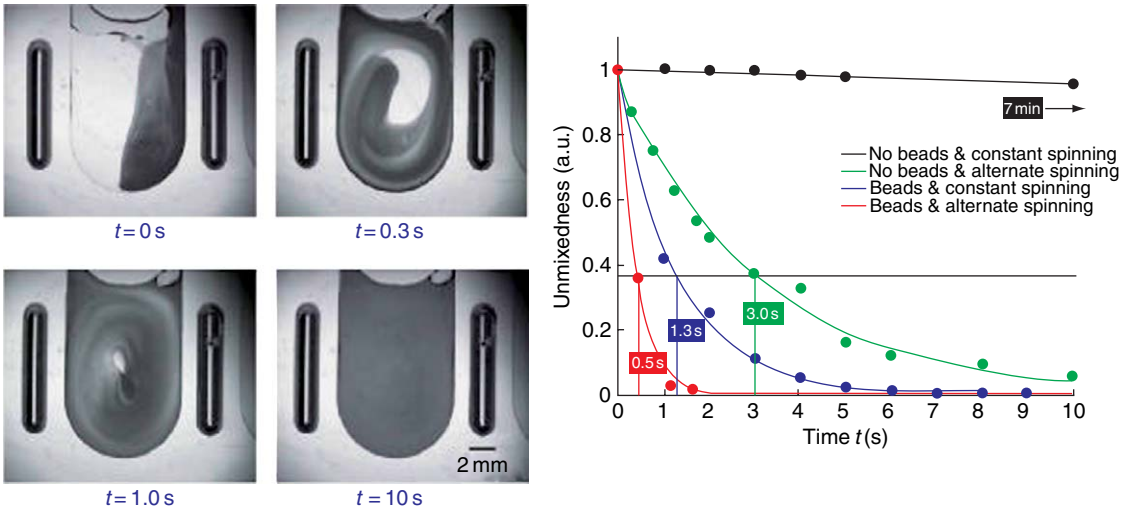


Figure 25 *Left*: Two initially separated liquids (pure water on the left and ink on the right) are mixed under the impact of the shake-mode mixing within 10 s. *Right*: Typical mixing times for different mixing schemes: no beads and constant spinning represent mixing via diffusion only. Alternate spinning describes the shake-mode mixing scheme. Additional magnetic beads within the chamber can be used to further decrease mixing times. (Reprinted with permission from Steigert J, Grumann M, Brenner T, Mittenbühler K, Nann T, Rühe J, Moser I, Haeberle S, Riegger L, Riegler J, Bessler W, Zengerle R, Ducrée J 2005 Integrated sample preparation, reaction, and detection on a high-frequency centrifugal microfluidic platform. *J. Assoc. Lab. Autom.* **10**, 331–41. Copyright 2005 Elsevier.)

be possible without the integration of any active or moving part. Following this philosophy, a liquid switch availing the transversal Coriolis force to guide liquid flows between two outlets at the branching of an inverse Y-shaped channel has been presented (Brenner *et al.* 2005).

The functional principle is depicted in **Figure 26**. Depending on the sense of rotation, the Coriolis force is either directed to the left (counterclockwise rotation) or to the right (clockwise rotation). One common radial channel is split toward a left and a right reservoir at the branching. The liquid stream is directed into one of these reservoirs depending on the direction of the Coriolis force, i.e., the sense of rotation. A fully reliable switching behavior has been reported for frequencies above ~55 Hz for a continuous flow. An improved version of Coriolis-based switching operates on individual droplets and enables switching of small flow rates of 160 nl s^{-1} only at low frequencies down to a few Hz (Haeberle *et al.* 2006b).

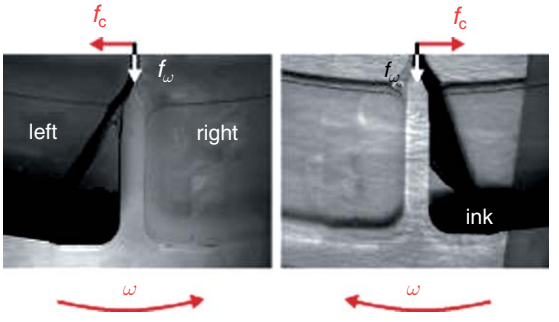


Figure 26 Liquid streams are switched between two possible outlet channels at the Coriolis flow switch, an inverse Y-shaped channel. (Source: Brenner T, Glatzel T, Zengerle R, Ducrée J 2005 Frequency-dependent transversal flow control in centrifugal microfluidics. *Lab Chip* **5**, 146–50. Reproduced by permission of The Royal Society of Chemistry.)

Another method for liquid routing based on different wetting properties of the continuative channels has been reported by Gyros AB (2006), Sweden. The liquid stream is initially guided toward a radial channel, exhibiting a hydrophobic patch at the beginning. Therefore, the liquid is deflected into another channel, not the hydrophobic channel next to the radial one. For high frequencies of rotation, the approaching liquid possesses sufficient energy to overcome the hydrophobic patch and is therefore routed into the radial channel. This enables the frequency-dependent routing of liquid flows between two consecutive channels (Ekstrand and Thorsen 2005).

2.13.5.2.6 Droplet formation

Monodisperse droplet emulsions can be generated in an adopted flow focusing structure under the centrifugal gravity field. Trains of water droplets dispersed into a continuous oil stream with droplet volumes

between 5 and 22 nl with a CV of the droplet diameter below 2% are reported (Haeberle *et al.* 2006c). The flow focusing structure contacts a central water-guiding channel to two symmetric oil flows at a junction. During the droplet formation process, a water plug is expelled out of the central channel into the junction area and is subsequently forced through a geometrical constriction. This leads to the controlled break-off of small water droplets into the continuous oil flow and thus a monodisperse droplet emulsion. Additional droplet-based unit operations like splitting of droplets or droplet sedimentation have also been realized on the centrifugal multiphase platform.

2.13.5.2.7 Separation

Within the artificial gravity field under rotation, different phases can be separated if they possess different mass densities. This process is called sedimentation and is used as a standard routine in laboratory centrifuges. Since lab-on-a-chip systems, especially within the field of point-of-care testing, deal with raw biological samples, which have to be pretreated before the analytical assay can be conducted, a system-immanent separation method is of great importance.

The extraction of plasma from a blood sample is the prevalent first step in a complete analytical protocol starting from an untreated whole blood sample. Since blood plasma is less dense than the white and red blood cells it is found in the upper phase, the so-called supernatant after sedimentation. The separation of the cellular and plasma phase in a sedimentation chamber, however, is just one part of a complete plasma extraction process. The second step is the spatial extraction of the gained plasma from the cellular pellet, located at the bottom of the chamber. This extraction is done by manual or automated pipetting in classic macroscopic plasma extraction protocols. However, to arrive at an extraction method capable of the seamless integration into automated centrifugal platform protocols, the plasma phase has to be extracted automatically within the microfluidic channel network.

Two methods have been proposed to achieve an integrated plasma extraction on the centrifugal platform. In the first method, plasma and cells are separated in a sedimentation chamber. A capillary extraction channel branches from the sedimentation chamber at radial position where only plasma is expected. It is used to extract the supernatant plasma phase at lower frequencies via capillary action (Schembri *et al.* 1995). The whole blood sample has to be metered prior the sedimentation process in order to guarantee proper function and also to result in a metered extracted plasma volume. A complete analyzing system using this extraction method is described in more detail in Section 2.13.4.3.

Another method uses a preseparation of the cellular and plasma phase during the sample flow through an azimuthal-aligned channel of just 300 µm radial width (Haeberle *et al.* 2006a). The obtained plasma fraction is thereafter split from the cellular components by a decanting process (**Figure 27**, *left*).

The whole extraction proceeds as follows: Initially, a raw blood sample is metered to a fixed volume defined by an overflow channel next to the outlet of the metering chamber. Subsequently, the metered sample is forwarded via the drain channel to the decanting structure. Within the drain channel the blood is preseparated with the plasma phase flowing on top of the cellular phase. The two joint streams separate when they enter the decant chamber. The more dense cellular phase sinks outward while the less dense plasma phase stays on top. When the decant chamber is entirely filled, pure plasma in the supernatant overflows into a plasma chamber (**Figure 27**, *right*). The course of separation concludes when the entire blood sample has been transported from the metering chamber into the

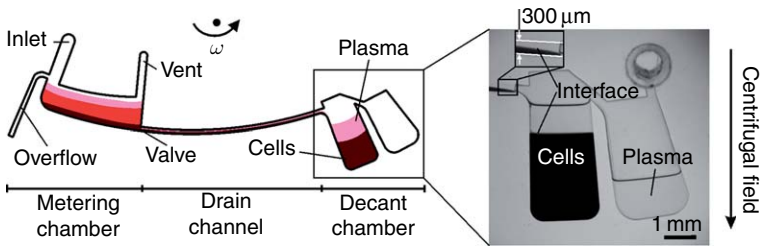


Figure 27 *Left*: Flow scheme of the decanting structure for plasma extraction from a raw blood sample. *Right*: Purified plasma is decanted into a separate reservoir while the cellular pellet is retained at the bottom of the decant chamber. (Source: Haeberle S, Brenner T, Zengerle R, Ducree J 2006a Centrifugal extraction of plasma from whole blood on a rotating disk. *Lab Chip* **6**, 776–81. Reproduced by permission of The Royal Society of Chemistry.)

decant and the plasma chamber. The extracted plasma volume is defined by the chamber geometries and can therefore be directly forwarded to further downstream processes, e.g., by a capillary duct.

This centrifugal flow separation technique extracts 2 µl of plasma from a raw blood sample that is initially metered to 5 µl. Typical separation times of 20 s can be achieved for moderate spinning frequencies of 40 Hz. The residual cell concentration in the bulk plasma is <1‰.

2.13.5.3 Application Examples

2.13.5.3.1 Blood test (Abaxis Picolo®)

The Abaxis Picolo® system (Abaxis Inc., 2006) is a centrifugal analyzer for near-patient testing (Schembri *et al.* 1992, 1995). The rotor, with a diameter of approximately 8 cm, is made of molded 2 cm-thick polymethylmethacrylate (PMMA) plastic (see **Figure 21**). It consists of a series of many interlinked chambers and passages. The movement of fluid is controlled by a series of stop junctions, capillaries, and siphons managed by the use of centrifugal force. The center of the rotor hosts a diluent container while dried reagent beads are placed in cuvettes in the periphery of the rotor. Fifteen cuvettes are reserved to analyze the patients' sample and further ten are used as internal quality control. The loading port for blood injection is coated with heparin to prevent blood coagulation.

The operator adds an unmetered blood sample between 90 and 120 µl. The sample preparation steps such as metering, blood separation, and mixing are performed in the rotor in an automatic fashion. Once the rotation starts, the centrifugal force propels the diluent outward into a chamber that contains a ball coated with dye. Simultaneously, the blood sample is metered to 75 µl and moves into the separation chamber, where 20 µl of plasma is obtained within 30 s. The blood plasma is afterwards diluted and distributed to the cuvettes for detection. The chemical reactions in the cuvettes are optically monitored and checked with an enzymatic control in the reference cuvettes before the results from the analysis are reported. The centrifugal platform enables point-of-care measurements of multiple analytes, e.g., glucose of whole blood samples in about 10 min. At present, Abaxis supplies the Piccolo platform for medical and veterinary diagnostics.

2.13.5.3.2 Enzyme-linked immunosorbent assay (LabCD)

Madou *et al.* from University of California, Irvine, CA, showed a series of capillary valves to perform enzyme-linked immunosorbent assays (ELISAs) on the centrifugal platform (Lai *et al.* 2004). An antibody is immobilized in the detection reservoir 2 in an off-chip protocol (**Figure 28**). Different reagents are then loaded into their corresponding reservoirs, numbered 3–7 on the CD, which is then mounted on a motor plate.

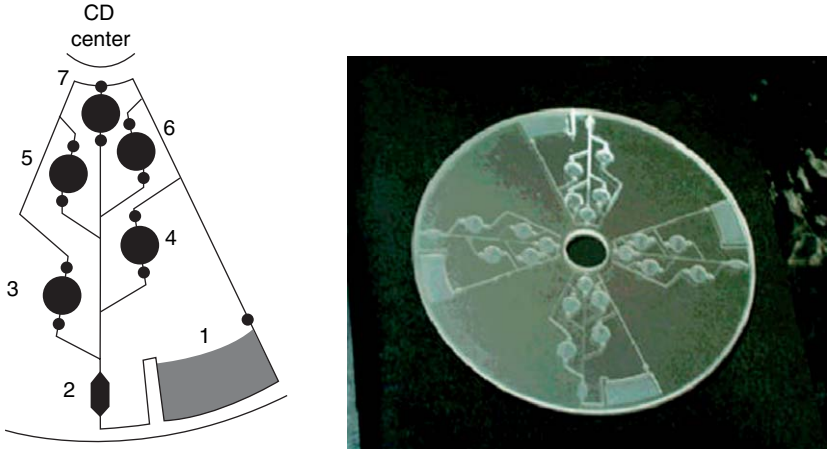


Figure 28 *Left*: Channel design for the enzyme-linked immunosorbent assay (ELISA) assay. Several reservoirs containing different sample and washing solutions can be opened serially depending on the frequency of rotation. *Right*: Picture of the compact disk fabricated by computer numerical control (CNC) milling with four identical assay structures. (Reprinted with permission from Lai S, Wang S, Luo J, Lee L J, Yang S T, Madou M J 2004 Design of a compact disk-like microfluidic platform for enzyme-linked immunosorbent assay. *Anal. Chem.* **76**, 1832–7. Copyright 2004 American Chemical Society.)

Under rotation (360 rpm), first the sample solution is released from reservoir 3 into reservoir 2 for the antigen–antibody binding process. The capillary valve at the outer radius of reservoir 3 is designed with the lowest burst frequency to accomplish an automated assay procedure. After an incubation step, reservoir 2 is washed with washing solution from reservoir 4 at a higher frequency of rotation (560 rpm, greater than burst frequency of reservoir 4). Following this serial flow-through assay scheme, the conjugate solution (reservoir 5) is released at 790 rpm, the detection chamber is washed (reservoir 6) at 1190 rpm, and finally the substrate solution (reservoir 7) is flushed through reservoir 2 at 1280 rpm. Immediately after the release of the substrate, the disk is stopped and the assay result is detected using an inverted fluorescent microscope.

It can be shown that the centrifugally conducted assay has the same performance in terms of detection range as the conventional method on the 96-well microtiter plate, while having advantages over the conventional method such as less reagent consumption and shorter assay time.

2.13.5.3.3 Protein quantification (Gyrolab Bioaffy®)

Gyros AB (2006), Sweden uses a flow-through sandwich immunoassay at the nanoliter scale to quantify proteins within the Gyrolab™ Workstation. Therefore, a column of prepacked and streptavidin-coated microparticles is integrated in each of the total 112 identical assay units on the microfluidic CD. Each unit has an individual sample inlet and a volume definition chamber that leads to an overflow channel. Defined volumes (200 nl) of samples and reagents can be added to the prepacked particle column (**Figure 29**). Samples and reagents can be added to a specific unit via an individual inlet port or via a common inlet feeding defined clusters. Capillary action draws samples and reagents into the microstructures where volumes are precisely defined. This enhances reproducibility and reliability as well as eliminates concerns about pipetting precision. After the sandwich assay protocol is completed, the assay result is read out using a detection reagent bound to the target protein and laser-induced fluorescence. The laser-induced fluorescent (LIF) detector is incorporated into the Gyrolab Workstation.

Using this technology, multiple immunoassays with 200 nl of sample have been carried out to determine the imprecision of the assay result. The day-to-day (total) imprecision (CV) of the immunoassays on the microfluidic CD is <20% (Honda *et al.* 2005). The assays were carried out within 50 min, while in comparison the traditional ELISA in a 96-well plate typically takes a few hours, with sample volumes of a few hundred microliters.

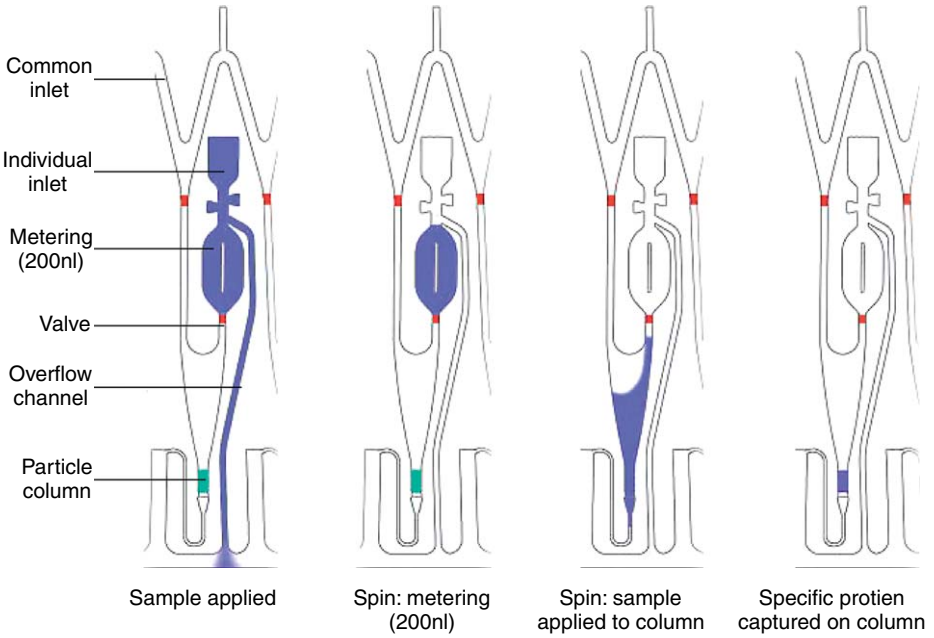


Figure 29 Metering step of the sandwich immunoassay protocol shown for one microstructure unit in Gyrolab Bioaffy® system. The sample is applied to the individual inlet port and metered to a volume of 200 nl. Subsequently, it flows through a column of functionalized particles. A certain protein specifically binds to a surface-immobilized capture protein and is read out in a subsequent assay step. (Source: GYROS AB 2006 Uppsala, Sweden, www.gyros.com, accessed 2006.)

2.13.5.3.4 Alcohol test (Bio-Disk)

A fully integrated colorimetric assay for the determination of the alcohol concentration in human whole blood has been shown on the centrifugal Bio-Disk platform (Steigert *et al.* 2006). After loading the reagents into the reagents reservoir, a droplet of untreated human blood derived from fingertip is loaded into the inlet port of the microstructure. All process steps of the assay are implemented by a fully automated frequency protocol.

First, a small 500 nl blood volume is metered from the sample and purged into the detection chamber at a frequency exceeding the burst frequency of the hydrophobic valves. At this frequency, the assay reagents also enter the reaction chamber. To enforce rapid mixing the sense of rotation is frequently reversed for 10 s, leading to a homogeneous mixture (shake-mode mixing). Thereby an enzymatic reaction is initiated by changing the color of the mixture depending on the alcohol concentration. During the subsequent sedimentation step, the disk is spun at a frequency of 30 Hz for 30 s so that all the cells contained within the blood sample settle at the radial outer end of the detection chamber. After sedimentation, the absorbance is monitored in real time during constant spinning for about 100 s to obtain the optical density of the mixture.

According to the Beer–Lambert law, the absorbance (or optical density) linearly depends on the molar extinction coefficient of the solution, which is governed by the products of the colorimetric reaction and thus the alcohol concentration. Since the absorbance scales with the optical path length l_{abs}, the sensitivity of this readout method can be enhanced by elongating the path of the laser beam through the analyte solution. A flat polymer substrate, however, enables only small optical lengths in readout schemes perpendicular to the disk surface. Therefore micromachined V-grooves placed on both sides of the detection chamber to deflect the laser beam by total internal reflection into the disk plane have been applied on one side of the disk (total internal reflection (TIR) set-up, **Figure 30**).

Using this automated assay and readout protocol the concentration of alcohol in human whole blood can be determined within 150 s only. The results are comparable to common point-of-care tests and require a minute blood volume of just 500 nl.

2.13.5.4 Strengths and Challenges of the Platform

The modular set-up of the system with cheap, disposable, and easy exchangeable plastic cartridges is certainly one major advantage of the centrifugal microfluidic platform, especially in the field of diagnostics. The cost-efficient fabrication predominantly originates in the simple and passive microfluidic elements that can be easily combined in a monolithic way within the same fabrication process. Those elements allow to implement all unit operations needed to perform complex assay protocols in an automated way. Due to the rotational symmetry of the disks optionally a high degree of parallelization can be achieved. All processes are controlled by the frequency of rotation

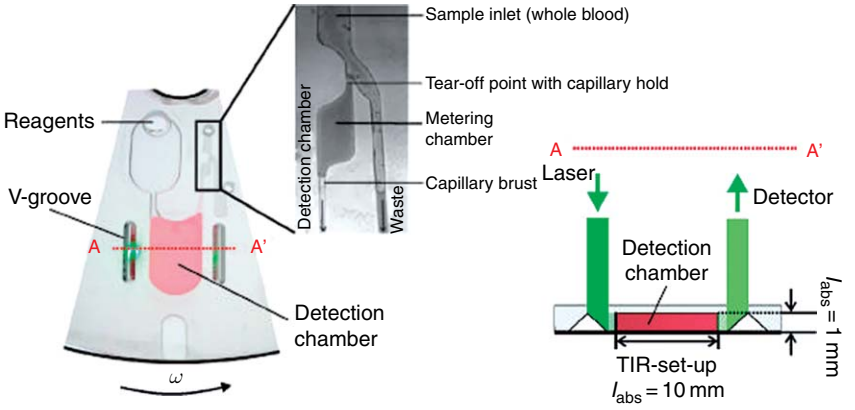


Figure 30 *Left*: Microfluidic structure for the fully integrated alcohol determination starting from a raw sample of human blood. First, the sample is metered and then mixed with the reagent within the detection chamber. *Right*: Readout is accomplished by the total inner reflection of a laser beam into the disk plane to elongate the optical path length l_{abs} from approximately 1 to 10 mm (TIR set-up). (Source: Steigert J, Grumann M, Brenner T, Riegger L, Harter J, Zengerle R, Ducrée J 2006 Fully integrated whole blood testing by real-time absorption measurement on a centrifugal platform. *Lab Chip* **6**, 1040–4. Reproduced by permission of The Royal Society of Chemistry.)

of one single macroscopic rotary engine only. In addition the centrifugal microfluidic platform can be easily applied to a wide range of different applications due to the fact that it allows scaling of the pulse-free flow rates by six orders of magnitude.

As soon as any additional actuation or sensing function is required on the module while rotating, things however become tricky from a technical point of view if a contact-free interfacing is not applicable. The platform also lacks flexibility compared to other that allow online programming of fluidic networks within one piece of hardware that fits all needs (see Sections 2.13.6.3 and 2.13.6.4). Most of the logic functions as well as their critical frequencies are imprinted into the channel network.

2.13.6 Droplet-based Microfluidic Platforms

2.13.6.1 Introduction

The principal idea behind droplet-based microfluidic systems is the use of single droplets as reaction confinements for biological assays or chemical reactions. Dominant interfacial and surface tensional forces in the microdimension enable the precise generation and spatial stabilization of these droplets. Since the droplets are kept isolated within an immiscible surrounding fluid like air or oil, lateral dispersion (Taylor dispersion) can be avoided while moving the droplets to different locations.

Droplets as reaction confinements can be regarded as nanoliter-sized batch reactors for mixing and reacting in contrast to the merely flow-through concept of the other platforms described so far. Many parallel screening reactions, each consuming only a minute amount of reagents, are enabled inside the small-sized droplets. The reproducibility of the reaction conditions is very high, since the droplet volume as well as the reaction conditions within the droplet, i.e., temperature and mixing conditions, can be controlled precisely. This is of paramount interest, especially for HTS applications in the pharmaceutical industry.

The droplet-based microfluidic systems can be fundamentally divided into two basic setups, the channel-based and the planar surface-based approach described in **Table 5**. The channel-based systems are mostly pressure-driven with the droplet generation and manipulation depending on the actuation via liquid flows within closed microchannels. On the planar surface-based platforms, droplets can be arbitrarily moved in two dimensions representing planar programmable laboratories on-chip. They are actuated by electrowetting-on-dielectric (EWOD) or surface acoustic waves (SAW).

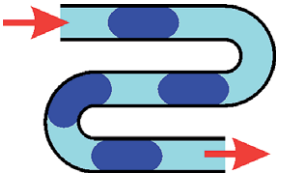
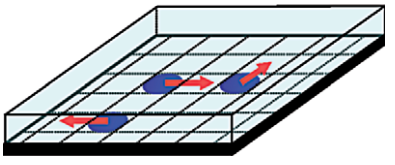
The different mechanisms to transport and manipulate droplets on the three different droplet-based microfluidic platforms, namely the pressure-driven, electrowetting-driven, and SAW-driven platform, are discussed in the following sections.

2.13.6.2 Pressure-driven Unit Operations

2.13.6.2.1 Basic set-up

The pressure-driven, droplet-based platform depends on the two-phase liquid flow through microchannels. The two immiscible phases are dispersed into each other so that a sample liquid (e.g., aqueous

Table 5 Description of the two basic set-ups for droplet-based microfluidic platforms

Approach	*Schematic sketch*	*Description*
Channel-based		Liquid flow-driven actuation (pressure-driven) Simultaneous motion in one dimension, droplet movement defined by geometry, closed microchannels, fixed droplet arrangement in channel under stable conditions possible, e.g., for droplet storage
Surface-based		Actuation by SAW or EWOD, individual motion in two dimensions, arbitrary droplet movement on surface, planar (open) surface, online control of assay protocol

solution) forms plugs of a certain length, separated by the carrier liquid (e.g., oil) along the channel. The flow scheme is called segmented flow, since the size of the inner phase droplet exceeds the cross-sectional dimensions of the channel leading to squeezed liquid plugs. The two-phase flow is pumped throughout the channels by an externally applied pressure.

2.13.6.2.2 Droplet generation and metering

The most elementary unit operation on the pressure-driven, droplet-based platform is the initial generation of the droplets. This step can also be considered as metering, since the liquid volumes involved in the latter reaction within the droplet are defined during the droplet formation process. Generally, two different microfluidic structures have been reported for a controlled droplet generation, the flow focusing structure (Anna *et al.* 2003, Joanicot and Ajdari 2005), and the T-junction (Nisisako *et al.* 2002) (**Figure 31**).

Besides these two most prominent methods, other microfluidic structures for droplet formation, mainly for emulsification in the field of microprocess engineering, have also been proposed. Vertical microchannels leading into a steady continuous phase (Kobayashi *et al.* 2005) or liquid plugs sliding down from a microterrace (Sugiura *et al.* 2001), for example, are used to generate droplets with a higher throughput. Also the production of double emulsions, e.g., water-in-oil-in-water (W/O/W), has been shown in a serial arrangement of T-junctions (Okushima *et al.* 2004) or within more complex interleaved microcapillaries arrangements (Utada *et al.* 2005).

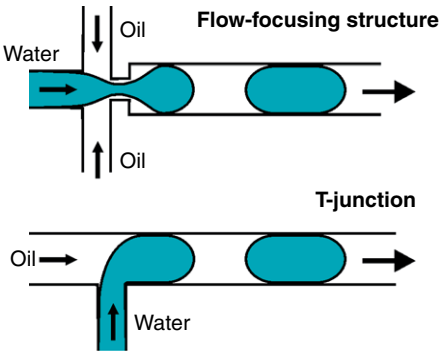


Figure 31 Both microfluidic channel arrangements, the flow-focusing structure depicted on the top and the T-junction depicted on the bottom, can be used to reproducibly initiate a droplet break-off. The sketches describe the arrangement of water as the sample and oil as the carrier phase.

The flow-focusing device and the T-junction, however, are more adequate for the injection of droplets into small microchannels inducing a segmented flow scheme. For both structures, the size of the droplet is influenced by the strength of the shear forces at the channel junction (higher shear forces lead to smaller droplets). Thus, the droplet diameter can be controlled via the flow rates and flow rate ratios of the two phases. In order to ensure a complete isolation of the single droplets, the wetting properties of the channel walls have to be assumed to be non-wetting for the sample phase. Thus, the sample phase droplets are not in contact with the walls and no leakage out of them can cause changing concentrations during a reaction.

Also the injection of gaseous bubbles into a continuous liquid stream at a flow focusing structure has been demonstrated (Garstecki *et al.* 2004, 2005b). Using these mechanisms, gas–liquid segmented flows can be generated within microchannels, which are, for example, used for the enhanced mixing of two liquid phases as described in Section 2.13.6.2.6.

2.13.6.2.3 Sample load

In order to use the droplets inside a channel as reaction confinements, the different liquid educts have to be loaded into the droplets first. A method to combine three different sample liquid streams by a sheath flow arrangement with subsequent injection as a common droplet into the carrier fluid has been shown by the group of Rustem F. Ismagilov at the University of Chicago, USA (Song *et al.* 2003b) (**Figure 32**). Different concentrations and ratios of two reagent substreams and a dilution buffer merge into one droplet and perform a so-called on-chip dilution (Song and Ismagilov 2003). The mixing ratios can be adjusted by the volume flow ratio of the three streams as depicted in the two exemplary photographs in **Figure 32**. Food dye has been used to color the reagents green and red and to visualize their portion within the generated droplet depending on the corresponding flow rates.

Using a combination of two opposing T-junctions connected to the same channel, the formation of droplets of alternating composition has been demonstrated (Zheng *et al.* 2004b). This method can be used for protein crystallization (see Section 2.13.6.2.9) or indexing within continuous flow screening experiments on the droplet-based platform. There, the first droplet of each droplet pair is used to conduct

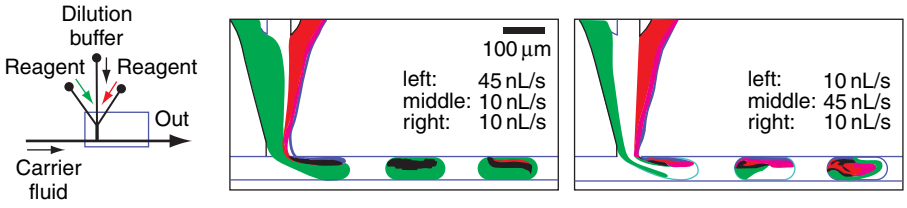


Figure 32 On-chip dilution depicted with two exemplary situations at the droplet formation T-junction. Different flow rates of the reagent streams (food dye colored green and red), separated by the transparent dilution buffer stream in the middle, are injected into the droplet. Different concentrations of the two reagents are then loaded into the droplet for reaction. (Reprinted with permission from Song H, Ismagilov R F 2003 Millisecond kinetics on a microfluidic chip using nanoliters of reagents. *J. Am. Chem. Soc.* **125**, 14613–19. Copyright 2003 American Chemical Society.)

the reaction, and the second droplet is used to index the composition of the first droplet. Using a similar technique, the injection of an additional reactant into a liquid plug moving through the channel at an additional downstream T-junction has also been demonstrated (Shestopalov *et al.* 2004).

However, not only liquid chemical reagents but also other components like cells have been loaded into droplets as reported in He *et al.* (2005). Therefore, a flow focusing device has been used with an aqueous cell suspension as the sample phase flowing through the central channel. The selective encapsulation of single cells into droplets of femtoliter volume has been demonstrated.

2.13.6.2.4 Merging and splitting of droplets

Merging of the droplets of different sizes and possessing different velocities to form single droplets has been demonstrated by Ismagilov *et al.* (Song *et al.* 2003b). In the same work, they also show the controlled splitting of droplets at a channel branching point. A constriction of the channels at the branching leads to longer and narrower droplets that could be split into two subdroplets. The size of the two split droplets was proportional to the relative flow rates in the two outlet channels. Using a similar method, the formation of droplet emulsions with controlled volume fractions and drop sizes has been demonstrated (Link *et al.* 2004).

2.13.6.2.5 Transport

The droplets are transported via a pressure-driven flow through the microchannels. Since the droplets typically are bigger in size than the channel cross section, they form a plug within the channel, segmenting the continuous phase (segmented flow). This keeps the phase configuration along the channel stable since no liquid is exchanged between the plug interspaces. This way, plugs can be moved spatially separated through a channel network.

This has been used, e.g., for a polymerization process. Liquid monomer droplets have been hardened by exposing UV light to the droplets- carrying microchannel. Using this method, nonspherical particles can also be generated by polymerizing within a microchannel of small cross section wherein the droplets are squeezed and thus become disk shaped (Dendukuri *et al.* 2005).

2.13.6.2.6 Mixing

The starting point of the mixing process is set to the time when the droplet is generated and released into the pressure-driven segmented flow (**Figure 32**) or when two droplets are merged. In the first case a buffer liquid separates the two reagent flows prior to the injection into a droplet. The two reagents are then mixed by a recirculating flow inside the droplets due to shear forces induced by the motion along the stationary channel wall (Tice *et al.* 2003). This effect is even more pronounced if two liquids of differing viscosities (2.0 and 18 mPa s) are mixed within the droplet as reported in Tice *et al.* (2004).

Based on the recirculation flow, a mixing scheme for the pressure-driven, droplet-based platform has been proposed using winding microchannels as depicted in **Figure 33** (Song *et al.* 2003a). Within each 90° channel curvature the orientation between the phase pattern in the droplet and the direction of motion is changed so that the inner recirculation leads to stretching and folding of the phases. Therewith, the number of phase lamellae within the droplet continuously increases by every bend along the channel. Under favorable conditions, submillisecond mixing can be achieved and has been employed for the multistep synthesis of

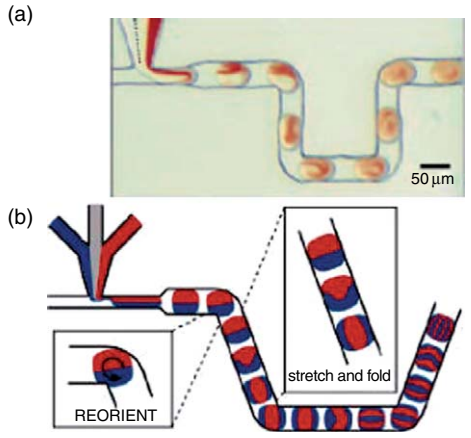


Figure 33 Mixing within liquid droplets moving through bending microfluidic channels shown (a) experimentally and (b) schematically. (Source: Song H, Bringer M R, Tice J D, Gerdts C J, Ismagilov R F 2003a Experimental test of scaling of mixing by chaotic advection in droplets moving through microfluidic channels. *Appl. Phys. Lett.* **83**, 4664–6. Reproduced by permission of American Institute of Physics.)

nanoparticles, for example (Shestopalov *et al.* 2004). A detailed and theoretical description of this mixing effect is given in Bringer *et al.* (2004).

Besides the mixing within liquid droplets, dispersed into another liquid carrier phase, mixing within the carrier phase can also be accelerated by a segmented flow. The injection of gas bubbles into a continuous liquid stream forming a segmented gas–liquid flow has been described by Klavs Jensen and his group at the MIT (Gunther *et al.* 2004, 2005). The gas bubbles are introduced into the liquid flow and initiate recirculation flows within the liquid segments in between due to the motion along the channel wall. The gas bubbles can be completely separated from the liquid stream using a planar capillary separator after the reaction is completed. Using this technology, the synthesis of colloidal silica particles has been demonstrated (Khan *et al.* 2004). Compared to a pressure-driven flow without segmentation, the axial dispersion is eliminated in the proposed gas–liquid flow scheme. Therefore, the residence time distribution of the reactants is also narrowed leading to less polydisperse particles.

Another microfluidic mixing scheme based on a gas–liquid segmented flow uses an additional repeated separation and recombining of the channel (Garstecki *et al.* 2005a). Homogenization of two aqueous streams has been demonstrated after 10 branching units for various flow rates (Reynolds numbers between 0.01 and 100) using this mixing scheme.

2.13.6.2.7 Incubation

The incubation time of the reagents combined inside a droplet at the injection position can easily be calculated at a certain point of observation from the traveling distance of the droplet divided by the droplet velocity. A unique feature of the platform is that a time-resolved monitoring of the incubation can easily be done by scanning along the channel from the injection point to more downstream positions. Thereby the kinetics of chemical reactions in the order of some milliseconds can easily be investigated and have been reported. On the other hand, incubation times on the order of a week for storing applications have been demonstrated. This is enabled by the droplet compartments that are separated by the carrier fluid, which prevents evaporation and diffusion. Using this approach, several 60 nl liquid droplets containing one or a few cells were generated within a microfluidic chip and were then flushed into a Teflon® capillary tube for cultivation. The cell densities were still as high as in conventional systems after 144 h of growth within the droplets (Martin *et al.* 2003).

2.13.6.2.8 Switching

Switching on the droplet-based and pressure-driven platform means the controlled routing of single droplets. All liquid routing technologies, available on the pressure-driven platform described in Section 2.13.3.2.5, can be used to switch droplets in the channel-type droplet-based microfluidic platform. If certain droplets out of a train of droplets should be separated, for example, an X-shaped channel crossing can be used. As soon as the droplet of interest passes the junction within the main flow, a short pressure pulse is applied at a side channel to direct the droplet into one of the two remaining outflow channels.

2.13.6.2.9 Application example: protein crystallization

Protein crystallization (see Section 2.13.4.3.1 for description) has been implemented on the pressure-driven, droplet-based platform by Ismagilov and coworkers (Zheng *et al.* 2003, 2004c). The basic principle is depicted in **Figure 34(a)**: droplets of three liquids namely the protein solution, the buffer, and the precipitant are dispersed into the oil carrier phase. The precipitant concentration inside the

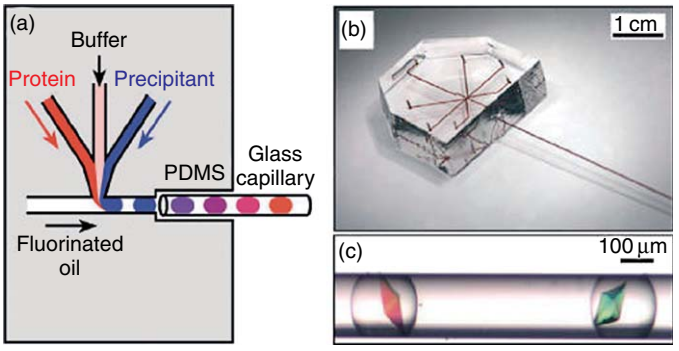


Figure 34 (a) Schematic illustration of the protein crystallization device and (b) the complete set-up consisting of a polydimethylsiloxane (PDMS) chip and a glass capillary. (c) Examples of thaumatin crystals that were grown inside droplets in a capillary. (Source: Zheng B, Tice J D, Roach L S, Ismagilov R F 2004c A droplet-based, composite PDMS/glass capillary microfluidic system for evaluating protein crystallization conditions by microbatch and vapor-diffusion methods with on-chip X-ray diffraction. *Angew. Chem. Int. Ed.* **43**, 2508–11. Reproduced by permission of Wiley VCH.)

droplet is adjusted via the buffer and precipitant flow rate. Therewith, different concentrations are generated and transferred into a glass capillary for later detection. Nonspecific protein adsorption onto the liquid–liquid interface can be suppressed by adding certain surfactants to the carrier phase (Roach *et al.* 2005). As depicted in **Figure 34(b)**, the droplets containing the protein crystals can be directly forwarded into a glass capillary ready for the X-ray analysis (Yadav *et al.* 2005).

As described earlier in this chapter, successful protein crystallization is initiated for a certain precipitation concentration only. Therefore, the conduction of many crystallization experiments within a short time and with a reduced amount of protein sample is favorable and is enabled by the described sample loading technique. Examples of crystallized thaumatin proteins are depicted in **Figure 34(c)**. The effect of mixing on the nucleation of protein crystallization has been investigated by combining the described crystallization structure with a winding mixing channel as described in Section 2.13.6.2.6 (Chen *et al.* 2005). Fast chaotic mixing has been found to be favorable for the formation of well-formed proteins within the droplets (Zheng *et al.* 2005).

Besides the described method for crystallization, an alternative process in which the concentration within one droplet is changed over time has also been developed by the group of Ismagilov, based on water diffusion between droplets (Zheng *et al.* 2004a, c). In this case, a carrier liquid that enables diffusion between the droplets, that is, a water-permeable liquid, is selected. Alternating droplets of protein and precipitant on the one hand and a high-concentration salt solution on the other hand are generated using two opposing droplet generation structures (Zheng *et al.* 2004b). Water diffuses through the oil carrier phase from the low salt concentration, i.e., from the protein-containing droplet to the high-concentration droplet. This steadily increases the concentration within the protein-containing droplet until the suitable crystallization condition is achieved.

Recent developments on the pressure-driven, droplet-based platform aim at HTS applications. Therefore a large number of droplets, each containing a different reagent, are separated and surrounded by a fluorinated carrier fluid within a microcapillary. In order to prevent coalescence during possibly long storage times, a gas bubble is injected between the droplets as an additional separation phase (Zheng and Ismagilov 2005). Based on this three-phase liquid–liquid–gas system, reliable and HTS assays can be performed, which could be an alternative for well plates in the future (Chen and Ismagilov 2006).

2.13.6.3 Electrowetting-driven Unit Operations

2.13.6.3.1 Basic set-up

The electrowetting effect was first described by Gabriel Lippmann in 1875 while the recent developments were initiated in the early 1990s by introducing the idea of using thin insulating layers to separate the conductive liquids from the metallic electrodes in order to eliminate the problem of electrolysis (taken from Mugele and Baret (2005) on electrowetting). This paved the way for the

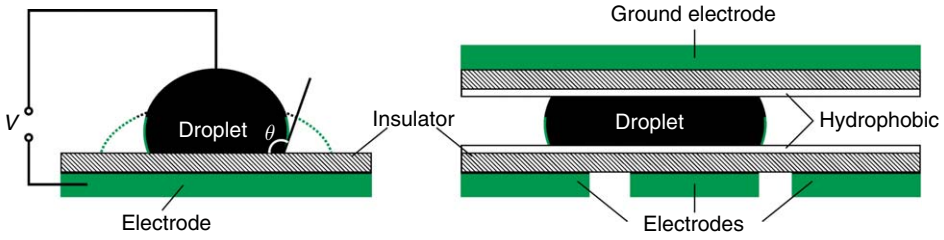


Figure 35 *Left*: The electrowetting principle. If a voltage V is applied between a conductive liquid and an electrode separated by an insulating layer, the contact angle θ of the liquid–solid interface can be decreased. *Right*: Schematic diagram of the electrowetting actuation of liquid droplets. Several individual addressable control electrodes (bottom) are used to locally change the contact angle and thereby initiate a droplet movement.

application of the electrowetting effect as a liquid propulsion principle for lab-on-a-chip systems.

The principal setup of an EWOD (Lee *et al.* 2002) setup is depicted in **Figure 35** (*left*) with a liquid droplet on top of an insulating dielectric layer. A voltage V can be applied between the conductive liquid droplet and the electrode leading to a change in the contact angle, θ

$$\gamma_{lg}\cos\theta = \gamma_{sg} - \gamma_{sl} + \frac{1}{2}\frac{\varepsilon_0\varepsilon_r}{d}V^2 \qquad [6]$$

depending on the interfacial tension at the liquid–gas interface γ_{lg}, solid–gas interface γ_{sg}, and solid–liquid interface γ_{sl}. The thickness of the insulator d and the dielectric parameters ε_0 and ε_r define the impact of a voltage change on the contact angle. So EWOD can be simply described as a tool to control the contact angle of conductive liquids (Mugele *et al.* 2005).

Using the EWOD set-up, a microfluidic actuation method for moving droplets between two parallel electrodes has been published by Pollack *et al.* (2000) from the Duke University in Durham and by Chang-Jin (CJ) Kim from the University of California, Los Angeles, CA (UCLA) (Lee and Kim 2000). A schematic cross-section of an EWOD actuator is depicted in **Figure 35** (*right*). Several individual addressable control electrodes are located on the bottom of the device to control the droplet path. They are typically arranged in 2D arrays. An additional hydrophobic layer (mostly Teflon) is added to the insulator surface to enhance the droplet movement. A common ground electrode in contact with the liquid droplet replaces the extended electrode.

The droplet, which is enclosed between the two electrode plates, features a certain volume so that it covers parts of two addressable electrodes at all times. If a voltage is applied on one of the control electrodes covered by the droplet, the contact angle is reduced at this part of the droplet. This initiates droplet movement along the paths given by the activated pads. Because the path of the droplets is determined by the pattern of electric potentials, the EWOD-driven, droplet-based platform is easily programmable. This allows different assays to be run by different programs on the same piece of hardware.

Besides the transport of aqueous solutions, several other liquids like organic solvents, ionic liquids, aqueous surfactants solutions (Chatterjee *et al.* 2006), and also biological fluids like whole blood, serum, plasma, urine, saliva, sweat, and tears (Srinivasan *et al.* 2004) have also been successfully transported on the EWOD droplet-based platform.

An alternative actuation principle for surface-based droplet manipulation that does not require a second electrode in contact with the droplet is dielectrophoresis. The basic unit operations, like droplet generation, movement, metering, and merging, using this alternative actuation have been demonstrated recently (Gascoyne *et al.* 2004, Schwartz *et al.* 2004). However, they are not discussed in detail in the following sections that focus on EWOD-based unit operations only.

2.13.6.3.2 Metering

The dispensing, i.e., initial metering unit operation, is probably the most critical step on the EWOD-driven, droplet-based platform. Although droplets may be formed by simply pulling liquid out of a reservoir, the location of the tear-off can hardly be predicted. In the method depicted in **Figure 36**, metered droplets are formed from an on-chip reservoir in three steps (Srinivasan *et al.* 2004). First, a liquid column is extruded from the reservoir by activating a series of electrodes adjacent to it. Second, once the column overlaps the electrode on which the droplet is to be formed, all the remaining electrodes are turned off to

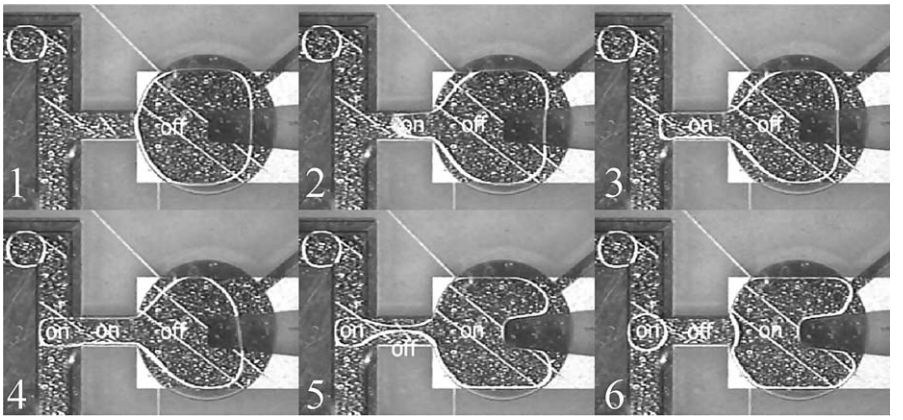


Figure 36 Droplet formation from an on-chip reservoir using electrowetting forces only. First, a liquid column is extruded from the reservoir and tears off after the connecting electrode is turned off. (Source: Srinivasan V, Pamula V K, Fair R B 2004 An integrated digital microfluidic lab-on-a-chip for clinical diagnostics on human physiological fluids. *Lab Chip* **4**, 310–15. Reproduced by permission of The Royal society of Chemistry.)

form a neck in the column. The reservoir electrode is then activated in the third and last step to pull back the liquid and break the neck completely to form a droplet. Using this droplet metering structure, droplets of 20 nl volume can be generated with a standard variation below 2% (Srinivasan *et al.* 2004).

Since the droplet dispensing is a crucial step for the performance and the accuracy of all assays on the EWOD platform, additional measures for a controlled liquid metering such as on-chip capacitance metering for volume control (Ren *et al.* 2004) or the use of numerical methods for the design of EWOD structures (Berthier *et al.* 2006) have been proposed.

2.13.6.3.3 Mixing

The most basic of miscible liquids within droplets on the EWOD platform is by oscillating the droplet between two electrodes. Before this active mixing scheme is used, the two droplets containing the liquids to be mixed have to be merged into a single droplet. This coalesced droplet is then moved along the electrodes in an oscillating fashion to induce advectional effects inside the droplet. An increasing frequency of droplet movement leads to reduced mixing times. The mixing process can be further accelerated by oscillating over a longer linear electrode array. The shortest mixing time for two 1.3-µl droplets in linear oscillation on four electrodes was about 4.6 s (Paik *et al.* 2003b). In another work, the mixing time was further reduced to less than 3 s using 2D arrays (Paik *et al.* 2003a).

2.13.6.3.4 Merging and splitting of droplets

Together with droplet generation from a reservoir and the droplet transport along electrode arrays, the controlled merging and splitting of droplets complete the four fundamental fluidic operations considered essential to build digital microfluidic circuits for lab-on-a-chip applications (Cho *et al.* 2003).

The realization of droplet splitting and merging on three linearly aligned EWOD electrodes is depicted in **Figure 37** (Pollack *et al.* 2002). The initial droplet volume is first elongated toward the middle electrode by applying a voltage on it (**Figure 37(b)**). After a certain time delay, required for the stabilization of the droplet, the voltage is switched from the middle to the right electrode, initiating a division of the droplet into two subdroplets placed on the left and the right electrode, as depicted in **Figure 37(c)** and **37(d)**. For a subsequent merging of the droplets, the voltage is again switched from the right to the middle electrode, moving the right subdroplet back again to the middle position and bringing it in contact with the left subdroplet (**Figure 37(e)**–**(f)**).

2.13.6.3.5 Readout

Different readout schemes for biochemical assays have been applied to the EWOD-driven, droplet-based platform. Colorimetric, enzymatic assays that are important for diagnostic applications have been successfully implemented and glucose concentration measurements on several biological fluids (serum, plasma, urine, and saliva) with good comparable results have been demonstrated (Srinivasan *et al.* 2004).

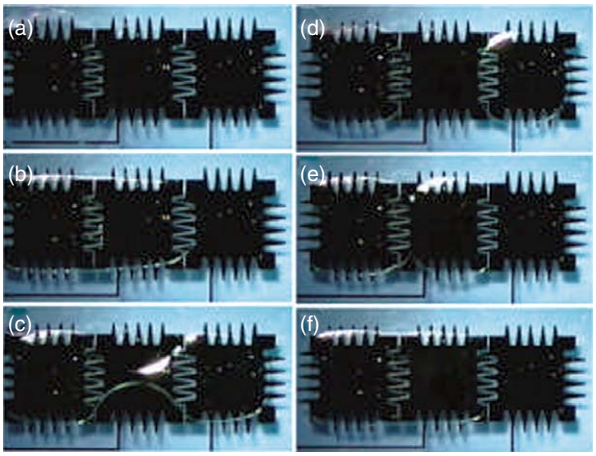


Figure 37 A sequence of photographs showing the droplet splitting and merging. (a) Initially, only the left electrode is activated. (b) The middle electrode is then activated and, (c) and (d) after a delay, the voltage is switched from the middle to the right electrode resulting in division of the droplet. (e) and (f) The original droplet is reassembled by switching the voltage on the right electrode back to the middle electrode. (Source: Pollack M G, Shenderov A D, Fair R B 2002 Electrowetting-based actuation of droplets for integrated microfluidics. *Lab Chip* **2**, 96–101. Reproduced by permission of The Royal Society of Chemistry.)

Another example with regard to the use of an EOWD system for the automated sample preparation of peptides and proteins for matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) is reported (Wheeler *et al.* 2004). In this work, standard MALDI-MS reagents, analytes, concentrations, and recipes have been demonstrated to be compatible with this technique, and mass spectra comparable to those collected by conventional methods were obtained.

2.13.6.4 SAW-driven Unit Operations

2.13.6.4.1 Basic set-up

An alternative to the electrowetting-based actuation of droplets on a plane surface has been proposed by Achim Wixforth and his group at the University of Augsburg, Germany (Wixforth 2003). The approach is based on SAW, which are mechanical waves with amplitudes of typically only a few nanometers. In contrast to ordinary sound waves, which are of longitudinal type, SAW are transversal waves. Their amplitude can be either parallel to the surface (Love waves) or perpendicular to the surface (Rayleigh waves) or of mixed type. For the manipulation of droplets on the SAW-based platform, typically Rayleigh waves are used.

The SAW are generated by a piezoelectric transducer chip (e.g., quartz), which can be fabricated by placing interdigital electrodes (interdigital transducer, IDT) on top of a piezoelectric layer, for example. Liquid droplets are placed on the hydrophobic surface of the chip and can be moved by the SAW on the surface. Appropriate AC signals are applied to the electrodes only at a certain region of the piezoelectric layer. Thus the SAW are exclusively generated at the desired location and can be focused on the droplet. If the acoustic pressure exerted on the liquid droplet is high enough, the droplet can be deformed or even be moved as depicted in **Figure 38** (Wixforth *et al.* 2004).

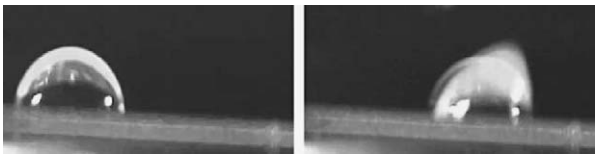


Figure 38 Side view of an approximately 50-nl droplet on the hydrophobic surface (contact angle of about 90°) of a piezoelectric substrate. *Left*: Resting droplet. *Right*: The droplet is hit by SAW impinging from the left resulting in a momentary asymmetry of the wetting angles of the droplet. (Source: Wixforth A, Strobl C, Gauer C, Toegl A, Scriba J, von Guttenberg Z 2004 Acoustic manipulation of small droplets. *Anal. Bioanal. Chem.* **379**, 982–91. With kind permission of Springer Science and Business Media.)

This approach is also sometimes referred to as flat fluidics, because no cover or slit is required as in the EOWD approach but only a flat surface where the SAW are generated is needed. Today, this technology is being pushed forward by the company Advalytix AG (2006), Germany.

2.13.6.4.2 Metering

Metering is accomplished by a combination of locally changed wetting properties of the surface and the actuation via SAW. One possible metering process is depicted in **Figure 39**. An initial water plug of approximately 100 nl is placed on an oval hydrophilic zone on the chip surface (**Figure 39(a)**). After the first actuator (IDT) is activated, SAW propagate horizontally toward the liquid plug, pushing it to the right across another small hydrophilic metering spot (**Figure 39(b)**). After the actuation, the liquid volume withdraws to the hydrophilic regions on the chip, leaving behind a small metered liquid portion (**Figure 39(c)**). This droplet is subsequently moved upward by another IDT for

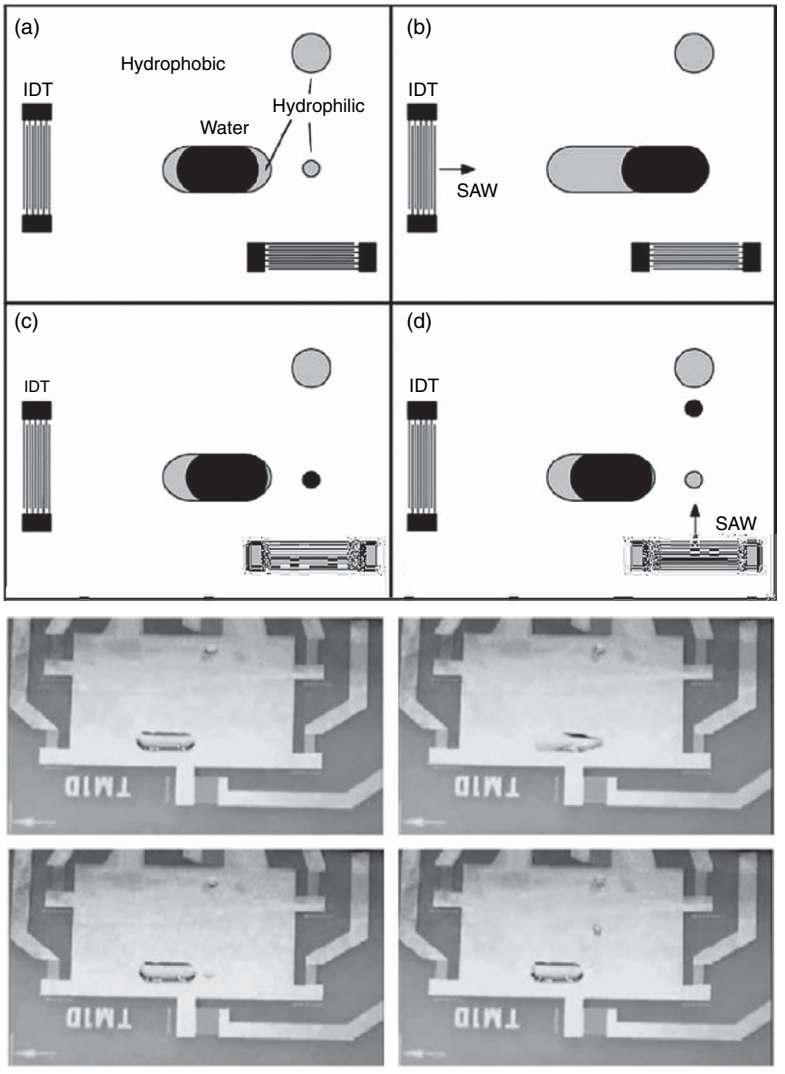


Figure 39 Acoustically driven metering process on the SAW platform. (a)–(d) Hydrophobic and hydrophilic areas are indicated in the schematic sketch on the top. First, the horizontally acting interdigital transducer (IDT) pushes the water droplet to the right across the hydrophobic substrate. After the actuation, the droplet withdraws meanwhile separating a small liquid volume, which remains at the small hydrophilic spot, as depicted in (c). In the last step, this metered liquid portion is pushed upward toward another hydrophilic zone. Corresponding photographs of the actual experiment are also depicted. (Source: Wixforth A, Strobl C, Gauer C, Toegl A, Scriba J, von Guttenberg Z 2004 Acoustic manipulation of small droplets. *Anal. Bioanal. Chem.* **379**, 982–91. With kind permission of Springer Science and Business Media.)

further processing (**Figure 39(d)**). At the bottom of **Figure 39**, a sequence of experimental photographs shows the metering process on the open chip surface.

2.13.6.4.3 Mixing

Mixing presents a system-immanent unit operation on the SAW-driven, droplet-based platform. A droplet, which is placed on the substrate and is hit by SAW, experiences an internal streaming due to the vibrating forces of the wave. If the amplitude is not large enough for a droplet movement, the liquid inside the droplet is efficiently stirred while its position on the substrate is retained. The internal streaming is depicted in **Figure 40** where a fluorescent dye is dissolved into a 50-nl droplet under SAW agitation.

2.13.6.4.4 Merging and splitting of droplets

Using the SAW-based mechanism for droplet movement in combination with a certain pattern of different wetting areas on the chip surface, droplet splitting as described in Section 2.13.6.4.2 as well as droplet merging can be accomplished.

The splitting of one droplet into several smaller volumes of the same size, i.e., aliquotting, is realized by moving the initial droplet over a hydrophobic/hydrophilic chessboard zone. Since the transport of a droplet depends nonlinearly on the droplet size, the initial droplet of bigger volume is constantly moved forward, while small, picoliter-sized droplets remain on the hydrophilic parts of the wetting pattern (Wixforth 2003).

2.13.6.4.5 Incubation and entrapment

For some assay protocols, longer incubation steps at elevated temperatures are also required, e.g., for a PCR amplification. The liquid plug is placed above a microheating element on the substrate surface for these incubation steps. However, since the nanoliter-sized droplet possesses a high surface-to-volume ratio, the liquid volume decreases rapidly due to evaporation. Hence to avoid evaporation, the aqueous liquid plug is covered with an oil plug having a smaller contact angle. This droplet in droplet configuration still can be moved via SAW on the substrate surface. Using this technology, a PCR assay within a 200-nl droplet enclosed in mineral oil was performed with an online monitoring of the DNA concentrations and provided a sensitivity of 0.1 ng (Guttenberg *et al.* 2005).

2.13.6.4.6 Readout

The fabrication of the electrodes for the actuators (IDT) as well as the fabrication of the wetting patterns on the substrate is done by high-precision lithographic processes. Therefore, heaters, electrodes, or different types of sensors based on surface micromachining can easily be integrated in a straightforward way. Also, optical readout schemes have been realized for the online fluorescent monitoring of the RT-PCR progress, for example (Guttenberg *et al.* 2005).

2.13.6.5 Strengths and Challenges of the Platform

General advantages of droplet-based microfluidics are the small liquid volumes of the droplets reducing reagent and sample consumption and thus paving the way for HTS applications. Additionally, the batch-mode operation scheme used in the nanoliter- to microliter-sized droplets represents a consistent further development of the classic assay protocols in, for example, well plates.

The pressure-driven approach combines these advantages with the high-throughput capabilities in a quasi-continuous operational scheme. The completely enclosed liquid droplets furthermore allow the incubation and storage of liquid assay

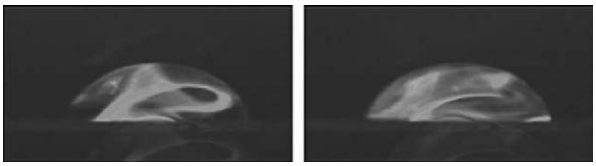


Figure 40 Internal streaming is induced within an approximately 50-nl droplet, hit by SAW. The internal mixing process is visualized by a fluorescent dye placed on the surface of the chip, which is dissolved by the SAW agitation. (Source: Wixforth A, Strobl C, Gauer C, Toegl A, Scriba J, von Guttenberg Z 2004 Acoustic manipulation of small droplets. *Anal. Bioanal. Chem.* **379**, 982–91. With kind permission of Springer Science and Business Media.)

results over a long period of time without evaporation. However, the microfluidic functionality is engraved by the channel design and cannot be adopted during an assay, for example.

In contrast, the surface-based actuation schemes (EWOD and SAW) come up with a high flexibility since liquid processing paths can be freely programmed. In addition, the simple setup without any moving parts can be fabricated very cost-efficiently using standard lithographic processes.

When comparing the EWOD and SAW principle for droplet actuation on the planar surface, the electrical change of the contact angle depends on the liquid properties and can cause electrolysis while the SAW principle allows easier adoption of the liquid properties. Evaporation of liquid and the long-term stability of the hydrophobic and hydrophilic surface coatings are the major drawbacks of the surface-based techniques. Last but not least, all planar droplet-based platforms do not really work properly in nonplanar positions and therefore, for example, a handheld operation is not recommendable.

2.13.7 Free Scalable Noncontact Dispensing

2.13.7.1 Introduction and Motivation

The free scalable noncontact dispensing platform allows to deliver liquids as free-flying droplets onto planar substrates (e.g., microarrays), conventional containers such as well plates, or any other target. This approach is closest to the traditional lab routine, which is based on conducting assays via successive pipetting steps, manually or by automated lab equipment like robots. In these fields, the dispensing of droplets of different volumes (picoliter to milliliter range), from a single channel or up to thousands of channels in parallel, with different pitch sizes and individual controllability is required. An overview of the different fields of operation regarding the droplet volume on the one hand and the number of different liquids to be dispensed at once on the other hand is depicted in **Figure 41**. Here we focus on the simultaneous dispensing of a large number of different reagents in parallel and will not focus on noncontact printing technology in general (for an overview of ink-jet printing technologies, see Le (1998)). All the three highlighted operating principles are based on one microfluidic platform as described in Section 2.13.7.2.

2.13.7.2 Unit Operations

2.13.7.2.1 Metered dispensing

One functional unit of the free scalable noncontact liquid dispensing platform is based on the combination of a reservoir for holding the liquid, a nozzle chamber with a nozzle from which the liquid is dispensed, and a capillary channel connecting reservoir and nozzle chamber. Depending on the arrangement of these components as well as the actuation principle, liquid volumes from several tens of picoliters to several microliters can be dispensed. Arranging several units on a so-called dosage chip enables the handling of up to thousands of different liquids in

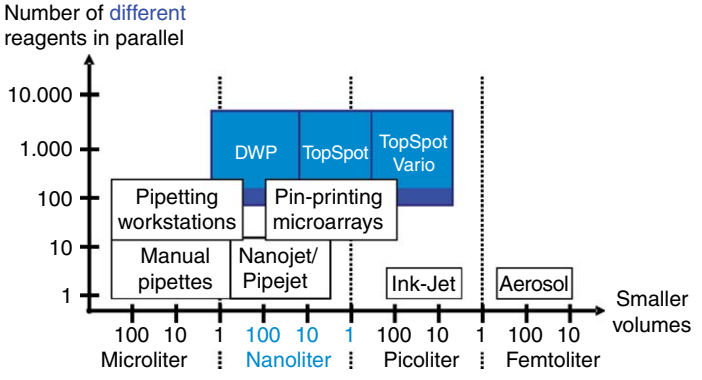


Figure 41 Overview of low volume dispensing. The presented platform for noncontact dispensing depends on three different techniques to dispense well-defined liquid volumes. These techniques are named according to the pilot application they are used for (dispensing well plate (DWP), TopSpot®, TopSpot® Vario). The platform covers the volume range from several tens of picoliters (pl) to several microliters (μl). Several thousands of dispensing units can principally be arranged in parallel to handle the same number of different liquids, e.g., different biological reagents in parallel.

parallel and has been demonstrated at a pitch ranging from several hundreds of micrometers to several millimeters. Three different actuation schemes based on the same geometrical arrangement, namely the dispensing well plate (DWP) (Koltay *et al.* 2004a, b), the TopSpot® (de Heij *et al.* 2003, 2004), and the TopSpot® Vario (Steinert *et al.* 2004) technology are depicted in **Figure 42**.

The noncontact dispensing process, strictly speaking, represents two unit operations, liquid metering on the one hand and liquid transfer on the other hand. The metered liquid volume is delivered as a free-flying droplet or a jet in a noncontact manner to any substrate, receiving vessel, or the reservoir of another dispensing unit. The volume of that liquid portion is determined by the nozzle geometry and the external actuation mechanism.

The DWP principle is based on the complete drainage of the liquid volume within the nozzle chamber. The micromachined nozzle chamber can be considered as a metering structure, which is filled with liquid from the reservoir by capillary forces via the capillary channel between two dispensing events. The total liquid present in the nozzle chamber is dispensed by applying a pneumatic pressure of 30–80 kPa for 3–10 ms. Since this pressure pulse is applied on the liquid–air interface of the nozzle chamber and the reservoir, no pressure gradient evolves along the capillary channel and thus no backflow of liquid from the nozzle chamber to the reservoir is observed. The dosed volume is hardly affected by the liquid properties like viscosity, density, and surface tension, but is affected only by the geometry of the nozzle chamber, making this dispensing method very robust.

The TopSpot principle also depends on a pneumatic pressure pulse, which in this case is on the submillisecond timescale. Such a short pneumatic pressure pulse can be generated only by compression of an enclosed gas volume. Therefore an assembly of a piezo stack actuator driving a piston into a closed actuation cavity above the nozzle chambers is used. The pressure pulse acts equally on all nozzle chambers within the pressurized actuation cavity and causes the simultaneous ejection of single droplets out of each nozzle. The droplet volume is typically on the order of 1 nl for a 50-µm nozzle, which is in contrast to the DWP principle just a small fraction of the nozzle chamber volume. The exact droplet volume depends on the liquid properties, the actuation parameters, and the nozzle dimensions (Gutmann *et al.* 2004b).

In contrast to the pneumatic technologies described so far, the TopSpot Vario principle uses the direct displacement of an incompressible but easily deformable elastomer for actuation. The elastomer inlay replaces the air volume in the set-up (see **Figure 42**, *right*) and is displaced into the nozzle chambers by the piston movement if the piezo stack is actuated. A well-defined volume of liquid in the nozzle chamber is displaced by the elastomer and a droplet of the corresponding volume is ejected out of the nozzle. This direct displacement principle allows the independent control over the droplet volume and the droplet speed by adjusting the stroke and the speed of the piezo actuator. The tunable volume range of the droplets is from 100 pl up to 1400 pl (1.4 nl) for a 50-µm nozzle and can easily be adopted by varying the control voltage of the piezo actuator.

The basic structures like reservoirs, capillary channel, nozzle chamber, and nozzles can be fabricated by different technologies using different materials. Dry etching of silicon (deep reactive ion etching, DRIE) (Steinert *et al.* 2004), lithographic fabrication in SU-8 (Bohl *et al.* 2005), and

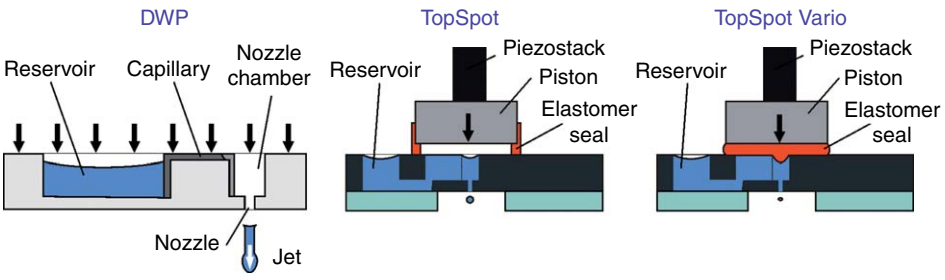


Figure 42 Dispensing well plate (DWP): Pressure-based actuation for dispensing from 10 nl up to several microliters; TopSpot®: pressure-based actuation for dispensing volumes in the lower nanoliter range; TopSpot® Vario: direct displacement principle via an elastomer for dispensing volumes from 50 to 1.000 pl.

micromilling of plastics (Steger *et al.* 2004) have been demonstrated. A high precision in the geometry of the dispensing units is needed for a high precision of the dispensed liquid volumes, and this results in the need for a high precision of the fabrication processes.

One unique feature of all the three dispensing principles described is the possibility to arrange many of them in parallel with a free scalable pitch of the nozzle chambers and the reservoirs. The capillary channel that connects these two substructures accomplishes the format conversion from a reservoir pitch of a few millimeters (enabling the filling of the dispensing chip using standard pipetting robots) to the pitch of the nozzle of a few hundred micrometers. This is important for the fabrication of microarrays, which is one application example for the platform, described in more detail in Section 2.13.7.3.1.

2.13.7.2.2 Incubation

Besides the noncontact dispensing of a metered liquid volume, incubation steps at defined temperatures are also important for conducting biological assays. This can be realized on the free scalable noncontact dispensing platform by a reversible sealing of the structure openings, namely the nozzle, nozzle chamber, and the reservoirs preventing any liquid flow through the capillaries. The reservoirs and the nozzle chambers act as standard reaction cavities, then enabling incubation steps as long as the sealing is not removed (**Figure 43**, *left*).

2.13.7.2.3 Amplification

Similar to the incubation described in Section 2.13.7.2.2, also temperature-initiated amplification using the polymerase chain reaction (PCR) method can be conducted within the cavities of the platform (**Figure 43**). Therefore, the sealed dispensing chip is mounted between an upper and a lower heating plate, which perform a certain temperature cycle between three temperatures (94°C, 53°C, and 72°C). This process is similar to the common PCR cycling within standard well plates. However, for the microcavities within the PCR slides the upper heating plate also has to change the temperature according to the cycling sequence in order to avoid temperature gradients within the PCR solution. Within the microwell plate, the upper heat plate can be fixed at a high temperature (to prevent condensation) since the air volume between the liquid and the heat plate acts as a thermal insulator. When the cycling is finished the PCR product (amplified DNA) can be dispensed in nanoliter portions into a microliter plate for further processing, onto a microarray for detection, or into another dispensing chip.

2.13.7.3 Application Examples

2.13.7.3.1 TopSpot for microarray spotting

A DNA microarray is an ensemble of microscopic DNA spots attached to a flat solid surface forming an array of different well-known capture molecules (probes) at well-defined positions. It can therefore be considered as a highly parallel biosensor, based on the lock and key principle. The probes on the surface react with a complex mixture of molecules (sample) during the hybridization phase of a microarray experiment. The sample molecules are equipped with fluorescent markers for later detection in a fluorescent readout device. The position of a positive fluorescent signal provides information about the identity of the sample (**Figure 44**).

The engineering challenge in the microarray fabrication is to immobilize up to 1000 different analytes as spots of approximately 1 nl volume, with a pitch of typically 500 µm and below, at high quality, high throughput, and low costs onto a substrate. One of the technical solutions to this problem is based on the TopSpot technology, a first application of the free scalable noncontact dispensing platform (Gutmann

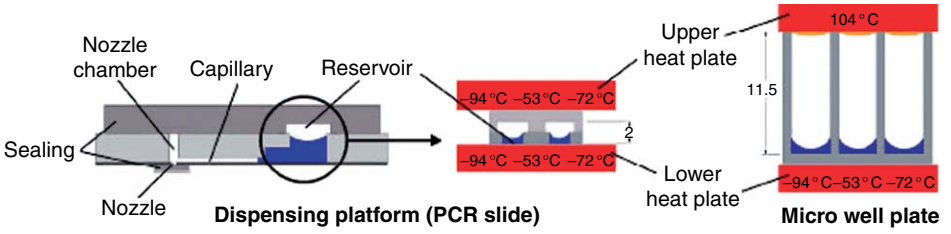


Figure 43 The nozzle and the reservoir opening are sealed for incubation and amplification. Thermal cycling within the thin substrate (polymerase chain reaction (PCR) slide) requires an equal temperature for the upper and the lower plate at any time. Otherwise, inhomogeneous liquid temperatures can influence the assay performance.

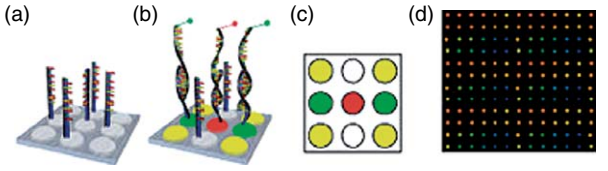


Figure 44 Illustration of the working principle of a DNA microarray. (a) Probe molecules are immobilized as different spots with a certain pitch on the solid surface; (b) a binding reaction takes place between the probe and the sample molecule if the chemical structures are complementary; (c) the fluorescent readout after the microarray experiment indicates positive binding events; and (d) fluorescent image from a real microarray experiment.

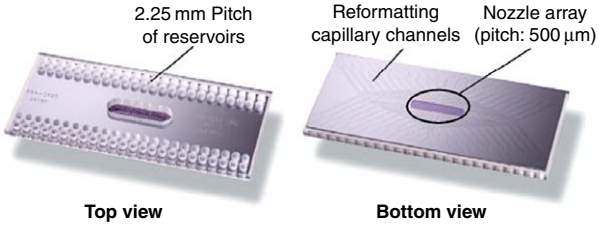


Figure 45 Picture of a 96 TopSpot® printhead. The top view shows the reservoirs that are filled via manual pipetting or lab robots. The capillary channels that connect each reservoir with a certain nozzle chamber ending up in an array of nozzles of pitch 500 μm, a feature called reformatting, can be seen on the underside of the printhead (*right*).

et al. 2004a, c). A dispensing chip or so-called printhead featuring 96 parallel dispensing channels is depicted in **Figure 45**.

The key advantage of the noncontact dispensing platform for microarray fabrication is the ease to perform passive format conversion in the system. The reservoirs on top of the printhead are arranged at a pitch of 2.25 mm corresponding to the pitch of 384 well plates enabling the filling with standard lab equipment, e.g., pipetting robots. Several microliters of liquids can be loaded into each of the 96 reservoirs, which is enough for several thousand dispensing events without the need for refilling. Each reservoir is connected to a certain nozzle chamber in the middle of the printhead via a capillary channel. The liquids are simply transported to the nozzle chamber by capillary forces and stop at the nozzle until a pressure pulse is applied. The nozzles are arranged in an array of 500-μm pitch on the underside of the printhead defining the later spot positions on the microarray.

The typical CV of spot diameters on the microarray is found to be <1% for a single dispensing unit and <1.5% between all nozzles of a printhead for all relevant printing buffers used (Gutmann *et al.* 2004a). Using the TopSpot technology, protein microarrays as well as living cell microarrays have also been fabricated (Gutmann *et al.* 2005).

2.13.7.3.2 Dispensing well plate

Well plates are the standard assay format in the life science industry, especially in HTS instrumentations for drug discovery. The assays are conducted by transferring different reagents between different containments called wells using pipetting robots. Different well plate standards such as the 96, 384, and 1536 format exist, differing in the pitch of the reservoirs (9.0, 4.5, and 2.25 mm, respectively) but all having the same outer dimensions (8 cm × 12 cm) according to the SBS standard. The trend toward smaller wells and consequently assay volumes also provokes the need for liquid handling systems being able to transfer small liquid volumes in the submicroliter range.

A method for the simultaneous and contact-free dispensing of typically 50-nl liquid jets into microwell plates or onto flat substrates has been realized on the dispensing platform. The so-called DWP principle (**Figure 42**) uses a set of dispensing units, arranged according to the well plate format, each featuring a liquid reservoir, a capillary channel, and a nozzle chamber (**Figure 46**).

The nozzle chamber is filled with the liquid from the reservoir by capillary forces only. Afterward, all dispensing units are actuated in parallel by applying a pressure pulse on top of the dispensing chip initiating the simultaneous and complete drainage of all nozzle chambers through the nozzle. As a result, an array of

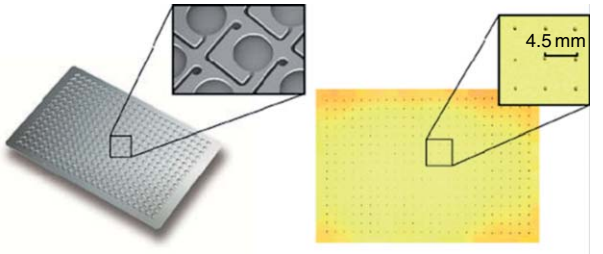


Figure 46 Picture of a 384 dispensing well plate (DWP) dispensing a chip micromachined in silicon. The 384 reservoirs are connected to 384 nozzle chambers next to the reservoirs via capillary channels. No format transformation (reformatting) but the metering and allocation of a defined liquid volume is accomplished within each of the 384 units. All these liquid portions are then transferred to another platform within the pneumatically driven noncontact dispensing process (**Figure 42**) or to a flat surface as depicted on the right.

individual spots is delivered on another well plate or a plane substrate (**Figure 46**, *right*). After switching off the driving pressure, the nozzle chambers refill again from the reservoirs via the capillary channels. Since the reservoir contains a multiple of the nozzle chamber volume, many dispensing events can be performed before the chip has to be refilled. This, for example, enables the fast replication of a certain well plate loading into other plates (compound reformatting) or the addition of nanoliter volumes to plate-based assays. The reproducibility (CV) of the mean dosage volume has been reported to be better than 3% (Steger *et al.* 2004).

2.13.7.4 Strengths and Challenges of the Platform

The wide range of dispensing volumes from 0.05 nl up to 1000 nl of the TopSpot Vario and the DWP, respectively, using the same basic geometric building blocks is certainly the main advantage of the free scalable noncontact dispensing platform. The three dispensing principles can easily be combined on one flat substrate with or without the reformatting by capillary channels and can be actuated in a highly parallel mode of operation.

However, the fabrication costs of the dispensing chips have to be reduced in the future making disposable printheads possible to avoid laborious washing procedures. Since cost reduction is not possible in silicon micromachining (no potential in downscaling of the footprint of the dispensing chips or printheads) microfabrication technologies for polymers like injection molding or hot embossing are likely the most promising alternatives. Today, this technology is pushed forward by the company Biofluidix GmbH (2006) in Germany.

2.13.8 Conclusion

The collection of examples of microfluidic platforms given in the previous sections shows that the platform idea has already been taken up by many groups within the microfluidics community. They not only work on individual components fabricated using diverse technologies, and above all focus on the combination of validated fluidic elements by simple proved technologies. This approach allows to design and fabricate application-specific systems easily and will lead to a paradigm shift from a component and technology-based research to a system-oriented approach. The platforms will allow the microfluidics community to leave today's device-oriented research in order to face the next challenge: the flexible and cost-efficient design of thousands of different applications that might be accessible by using the full potential of microfluidic platforms without starting always from scratch.

A good indicator for the growing interest in microfluidic platform technologies can be also seen in the remarkable number of spin-off companies that arose during the last years trying to commercialize lab-on-a-chip products based on microfluidic platform concepts. Some of them have been mentioned in the previous sections, but there are many more in business already or to emerge in the near future.

References

Abaxis Inc. 2006 Union City, CA, USA, www.abaxis.com, accessed 2006
Abbott P-2006 East Windsor, NJ, USA, www.abbottpointofcare.com, accessed 2006
Advalytix AG 2006 Brunnthal, Germany, www.advalytix.de

Anderson N G 1969 Computer interfaced fast analyzers. *Science* **166**, 317–24

Andersson H, van den Berg A 2004 *Lab-on-Chips for Cellomics, Micro and Nanotechnologies for Life Science.* Kluwer Academic Publishers, Dordrecht, The Netherlands

Anna S L, Bontoux N, Stone H A 2003 Formation of dispersions using ''flow focusing'' in microchannels. *Appl. Phys. Lett.* **82**, 364–6

Auroux P A, Iossifidis D, Reyes D R, Manz A 2002 Micro total analysis systems. 2. Analytical standard operations and applications. *Anal. Chem.* **74**, 2637–52

Berthier J, Clementz P, Raccurt O, Jary D, Claustre P, Peponnet C, Fouillet Y 2006 Computer aided design of an EWOD microdevice. *Sens. Actuators A Phys.* **127**, 283–94

Bioluidix GmbH 2006 Freiburg, Germany, www.biofluidix.com

Bohl B, Steger R, Zengerle R, Koltay P 2005 Multi-layer SU-8 lift-off technology for microfluidic devices. *J. Micromech. Microeng.* **15**, 1125–30

Brenner T, Glatzel T, Zengerle R, Ducrée J 2005 Frequency-dependent transversal flow control in centrifugal microfluidics. *Lab Chip* **5**, 146–50

Bringer M R, Gerdts C J, Song H, Tice J D, Ismagilov R F 2004 Microfluidic systems for chemical kinetics that rely on chaotic mixing in droplets. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* **362**, 1087–104

Burns M A, Johnson B N, Brahmasandra S N, Handique K, Webster J R, Krishnan M, Sammarco T S, Man P M, Jones D, Heldsinger D, Mastrangelo C H, Burke D T 1998 An integrated nanoliter DNA analysis device. *Science* **282**, 484–7

Burtis C A, Anderson N G, Mailen J C, Scott C D, Tiffany T O, Johnson W F 1972 Development of a miniature fast analyzer. *Clin. Chem.* **18**, 753–61

Chatterjee D, Hetayothin B, Wheeler A R, King D J, Garrell R L 2006 Droplet-based microfluidics with nonaqueous solvents and solutions. *Lab Chip* **6**, 199–206

Chen D L, Gerdts C J, Ismagilov R F 2005 Using microfluidics to observe the effect of mixing on nucleation of protein crystals. *J. Am. Chem. Soc.* **127**, 9672–3

Chen D L L, Ismagilov R F 2006 Microfluidic cartridges preloaded with nanoliter plugs of reagents: An alternative to 96-well plates for screening. *Curr. Opin. Chem. Biol.* **10**, 226–31

Chen H, Meiners J C 2004 Topologic mixing on a microfluidic chip. *Appl. Phys. Lett.* **84**, 2193–5

Cheng Y T, Lin L W, Najafi K 2001 A hermetic glass-silicon package formed using localized aluminum/silicon-glass bonding. *J. Microelectromech. Syst.* **10**, 392–9

Cho S K, Moon H J, Kim C J 2003 Creating, transporting, cutting, and merging liquid droplets by electrowetting-based actuation for digital microfluidic circuits. *J. Microelectromech. Syst.* **12**, 70–80

Chou H P, Unger M A, Quake S R 2001 A microfabricated rotary pump. *Biomed. Microdevices* **3**, 323–30

Clark T J, McPherson P H, Buechler K F 2002 The triage cardiac panel. *Point Care* **1**, 42–6

Delamarche E, Bernard A, Schmid H, Michel B, Biebuyck H 1997 Patterned delivery of immunoglobulins to surfaces using microfluidic networks. *Science* **276**, 779–81

deMello A J 2003 Microfluidics – DNA amplification moves on. *Nature* **422**, 28–9

Dendukuri D, Tsoi K, Hatton T A, Doyle P S 2005 Controlled synthesis of nonspherical microparticles using microfluidics. *Langmuir* **21**, 2113–16

Ducrée J, Zengerle R 2004 *FlowMap – Microfluidics Roadmap for the Life Sciences.* Books on Demand GmbH, Norderstedt, Germany

Ducrée J, Haeberle S, Brenner T, Glatzel T, Zengerle R 2005 Patterning of flow and mixing in rotating radial microchannels. *Microfluidics Nanofluidics* **2**, 97–105

Ducrée J, Brenner T, Haeberle S, Glatzel T, Zengerle R 2006 Multilamination of flows in planar networks of rotating microchannels. *Microfluidics Nanofluidics* **2**, 78–84

Duffy D C, McDonald J C, Schueller O J A, Whitesides G M 1998 Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). *Anal. Chem.* **70**, 4974–84

Duffy D C, Gillis H L, Lin J, Sheppard N F Jr., Kellogg G J 1999 Microfabricated centrifugal microfluidic systems: Characterization and multiple enzymatic assays. *Anal. Chem.* **71**, 4669–78

Effenhauser C S, Manz A, Widmer H M 1993 Glass chips for high-speed capillary electrophoresis separations with submicrometer plate heights. *Anal. Chem.* **65**, 2637–42

Effenhauser C S, Bruin G J M, Paulus A, Ehrat M 1997 Integrated capillary electrophoresis on flexible silicone microdevices: Analysis of DNA restriction fragments and detection of single DNA molecules on microchips. *Anal. Chem.* **69**, 3451–7

Ehrfeld W, Golbig K, Hessel V, Lowe H, Richter T 1999 Characterization of mixing in micromixers by a test reaction: Single mixing units and mixer arrays. *Ind. Eng. Chem. Res.* **38**, 1075–82

Ekstrand G, Thorsen T 2005 Liquid router. US 2005/0141344 A1

Ekstrand G, Holmquist C, Örlefors A E, Hellman B, Larsson A, Andersson P 2000 Microfluidics in a rotating CD. In: Van den Berg A, Olthius W, and Bergveld P (eds.) *Proc. Micro Total Analysis Systems 2000,* Enschede, The Netherlands, pp. 311–14

Erickson D 2005 Towards numerical prototyping of labs-on-chip: Modeling for integrated microfluidic devices. *Microfluidics Nanofluidics* **1**, 301–18

Fluidigm Corporation 2006 San Francisco, USA, www.fluidigm.com, accessed 2006

Fredrickson C K, Fan Z H 2004 Macro-to-micro interfaces for microfluidic devices. *Lab Chip* **4**, 526–33

Fu A Y, Spence C, Scherer A, Arnold F H, Quake S R 1999 A microfabricated fluorescence-activated cell sorter. *Nat. Biotechnol.* **17**, 1109–11

Garstecki P, Gitlin I, DiLuzio W, Whitesides G M, Kumacheva E, Stone H A 2004 Formation of monodisperse bubbles in a microfluidic flow-focusing device. *Appl. Phys. Lett.* **85**, 2649–51

Garstecki P, Fischbach M A, Whitesides G M 2005a Design for mixing using bubbles in branched microfluidic channels. *Appl. Phys. Lett.* **86**, 244108

Garstecki P, Stone H A, Whitesides G M 2005b Mechanism for flow-rate controlled breakup in confined geometries: A route to monodisperse emulsions. *Phys. Rev. Lett.* **94**, 164501

Gascoyne P R C, Vykoukal J V, Schwartz J A, Anderson T J, Vykoukal D M, Current K W, McConaghy C, Becker F F, Andrews C 2004 Dielectrophoresis-based programmable fluidic processors. *Lab Chip* **4**, 299–309

Geschke O, Klank H, Telleman P 2004 *Microsystem Engineering of Lab-on-a-Chip Devices.* Wiley-VCH, Weinheim

Goettsche T, Wolff A 2006 IntelliDrug – An integrated intelligent oral drug delivery system. *mst-news* **06**, 36–7

Gravesen P, Braneberg J, Jensen O S 1993 Microfluidics – A review. *J. Micromech. Microeng.* **3**, 168–82

Grumann M, Geipel A, Riegger L, Zengerle R, Ducrée J 2005 Batch-mode mixing on centrifugal microfluidic platforms. *Lab Chip* **5**, 560–5

Gunther A, Khan S A, Thalmann M, Trachsel F, Jensen K F 2004 Transport and reaction in microscale segmented gas–liquid flow. *Lab Chip* **4**, 278–86

Gunther A, Jhunjhunwala M, Thalmann M, Schmidt M A, Jensen K F 2005 Micromixing of miscible liquids in segmented gas–liquid flow. *Langmuir* **21**, 1547–55

Gutmann O, Kuehlewein R, Reinbold S, Niekrawietz R, Steinert C P, De H B, Zengerle R, Daub M 2004a A highly parallel nanoliter dispenser for microarray fabrication. *Biomed. Microdevices* **6**, 131–7

Gutmann O, Niekrawietz R, Kuehlewein R, Steinert C P, de Heij B, Zengerle R, Daub M 2004b Impact of medium properties on droplet release in a highly parallel nanoliter dispenser. *Sens. Actuators A Phys.* **116**, 187–94

Gutmann O, Niekrawietz R, Kuehlewein R, Steinert C P, Reinbold S, De H B, Daub M, Zengerle R 2004c Non-contact production of oligonucleotide microarrays using the highly integrated TopSpot nanoliter dispenser. *Analyst* **129**, 835–40

Gutmann O, Kuehlewein R, Reinbold S, Niekrawietz R, Steinert C P, de Heij B, Zengerle R, Daub M 2005 Fast and reliable protein microarray production by a new drop-in-drop technique. *Lab Chip* **5**, 675–81

Guttenberg Z, Muller H, Habermuller H, Geisbauer A, Pipper J, Felbel J, Kielpinski M, Scriba J, Wixforth A 2005 Planar chip device for PCR and hybridization with surface acoustic wave pump. *Lab Chip* **5**, 308–17

GYROS AB 2006 Uppsala, Sweden, www.gyros.com accessed 2006

Haeberle S, Brenner T, Schlosser H P, Zengerle R, Ducrée J 2005 Centrifugal micromixer. *Chem. Eng. Tech.* **28**, 613–16

Haeberle S, Brenner T, Zengerle R, Ducree J 2006a Centrifugal extraction of plasma from whole blood on a rotating disk. *Lab Chip* **6**, 776–81

Haeberle S, Naegele L, Zengerle R, Ducrée J 2006b A Digital Centrifugal Droplet Switch For Routing of Liquids. In: Kitamori T, Fujita H, Hasebe S (eds), *Proceedings of* μ*TAS 2006,* November 5–9, Tokyo, Japan, pp. 570–2.

Haeberle S, Zengerle R, Ducrée J 2007 Centrifugal generation and manipulation of droplet emulsions. *Microfluidics Nanofluidics* **3**, 65–75

Hansen C, Quake S R 2003 Microfluidics in structural biology: Smaller, faster··· better. *Curr. Opin. Struct. Biol.* **13**, 538–44

Hansen C L, Skordalakes E, Berger J M, Quake S R 2002 A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *Proc. Natl. Acad. Sci. USA* **99**, 16531–6

Harrison D J, Manz A, Fan Z H, Ludi H, Widmer H M 1992 Capillary electrophoresis and sample injection systems integrated on a planar glass chip. *Anal. Chem.* **64**, 1926–32

Harrison D J, Fluri K, Seiler K, Fan Z H, Effenhauser C S, Manz A 1993 Micromachining – A miniaturized capillary electrophoresis-based chemical-analysis system on a chip. *Science* **261**, 895–7

He M Y, Edgar J S, Jeffries G D M, Lorenz R M, Shelby J P, Chiu D T 2005 Selective encapsulation of single cells and subcellular organelles into picoliter- and femtoliter-volume droplets. *Anal. Chem.* **77**, 1539–44

de Heij B, Steinert C, Sandmaier H, Zengerle R 2003 A tunable and highly-parallel picolitre-dispenser based on direct liquid displacement. *Sens. Actuators A Phys.* **103**, 88–92

de Heij B, Daub M, Gutmann O, Niekrawietz R, Sandmaier H, Zengerle R 2004 Highly parallel dispensing of chemical and biological reagents. *Anal. Bioanal. Chem.* **378**, 119–22

Hessel V, Hardt S, Lowe H, Schonfeld F 2003 Laminar mixing in different interdigital micromixers: I. Experimental characterization. *AIChE J.* **49**, 566–77

Hessel V, Lowe H, Schonfeld F 2005 Micromixers – A review on passive and active mixing principles. *Chem. Eng. Sci.* **60**, 2479–501

Honda N, Lindberg U, Andersson P, Hoffman S, Takei H 2005 Simultaneous multiple immunoassays in a compact disc-shaped microfluidic device based on centrifugal force. *Clin. Chem.* **51**, 1955–61

Hong J W, Quake S R 2003 Integrated nanoliter systems. *Nat. Biotechnol.* **21**, 1179–83

Hong J W, Studer V, Hang G, Anderson W F, Quake S R 2004 A nanoliter-scale nucleic acid processor with parallel architecture. *Nat. Biotechnol.* **22**, 435–9

Hosokawa K, Fujii T, Endo I 1999 Handling of picoliter liquid samples in a poly(dimethylsiloxane)-based microfluidic device. *Anal. Chem.* **71**, 4781–5

Huang L R, Cox E C, Austin R H, Sturm J C 2004 Continuous particle separation through deterministic lateral displacement. *Science* **304**, 987–90

Jandik P, Weigl B H, Kessler N, Cheng J, Morris C J, Schulte T, Avdalovic N 2002 Initial study of using a laminar fluid diffusion interface for sample preparation in high-performance liquid chromatography. *J. Chromatogr. A* **954**, 33–40

Jiang F, Drese K S, Hardt S, Kupper M, Schonfeld F 2004 Helical flows and chaotic mixing in curved micro channels. *AIChE J.* **50**, 2297–305

Joanicot M, Ajdari A 2005 Applied physics – Droplet control for microfluidics. *Science* **309**, 887–8

Johnson T J, Ross D, Locascio L E 2002 Rapid microfluidic mixing. *Anal. Chem.* **74**, 45–51

Kamholz A E 2004 Proliferation of microfluidics in literature and intellectual property. *Lab Chip* **4**, 16N–20N

Khan S A, Gunther A, Schmidt M A, Jensen K F 2004 Microfluidic synthesis of colloidal silica. *Langmuir* **20**, 8604–11

Knight J B, Vishwanath A, Brody J P, Austin R H 1998 Hydrodynamic focusing on a silicon chip: Mixing nanoliters in microseconds. *Phys. Rev. Lett.* **80**, 3863–6

Kobayashi I, Mukataka S, Nakajima M 2005 Effects of type and physical properties of oil phase on oil-in-water emulsion droplet formation in straight-through microchannel emulsification, experimental and CFD studies. *Langmuir* **21**, 5722–30

Koltay P, Kalix J, Zengerle R 2004a Theoretical evaluation of the dispensing well plate method (DWP part II). *Sens. Actuators A Phys.* **116**, 472–82

Koltay P, Steger R, Bohl B, Zengerle R 2004b The dispensing well plate: A novel nanodispenser for the multiparallel delivery of liquids (DWP Part I). *Sens. Actuators A Phys.* **116**, 483–91

Kopp M U, de Mello A J, Manz A 1998 Chemical amplification: Continuous-flow PCR on a chip. *Science* **280**, 1046–8

Lai S, Wang S, Luo J, Lee L J, Yang S T, Madou M J 2004 Design of a compact disk-like microfluidic platform for enzyme-linked immunosorbent assay. *Anal. Chem.* **76**, 1832–7

Laser D J, Santiago J G 2004 A review of micropumps. *J. Micromech. Microeng.* **14**, R35–64

Le H P 1998 Progress and trends in ink-jet printing technology. *J. Imaging Sci. Technol.* **42**, 49–62

Lee G B, Hung C I, Ke B J, Huang G R, Hwei B H 2001a Micromachined pre-focused $1\times N$ flow switches for continuous sample injection. *J. Micromech. Microeng.* **11**, 567–73

Lee G B, Hwei B H, Huang G R 2001b Micromachined pre-focused $M\times N$ flow switches for continuous multi-sample injection. *J. Micromech. Microeng.* **11**, 654–61

Lee J, Kim C J 2000 Surface-tension-driven microactuation based on continuous electrowetting. *J. Microelectromech. Syst.* **9**, 171–80

Lee J, Moon H, Fowler J, Schoellhammer T, Kim C J 2002 Electrowetting and electrowetting-on-dielectric for microscale liquid handling. *Sens. Actuators A Phys.* **95**, 259–68

Li P C H 2006 *Microfluidic Lab-on-a-Chip for Chemical and Biological Analysis and Discovery.* Taylor & Francis Group, Boca Raton, FL

Link D R, Anna S L, Weitz D A, Stone H A 2004 Geometrically mediated breakup of drops in microfluidic devices. *Phys. Rev. Lett.* **92**, 054503

Lion N, Rohner T C, Dayon L, Arnaud I L, Damoc E, Youhnovski N, Wu Z Y, Roussel C, Josserand J, Jensen H, Rossier J S, Przbylski M, Girault H H 2003 Microfluidic systems in proteomics. *Electrophoresis* **24**, 3533–62

Liu J, Hansen C, Quake S R 2003 Solving the ''world-to-chip'' interface problem with a microfluidic matrix. *Anal. Chem.* **75**, 4718–23

Lotters J C, Olthuis W, Veltink P H, Bergveld P 1997 The mechanical properties of the rubber elastic polymer polydimethylsiloxane for sensor applications. *J. Micromech. Microeng.* **7**, 145–7

Madou M, Kellogg G J 1998 The LabCD: A centrifuge-based microfluidic platform for diagnostics. *Proc. SPIE* **3259**, 80–93, Systems and technologies for clinical diagnostics and drug discovery

Madou M, Lee J, Daunert S, Lai S, Shih C-H 2001 Design and fabrication of CD-like microfluidic platforms for diagnostics: Microfluidic functions. *Biomed. Microdevices* **3**, 245–54

Manz A, Graber N, Widmer H M 1990a Miniaturized total chemical-analysis system – A novel concept for chemical sensing. *Sens. Actuators B Chem.* **1**, 244–8

Manz A, Miyahara Y, Miura J, Watanabe Y, Miyagi H, Sato K 1990b Design of an open-tubular column liquid chromatograph using silicon chip technology. *Sens. Actuators B Chem.* **1**, 249–55

Manz A, Harrison D J, Verpoorte E M J, Fettinger J C, Paulus A, Ludi H, Widmer H M 1992 Planar chips technology for miniaturization and integration of separation techniques into monitoring systems – Capillary electrophoresis on a chip. *J. Chromatogr.* **593**, 253–8

Marcus J S, Anderson W F, Quake S R 2006a Microfluidic single-cell mRNA isolation and analysis. *Anal. Chem.* **78**, 3084–9

Marcus J S, Anderson W F, Quake S R 2006b Parallel picoliter RT-PCR assays using microfluidics. *Anal. Chem.* **78**, 956–8

Marko-Varga G, Nilsson J, Laurell T 2003 New directions of miniaturization within the proteomics research area. *Electrophoresis* **24**, 3521–32

Martin K, Henkel T, Baier V, Grodrian A, Schon T, Roth M, Kohler J M, Metze J 2003 Generation of larger numbers of separated microbial populations by cultivation in segmented-flow microdevices. *Lab Chip* **3**, 202–7

de Mello A J, Beard N 2003 Dealing with 'real' samples: Sample pre-treatment in microfluidic systems. *Lab Chip* **3**, 11N–19N

Micronics Inc. 2006 Redmond, WA, USA, www.micronics.net, accessed 2006

Mugele F, Baret J C 2005 Electrowetting: From basics to applications. *J. Phys. Condens. Matter* **17**, R705–74

Mugele F, Klingner A, Buehrle J, Steinhauser D, Herminghaus S 2005 Electrowetting: A convenient way to switchable wettability patterns. *J. Phys. Condens. Matter* **17**, S559–76

Nguyen N T, Wereley S T 2002 *Fundamentals and Applications of Microfluidics.* Artech House Books, Boston, MA

Nguyen N T, Wu Z G 2005 Micromixers – A review. *J. Micromech. Microeng.* **15**, R1–16

Nisisako T, Torii T, Higuchi T 2002 Droplet formation in a microchannel network. *Lab Chip* **2**, 24–6

Oh K W, Ahn C H 2006 A review of microvalves. *J. Micromech. Microeng.* **16**, R13–39

Okushima S, Nisisako T, Torii T, Higuchi T 2004 Controlled production of monodisperse double emulsions by two-step droplet breakup in microfluidic devices. *Langmuir* **20**, 9905–8

Oosterbroek R E, van den Berg A 2003 *Lab-on-a-Chip: Miniaturized Systems for (Bio)Chemical Analysis and Synthesis.* Elsevier Science, Amsterdam

Paik P, Pamula V K, Fair R B 2003a Rapid droplet mixers for digital microfluidic systems. *Lab Chip* **3**, 253–9

Paik P, Pamula V K, Pollack M G, Fair R B 2003b Electrowetting-based droplet mixers for microfluidic systems. *Lab Chip* **3**, 28–33

Pollack M G, Fair R B, Shenderov A D 2000 Electrowetting-based actuation of liquid droplets for microfluidic applications. *Appl. Phys. Lett.* **77**, 1725–6

Pollack M G, Shenderov A D, Fair R B 2002 Electrowetting-based actuation of droplets for integrated microfluidics. *Lab Chip* **2**, 96–101

Quake S R, Scherer A 2000 From micro- to nanofabrication with soft materials. *Science* **290**, 1536–40

Raschke G 2005 Molekulare Erkennung mit einzelnen Gold-Nanopartikeln. PhD-thesis, TU Muenchen

Ren H, Fair R B, Pollack M G 2004 Automated on-chip droplet dispensing with volume control by electro-wetting actuation and capacitance metering. *Sens. Actuators B Chem.* **98**, 319–27

Reyes D R, Iossifidis D, Auroux P A, Manz A 2002 Micro total analysis systems. 1. Introduction, theory, and technology. *Anal. Chem.* **74**, 2623–36

Roach L S, Song H, Ismagilov R F 2005 Controlling nonspecific protein adsorption in a plug-based microfluidic system by controlling interfacial chemistry using fluorous-phase surfactants. *Anal. Chem.* **77**, 785–96

Roche 2006 Basel, CH, www.roche.com, accessed 2006

Salemme F R 1972 Free interface diffusion technique for crystallization of proteins for X-ray crystallography. *Arch. Biochem. Biophys.* **151**, 533–9

Schembri C T, Ostoich V, Lingane P J, Burd T L, Buhl S N 1992 Portable simultaneous multiple analyte whole-blood analyzer for point-of-care testing. *Clin. Chem.* **38**, 1665–70

Schembri C T, Burd T L, Kopfsill A R, Shea L R, Braynin B 1995 Centrifugation and capillarity integrated into a multiple analyte whole-blood analyzer. *J. Automatic Chem.* **17**, 99–104

Schonfeld F, Hessel V, Hofmann C 2004 An optimised split-and-recombine micro-mixer with uniform 'chaotic' mixing. *Lab Chip* **4**, 65–9

Schwartz J A, Vykoukal J V, Gascoyne P R C 2004 Droplet-based chemistry on a programmable micro-chip. *Lab Chip* **4**, 11–17

Shestopalov I, Tice J D, Ismagilov R F 2004 Multi-step synthesis of nanoparticles performed on millisecond time scale in a microfluidic droplet-based system. *Lab Chip* **4**, 316–21

Shoji S, Esashi M 1994 Microflow devices and systems. *J. Micromech. Microeng.* **4**, 157–71

Shoji S, Esashi M, Matsuo T 1988 Prototype miniature blood-gas analyzer fabricated on a silicon-wafer. *Sens. Actuators* **14**, 101–7

Sia S K, Whitesides G M 2003 Microfluidic devices fabricated in poly(dimethylsiloxane) for biological studies. *Electrophoresis* **24**, 3563–76

Song H, Ismagilov R F 2003 Millisecond kinetics on a microfluidic chip using nanoliters of reagents. *J. Am. Chem. Soc.* **125**, 14613–19

Song H, Bringer M R, Tice J D, Gerdts C J, Ismagilov R F 2003a Experimental test of scaling of mixing by chaotic advection in droplets moving through microfluidic channels. *Appl. Phys. Lett.* **83**, 4664–6

Song H, Tice J D, Ismagilov R F 2003b A microfluidic system for controlling reaction networks in time. *Angew. Chem. Int. Ed.* **42**, 768–72

Song Y J, Zhao T S 2001 Modelling and test of a thermally-driven phase-change nonmechanical micropump. *J. Micromech. Microeng.* **11**, 713–19

SpinX Technologies 2006 Geneva, Switzerland, www.spinx-technologies.com accessed 2006

Squires T M, Quake S R 2005 Microfluidics: Fluid physics at the nanoliter scale. *Rev. Mod. Phys.* **77**, 977–1026

Srinivasan V, Pamula V K, Fair R B 2004 An integrated digital microfluidic lab-on-a-chip for clinical diagnostics on human physiological fluids. *Lab Chip* **4**, 310–15

Steger R, Bohl B, Zengerle R, Koltay P 2004 The dispensing well plate: A novel device for nanoliter liquid handling in ultra high-throughput screening. *J. Assoc. Lab. Autom.* **9**, 291–9

Steigert J, Grumann M, Brenner T, Mittenbühler K, Nann T, Rühe J, Moser I, Haeberle S, Riegger L, Riegler J, Bessler W, Zengerle R, Ducrée J 2005 Integrated sample preparation, reaction, and detection on a high-frequency centrifugal microfluidic platform. *J. Assoc. Lab. Autom.* **10**, 331–41

Steigert J, Grumann M, Brenner T, Riegger L, Harter J, Zengerle R, Ducrée J 2006 Fully integrated whole blood testing by real-time absorption measurement on a centrifugal platform. *Lab Chip* **6**, 1040–4

Steinert C P, Goutier I, Gutmann O, Sandmaier H, Daub M, de Heij B, Zengerle R 2004 A highly parallel picoliter dispenser with an integrated, novel capillary channel structure. *Sens. Actuators A Phys.* **116**, 171–7

Steinert C P, Mueller-Dieckmann J, Weiss M, Roessle M, Zengerle R, Koltay P 2007 Miniaturized and highly parallel protein crystallization on a microfluidic disc. *Proceedings of the IEEE MEMS 2007*, January 21–25, Kobe, Japan, pp. 561–4

Stone H A, Stroock A D, Ajdari A 2004 Engineering flows in small devices: Microfluidics toward a lab-on-a-chip. *Annu. Rev. Fluid Mech.* **36**, 381–411

Stroock A D, Dertinger S K W, Ajdari A, Mezic I, Stone H A, Whitesides G M 2002 Chaotic mixer for microchannels. *Science* **295**, 647–51

Sudarsan A P, Ugaz V M 2006 Fluid mixing in planar spiral microchannels. *Lab Chip* **6**, 74–82

Sugiura S, Nakajima M, Iwamoto S, Seki M 2001 Interfacial tension driven monodispersed droplet formation from microfabricated channel array. *Langmuir* **17**, 5562–6

Takagi J, Yamada M, Yasuda M, Seki M 2005 Continuous particle separation in a microchannel having asymmetrically arranged multiple branches. *Lab Chip* **5**, 778–84

Tay F E H 2002 *Microfluidics and BioMEMS Applications.* Kluwer Academic Publishers, Boston, MA

Terry S C, Jerman J H, Angell J B 1979 Gas-chromatographic air analyzer fabricated on a silicon-wafer. *IEEE Trans. Electron Devices* **26**, 1880–6

Thorsen T, Maerkl S J, Quake S R 2002 Microfluidic large-scale integration. *Science* **298**, 580–4

Tice J D, Song H, Lyon A D, Ismagilov R F 2003 Formation of droplets and mixing in multiphase microfluidics at low values of the Reynolds and the capillary numbers. *Langmuir* **19**, 9127–33

Tice J D, Lyon A D, Ismagilov R F 2004 Effects of viscosity on droplet formation and mixing in microfluidic channels. *Anal. Chim. Acta* **507**, 73–7

Unger M A, Chou H P, Thorsen T, Scherer A, Quake S R 2000 Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113–16

Urban G A 2006 *BioMEMS.* Springer, Berlin

Utada A S, Lorenceau E, Link D R, Kaplan P D, Stone H A, Weitz D A 2005 Monodisperse double emulsions generated from a microcapillary device. *Science* **308**, 537–41

Van Lintel H T G, Vandepol F C M, Bouwstra S 1988 A piezoelectric micropump based on micromachining of silicon. *Sens. Actuators* **15**, 153–67

Vilkner T, Janasek D, Manz A 2004 Micro total analysis systems. Recent developments. *Anal. Chem.* **76**, 3373–85

Waibel G, Kohnle J, Cernosa R, Storz M, Schmitt M, Ernst H, Sandmaier H, Zengerle R, Strobelt T 2003 Highly integrated autonomous microdosage system. *Sens. Actuators A Phys.* **103**, 225–30

Weigl B H, Bardell R L, Cabrera C R 2003 Lab-on-a-chip for drug development. *Adv. Drug Deliv. Rev.* **55**, 349–77

Wheeler A R, Moon H, Kim C J, Loo J A, Garrell R L 2004 Electrowetting-based microfluidics for analysis of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **76**, 4833–8

Wixforth A 2003 Acoustically driven planar microfluidics. *Superlattices Microstruct.* **33**, 389–96

Wixforth A, Strobl C, Gauer C, Toegl A, Scriba J, von Guttenberg Z 2004 Acoustic manipulation of small droplets. *Anal. Bioanal. Chem.* **379**, 982–91

Woias P 2005 Micropumps – Past, progress and future prospects. *Sens. Actuators B Chem.* **105**, 28–38

Xia Y N, Whitesides G M 1998a Soft lithography. *Annu. Rev. Mater. Sci.* **28**, 153–84

Xia Y N, Whitesides G M 1998b Soft lithography. *Angew. Chem. Int. Ed.* **37**, 551–75

Xia Y N, Kim E, Zhao X M, Rogers J A, Prentiss M, Whitesides G M 1996 Complex optical surfaces formed by replica molding against elastomeric masters. *Science* **273**, 347–9

Yadav M K, Gerdts C J, Sanishvili R, Smith W W, Roach L S, Ismagilov R F, Kuhn P, Stevens R C 2005 In situ data collection and structure refinement from microcapillary protein crystallization. *J. Appl. Crystallogr.* **38**, 900–5

Yamada M, Seki M 2004 Nanoliter-sized liquid dispenser array for multiple biochemical analysis in microfluidic devices. *Anal. Chem.* **76**, 895–9

Yamada M, Seki M 2005 Hydrodynamic filtration for on-chip particle concentration and classification utilizing microfluidics. *Lab Chip* **5**, 1233–9

Yamada M, Seki M 2006 Microfluidic particle sorter employing flow splitting and recombining. *Anal. Chem.* **78**, 1357–62

Yamada M, Nakashima M, Seki M 2004 Pinched flow fractionation: Continuous size separation of particles utilizing a laminar flow profile in a pinched microchannel. *Anal. Chem.* **76**, 5465–71

Yokoyama Y, Takeda M, Umemoto T, Ogushi T 2004 Thermal micro pumps for a loop-type micro channel. *Sens. Actuators A Phys.* **111**, 123–8

Zheng B, Ismagilov R F 2005 A microfluidic approach for screening submicroliter volumes against multiple reagents by using preformed arrays of nanoliter plugs in a three-phase liquid/liquid/gas flow. *Angew. Chem. Int. Ed.* **44**, 2520–3

Zheng B, Roach L S, Ismagilov R F 2003 Screening of protein crystallization conditions on a microfluidic chip using nanoliter-size droplets. *J. Am. Chem. Soc.* **125**, 11170–1

Zheng B, Tice J D, Ismagilov R F 2004a Formation of arrayed droplets of soft lithography and two-phase fluid flow, and application in protein crystallization. *Adv. Mater.* **16**, 1365–8

Zheng B, Tice J D, Ismagilov R F 2004b Formation of droplets of in microfluidic channels alternating composition and applications to indexing of concentrations in droplet-based assays. *Anal. Chem.* **76**, 4977–82

Zheng B, Tice J D, Roach L S, Ismagilov R F 2004c A droplet-based, composite PDMS/glass capillary microfluidic system for evaluating protein crystallization conditions by microbatch and vapor-diffusion methods with on-chip X-ray diffraction. *Angew. Chem. Int. Ed.* **43**, 2508–11

Zheng B, Gerdts C J, Ismagilov R F 2005 Using nanoliter plugs in microfluidics to facilitate and understand protein crystallization. *Curr. Opin. Struct. Biol.* **15**, 548–55

Zimmermann M, Bentley S, Schmid H, Hunziker P, Delamarche E 2005 Continuous flow in open microfluidics using controlled evaporation. *Lab Chip* **5**, 1355–9

Zoval J V, Madou M J 2004 Centrifuge-based fluidic platforms. *Proc. IEEE* **92**, 140–53

Biographies

Stefan Haeberle is a Ph.D. student in the laboratory for MEMS applications at the Department of Microsystems Engineering (IMTEK) at the University of Freiburg, Germany. He received his diploma degree in Microsystem Engineering in 2004 from the University of Freiburg. His research concentrates on the development of lab-on-a-chip systems based on the centrifugal microfluidics platform. Freiburg, 02.11.2006 Stefan Haeberle.

Prof. Dr. Roland Zengerle is the head of the laboratory for MEMS applications at the Department of Microsystems Engineering (IMTEK) at the University of Freiburg, Germany. In addition, he is a director at the Institute for Micro- and Information Technology of the Hahn-Schickard-Gesellschaft (HSG-IMIT). HSG-IMIT is a nonprofit organization supporting industries in the development of new products based on MEMS technologies.

The research of Dr. Zengerle and his team of 50 engineers is focused on microfluidics and covers topics like miniaturized and autonomous dosage systems, implantable drug delivery systems, nanoliter and picoliter dispensing, lab-on-a-chip systems, thermal sensors, miniaturized fuel cells as well as micro- and nanofluidics simulation. Dr. Zengerle coauthored more than 200 technical publications and 25 patents. He is the European editor of the *Springer Journal of Microfluidics and Nanofluidics*. Dr. Zengerle serves on the international steering committee of the IEEE-MEMS conference as well as on the technical program committees of several other international conferences.

2.14 Ultrasonic Transduction

B. T. Khuri-Yakub, A. S. Ergun, O. Oralkan, and G. G. Yaralioglu, E. L. Ginzton Laboratory, Stanford University, Stanford, CA, USA

2.14.1 Introduction

Ultrasonic transducers can be classified based on the physical mechanism involved in the conversion of electrical energy into ultrasonic energy or *vice versa*. Magnetostriction, piezoelectricity, and electrostatics are some of the physical mechanisms used to generate and detect ultrasound.

Throughout the history of ultrasound imaging, piezoelectric crystals, ceramics, polymers, and recently piezocomposite materials have been used to generate and detect ultrasound. Although the idea of electrostatic transducers is as old as the early piezoelectric transducers, piezoelectric materials have dominated ultrasonic transducer technology. Electrostatic transduction can best be explained by means of a parallel plate capacitor, where one plate is kept stationary and the other is displaced either due to an electric field (transmit) or by the reflection of a wave at its surface (receive). The reason why electrostatic or capacitive transducers had not been popular is the need for high electric fields to be maintained to achieve acceptable efficiencies. Recent advances in microfabrication technology have made it possible to build capacitive ultrasound transducers that can compete with piezoelectric transducers.

Micromachined electrostatic transducers were first reported in the late 1980s (Hohm and Hess 1989, Suzuki *et al.* 1989) and the early 1990s (Rafiq and Wykes 1991). These devices were not well characterized, nor easily fabricated. The performance was also not at a level to compete with the piezoelectric transducers. A capacitive micromachined ultrasonic transducer (cMUT) with more advanced fabrication technology and improved performance was introduced in 1993 (Haller and Khuri-Yakub 1994). Advanced microfabrication processes enable realization of the submicron gaps between the electrodes, which makes it possible to achieve high electric fields. These processes also provide a precise control over device dimensions in the vertical and lateral directions. The wide bandwidth and the potential for integration with electronic circuits are other advantages associated with cMUTs. Since the first demonstration of cMUTs in the early 1990s, extensive research has been conducted on fabrication and modeling of this new transducer technology (Eccardt *et al.* 1996, Khuri-Yakub *et al.* 2000, Schindel and Hutchins 1995).

This chapter reviews the theory and modeling of this modern type of electrostatic transducers, explains fabrication methods, and gives examples of imaging systems based on this new technology.

2.14.2 Theory and Modeling of Capacitive Ultrasonic Transduction

Both analytical and numerical methods have been used to analyze cMUTs. A cMUT membrane can be modeled by a parallel plate capacitor with a moving top electrode as shown in **Figure 1**.

Under DC bias, the top electrode is attracted toward the bottom electrode. At equilibrium, the deflection due to the electrostatic force is counterbalanced by the mechanical spring force of the membrane. Assuming that the top electrode is displaced by x, the capacitance of the parallel plate capacitor is given by

$$C(x) = \frac{S\varepsilon_0}{\frac{t_i + t_m}{\varepsilon_r} + g_0 - x} \quad [1]$$

or

$$C(x) = \frac{S\varepsilon_0}{g_{eff} - x} \quad [2]$$

where S is the area of the top electrode, ε_0 the permittivity of vacuum, ε_r the relative permittivity of the insulator and the membrane material (assumed to be the same here), g_0 the initial gap distance under zero bias voltage, t_i the insulator thickness, t_m the membrane thickness, and g_{eff} the effective gap height.

For a parallel plate capacitor, the electrostatic attractive force applied on the top plate is given by

$$F_E = \frac{1}{2}QE \quad [3]$$

where $(1/2)E$ is the electric field due to the charge on the bottom plate. We also assume that $+Q$ and $-Q$ are the charges on the top and bottom plates, respectively. Since Q is given by the product of the capacitance and the applied voltage, the electrostatic force is

$$F_E = \frac{1}{2}\frac{C(x)}{g_{eff} - x}V^2 \quad [4]$$

at the applied bias voltage V. The total net force on the top plate is given by

$$F_N = \frac{1}{2}\frac{C(x)}{g_{eff} - x}V^2 + k_s x \quad [5]$$

where the second term in the right-hand side of the equation is the mechanical force of a membrane whose spring constant is k_s. At equilibrium, the electrical attraction and mechanical restoring forces are equal, and the resulting F_N is zero. Therefore, the relation between the applied voltage and the displacement is given by

$$V = \sqrt{\frac{2k_s x}{S\varepsilon_0}}(g_{eff} - x) \quad [6]$$

As expected, the top electrode displacement increases as the bias voltage increases according to eqn [6]. Also, as the top electrode moves closer to the bottom electrode due to the applied voltage, the electric field increases and the top electrode displaces further, acting as if the spring constant of the top electrode decreases under the influence of the applied voltage. This is called spring softening, which we will evaluate quantitatively later.

The capacitor in **Figure 1** exhibits an important phenomenon. If the bias voltage is increased beyond a certain value, the top electrode collapses on the bottom electrode. This means that a small displacement of the top electrode results in an increase in the electric field and an increase in the attractive force that cannot be balanced by the spring force, with the end result being the collapse of the top electrode onto the bottom electrode. Mathematically, this occurs when the electrostatic force gradient is larger than the mechanical force gradient. One can calculate collapse voltage from eqn [5] by equating the force gradient to zero.

$$\frac{d}{dx}F_N = -\frac{1}{2}\frac{d}{dx}\left(\frac{C(x)}{g_{eff} - x}\right)V^2 + k_s = -\frac{S\varepsilon_0 V^2}{(g_{eff} - x)^3} + k_s \quad [7]$$

The force gradient can also be viewed as the spring constant of the membrane in an electric field.

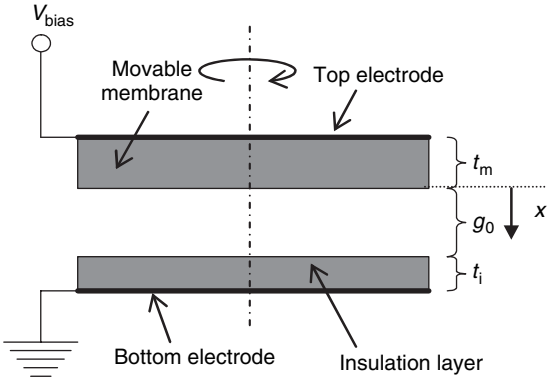


Figure 1 Parallel plate capacitor with moving top electrode.

By substituting the voltage from eqn [6] into eqn [7], the softened spring constant of the membrane can be found in terms of x:

$$k_s' = k_s\left(1 - \frac{2x}{g_{\text{eff}} - x}\right) \qquad [8]$$

The collapse occurs when the gradient of the total force F_N is negative. At that point, incremental increase in electrostatic force overcomes the incremental increase in the mechanical force. On equating eqn [8] to zero, it is found that the displacement is one-third of the effective gap at the collapse voltage:

$$x = \frac{g_{\text{eff}}}{3} \qquad [9]$$

and the collapse voltage is given by

$$V_{\text{col}} = \sqrt{\frac{8k_s g_{\text{eff}}^3}{27S\varepsilon_0}} \qquad [10]$$

By using the above equations, one can calculate the membrane deflection for a given DC voltage. However, for time-varying voltages the equation of motion given in eqn [11] needs to be solved:

$$m\frac{\mathrm{d}^2x(t)}{\mathrm{d}t^2} + c\frac{\mathrm{d}x(t)}{\mathrm{d}t} + k_s x(t) = \frac{\varepsilon_0 S V(t)^2}{2(g_{\text{eff}} - x)^2} \qquad [11]$$

where c is the damping term accounting for energy loss from the vibrating top electrode. Eqn [11] is a nonlinear differential equation. For an arbitrary applied voltage, there is no analytical solution; therefore, eqn [11] should be solved by numerical methods. However, for small voltages one can linearize the differential equation and use the equivalent circuit approach.

2.14.2.1 Equivalent Circuit Model

Mechanical systems can be converted into electrical circuits by using the analogy between the mechanical and the electrical domain. One way to implement this analogy is to replace the forces in the mechanical domain by voltage sources and velocities by electrical currents. Then, an equivalent circuit of the system is constructed. This method becomes an even more powerful tool for the analysis of electromechanical systems where some parts of the system are already in the electrical domain. For example, equivalent circuit analysis is successfully used for piezoelectric transducers for their design and optimization (Mason 1948).

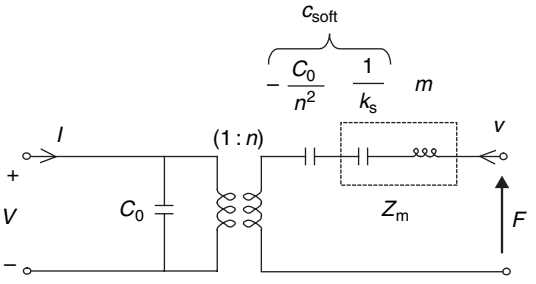


Figure 2 Equivalent circuit model.

Recently, the equivalent circuit model was employed for the characterization of cMUTs (Khuri-Yakub *et al.* 2000, Ladabaum *et al.* 1998). **Figure 2** shows the equivalent circuit for a cMUT transducer. In the electrical part, C_0 is the clamped capacitance of the device at the bias voltage. Spring softening capacitance and the mechanical membrane impedance constitute the mechanical part. The two parts are coupled together through an electromechanical transformer. For a parallel plate capacitor, the electric field and transformer ratio are given by

$$E_0 = \frac{V}{g_{\text{eff}} - x_0} \text{ and } n = E_0 C_0 \qquad [12]$$

respectively where x_0 is the equilibrium displacement under applied bias voltage V and can be calculated using eqn [6]). Note the negative sign in front of the spring softening capacitor.

When the transducer is operated in vacuum, the mechanical port of the circuit is short-circuited. For immersion devices, the mechanical port is simply terminated by the radiation impedance.

The maximum small-signal output pressure of the transducer can be easily calculated using the equivalent circuit method. The maximum output pressure is obtained at the resonant frequency of the membrane where all the reactive elements cancel each other out in the mechanical part of the circuit. At this frequency the output pressure per volt is given by

$$P_{\max} = \frac{n}{S} \qquad [13]$$

If eqn [12] is substituted in eqn [13] assuming that the electrode size is as large as the membrane itself, one obtains the following equation:

$$P_{\max} = \frac{\varepsilon_0 V}{(g_{\text{eff}} - x)^2} \qquad [14]$$

If the maximum DC voltage that can be applied to the cMUT is $V_{col,}$ then the maximum pressure per volt becomes

$$P_{max} = \sqrt{\frac{3k_s\varepsilon_0}{2Sg_{eff}}} \quad [15]$$

According to eqn [15], a smaller gap height and a higher membrane stiffness as well as membrane area improve small-signal output pressure.

Note that the aforementioned equivalent circuit analysis is only valid for voltages that are small compared with the bias voltage. As mentioned earlier, for large signals nonlinear equation of motion needs to be solved. In addition, the results derived can also be used for the receive mode where the membrane displacement is small compared with the gap height. In this case, stiffer and small membranes with small area result in improved sensitivity. The smaller gap height provides also higher sensitivity.

For a piston transducer, the equivalent circuit approach explains the basic operation principles and gives ideas about how to improve the performance. Recently, there were improvements to equivalent circuit to predict the sensitivity, output pressure, and bandwidth of immersion transducer (Lohfink *et al.* 2003, Yaralioglu *et al.* 2003). These models predict the behavior of cMUT when the membrane motion is parallel or around the first resonant frequency. However, more accurate results require finite element model (FEM) analysis.

2.14.2.2 Finite Element Model

FEMs (3D and 2D) have been developed for the cMUTs to calculate collapse voltage, output pressure, bandwidth, and crosstalk. Since the modeling of the whole transducer requires large computer resources, the symmetry of the transducer is almost always used for smaller and more practical models. The simplest FEM representation is a 2D axisymmetric model composed of a membrane and a fluid waveguide. This model represents a single membrane of an infinitely large transducer, and it is assumed that all the membranes move in phase (**Figure 3**).

A commercially available FEM software (ANSYS) was used in this study. The membrane and the fluid waveguide were meshed by planar elements. The electrical ports were added to the membrane by segmenting the gap into many parallel plate capacitors as shown in **Figure 3**. This approach neglects the fringing fields and assumes that the electric field is always perpendicular to the electrodes. Another important component of the model is the absorbing boundary that truncates the mesh at a certain distance from the membrane. This allows simulating the sound radiation into a fluid half space.

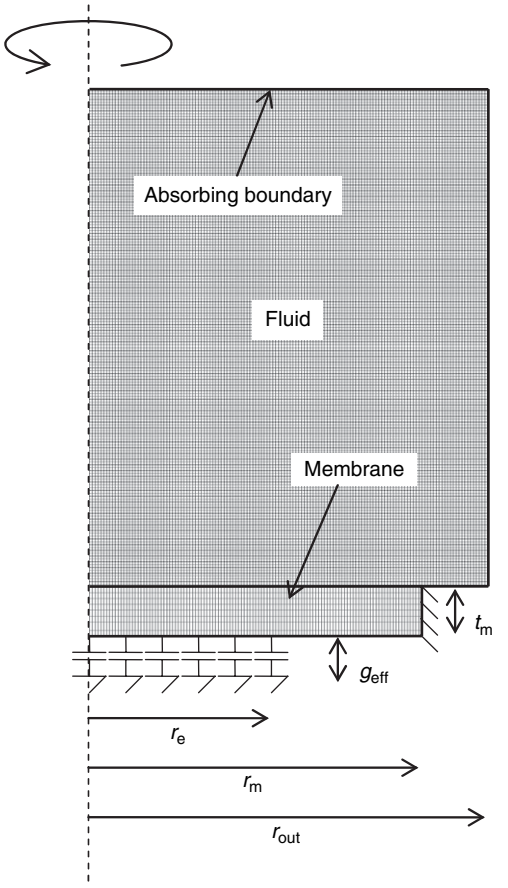


Figure 3 Schematic representation of the finite element model (FEM) model.

The model shown in **Figure 3** generates sound waves in the fluid when the membrane is excited. The field quickly converges into plane waves as it propagates away from the membrane and the wave fronts become parallel to the membrane plane and the absorbing boundary. Parallel incidence of waves on the absorbing boundary is essential to eliminate the reflections. Below a certain cutoff frequency, the fluid waveguide supports only the plane waves. However, if the frequency is high enough, there are also waves propagating at an oblique angle. The cutoff frequency of the waveguide is given by

$$f_c = 1.22\frac{v_L}{2r_{out}} \quad [16]$$

where v_L is the sound velocity in the fluid and r_{out} the radius of the waveguide. Above this frequency, due to the oblique incidence on the absorbing boundary, some part of the incidence wave gets reflected at the absorbing boundary. This results in standing waves in the waveguide. Therefore, care must be taken while examining the results of the model shown in **Figure 3**.

By using the above model, one can perform static, harmonic (small-signal frequency domain), and transient (large-signal time domain) analyses. Static calculations can be used to calculate the collapse voltage of the membrane. Static analysis is also needed to prestress the membrane before harmonic analysis. In this section, we will present the results of harmonic analysis.

The model shown in **Figure 3** was used to calculate the output pressure. The frequency spectra for various membrane radius and thickness combinations were determined. Initially, the collapse voltage was calculated. Then the membranes were biased at 80% of their collapse voltages using static analysis, and an AC voltage of 1 V was applied to the electrical ports for harmonic analysis. **Figure 4** shows the average pressure over the bottom surface of the waveguide including the membrane surface and the nonmembrane area between r_m and r_{out} for four different designs. For each design the thickness of the membrane is adjusted such that the resonant frequencies in vacuum are the same (19.14 MHz). The collapse voltage of the largest membrane ($r_m = 26\,\mu m$) was the highest (98.9 V) among the four designs. As the membrane gets smaller and thinner, the spring constant and the collapse voltage decrease. The collapse voltages of the other three designs are 59.8, 44.8, and 29.7 V. The membranes in this example are made of silicon. For silicon membranes, as the radius decreases, the spring constant-to-area ratio also decreases. To check the validity of eqn [15], we calculated the spring constants of the four designs and compared them with the graph shown in **Figure 4**. For example, eqn [15] predicts 1.7 times higher pressure for a 26-μm membrane than it does for a 16-μm membrane. The calculated result from **Figure 4** is 1.6.

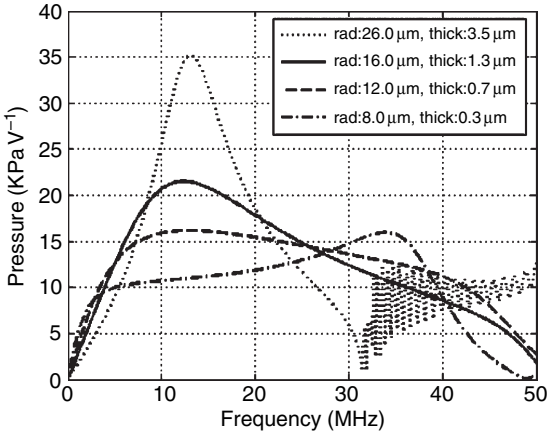


Figure 4 The gap height is 0.1 μm. The insulation layer thickness is also 0.1 μm. The top electrode is assumed to be under the membrane. The relative permittivity of the insulation layer is 3.9. (The equivalent gap height is 0.126 μm.) The membrane material is silicon ($E = 150$ GPa, $\sigma = 0.17$, $\rho = 2332$ kg m^{-3}). The radius of the fluid column is 12% larger than the membrane radius, resulting active (membrane) to total area (membrane + substrate surface) ratio of 80%.

The bandwidth is another important transducer characteristic. The lower cutoff (−3 dB) frequency is determined by the spring constant. As the spring constant increases, the bandwidth of the transducer decreases. The higher cutoff frequency is determined by the membrane mass and the complex loading impedance of the immersion medium as well as the second resonance of the membrane. For relatively softer membranes, the effect of the second resonance is more prominent. By carefully selecting the radius and the thickness combination, one can obtain a very large bandwidth as shown for the 12-μm design.

In the above analysis, we demonstrated small-signal calculations. The pressure output per volt increases for stiffer designs. However, the maximum attainable pressure is limited by the gap height. Calculations for large membrane displacements require nonlinear analysis.

Figure 5 shows large-signal response of the 16-μm membrane. Negative and positive 20-ns pulses (5 and 30 V) were applied to the membrane. The average and center displacements as well as the pulse levels were plotted in the figure. For small amplitudes, negative and positive voltage pulses generate similar amplitude displacement pulses. However, for large pulses, the pulse in the direction of the increased electric field over the gap generates higher amplitude displacement pulse than the pulse going in the other direction. Also, the average membrane displacement over the total area is four times less than the center displacement. This should be kept in mind while choosing the gap height to get the desired maximum output pressure.

Figure 6 plots the Fourier transform of the pressure pulses corresponding to the displacement pulses

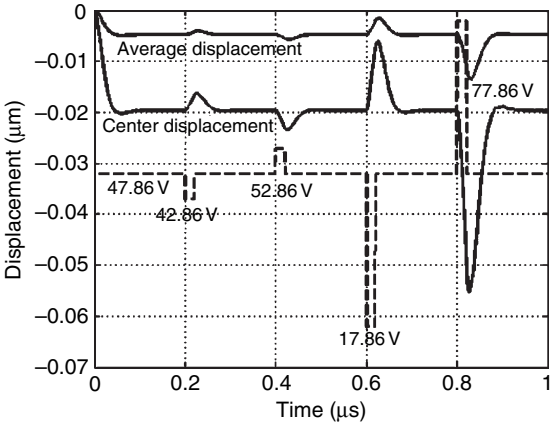


Figure 5 Average and center membrane displacements. The applied voltage waveform is also added on the graph. The membrane is biased at the 80% of the collapse voltage.

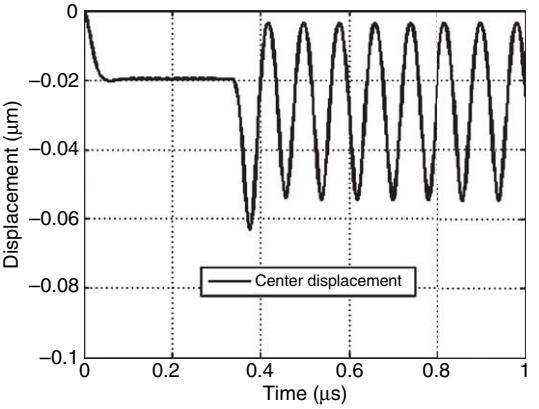


Figure 7 Center membrane displacement. The membrane is biased (47.86 V) at 80% of collapse voltage, and 30 V sine wave at 12.4 MHz is applied.

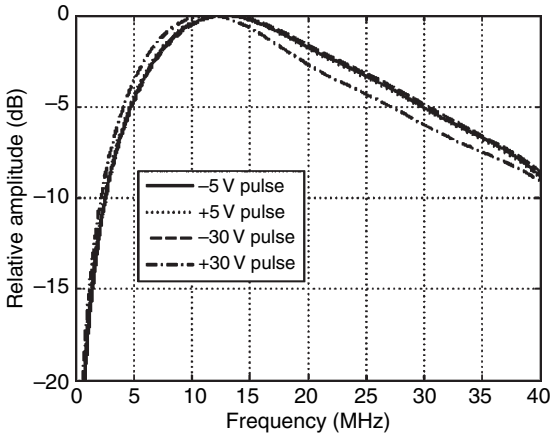


Figure 6 Frequency spectrum of corresponding pressure pulses of **Figure 5**.

shown in **Figure 5**. For small pulses and high negative pulses, the spectrum is very similar to the one obtained using linear analysis. However, for large positive pulses, the spectrum is shifted to the lower frequencies. Due to the nonlinearities, the DC bias voltage increases over the membrane for large positive pulses. This shifts the response of the membrane to the lower frequencies due to the spring softening.

By using large-signal analysis, one can also calculate the harmonic distortion or the amplitudes of the harmonic signals. Second harmonic generation is important for harmonic imaging. To demonstrate the calculation, a long sine wave was applied to the 16-μm membrane shown in **Figure 4**. The center displacement is plotted in **Figure 7**. Note that the instantaneous total voltage over the membrane exceeds the collapse voltage. However, the membrane does not collapse because there is a phase shift between the applied voltage and the displacement.

The fast Fourier transform (FFT) of the pressure waveform reveals the harmonic content of the output pressure. The results are shown in **Figure 8**. The figure depicts the first three harmonics. As expected, the harmonic distortion increases as the applied voltage amplitude gets closer to the bias voltage. The transducer generates a 15-dB second harmonic at high drive levels.

Figure 9 shows the peak-to-peak pressure amplitude and the amplitude of the fundamental frequency as well as the result of the linear calculation shown in **Figure 4**. The fundamental amplitude was calculated by filtering the harmonic amplitudes from the total

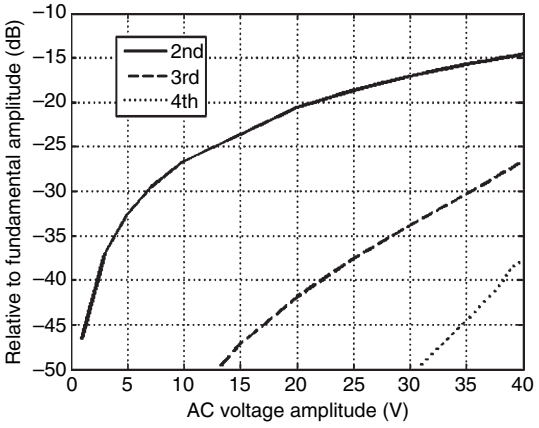


Figure 8 Relative harmonic amplitude as a function of applied AC voltage amplitude. The bias voltage is 47.86 V.

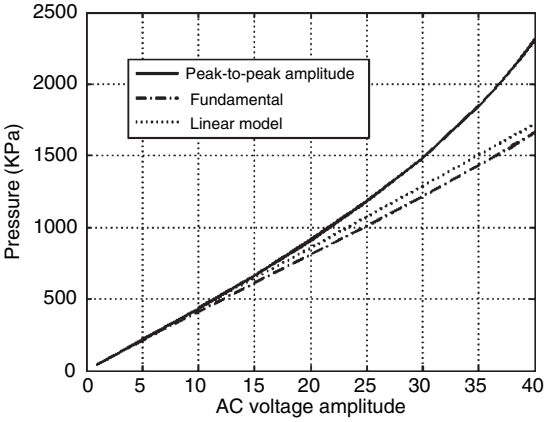


Figure 9 Peak-to-peak amplitude of the pressure waveform. The bias voltage is 47.86 V.

peak-to-peak amplitude. The linear curve is obtained by multiplying the peak-to-peak AC voltage amplitude by the pressure per volt value ($21.5\,\mathrm{k\,Pa\,V^{-1}}$) shown in **Figure 4**. The increase in the peak-to-peak pressure amplitude accelerates as the AC amplitude gets close to the bias voltage. However, the harmonic distortion in the pressure waveform also increases. The fundamental amplitude increases more linearly, and it matches very well with the linear calculation.

2.14.3 Fabrication Technologies

Various methods have been proposed for the fabrication of cMUTs. These methods use one of two manufacturing processes: the sacrificial release process or the wafer bonding process.

2.14.3.1 Surface Micromachining – Sacrificial Release Process

This process is called the sacrificial release process since the membrane material is deposited over a sacrificial layer that is later removed to leave a gap between the membrane and the substrate.

In this method, a conductive silicon wafer is used as the substrate. The wafer is doped to achieve high conduction at its surface, which is going to be the back electrode of the transducer. A silicon-rich, low-stress silicon nitride film is then deposited by using low-pressure chemical vapor deposition (LPCVD). Typically, the deposited films have a tensile stress of less than 100 MPa and a refractive index of 2.4. This layer, also called the insulation layer, must be sufficiently thick to protect the silicon wafer from the etchant during the long wet sacrificial layer etch. But, since the capacitance of the insulation layer is in series with the active gap capacitance, it should not be arbitrarily thick. Otherwise, the collapse voltage will be increased beyond practical limits. Typically, an insulation layer that is 1000 Å thick is sufficient to avoid problems with possible pinholes that are common in thin silicon nitride layers.

The next step is the deposition of the sacrificial layer. Various materials can be used for the sacrificial layer. Polysilicon can be etched easily in potassium hydroxide (KOH) while silicon nitride is very durable in KOH. Polysilicon is also available in most of the fabrication facilities. For these reasons, polysilicon is widely used as a sacrificial material.

Sacrificial layer deposition is done in two steps. The first deposition is followed by lithography and etch steps, so that the channels used by KOH to remove the sacrificial layer are thinner than the cavity thickness. The subsequent dry etch removes all the polysilicon except in the regions where the cavity will be formed. The dry etching stops at the silicon nitride layer (**Figure 10(a)**). Then, a second layer of polysilicon is deposited (**Figure 10(b)**). The thickness of the second polysilicon layer determines the thickness of the channels; the total polysilicon thickness from the first and second depositions determines the cavity height. For polysilicon, low-temperature depositions are preferred for better thickness control.

The cavity, membrane shape, and the etch channels are defined by another photolithography and dry etch steps. While the membranes can be in any shape, circles and hexagons are easier to model and thus are used more often. Because they determine the frequency response of the element, both the shape and the size of the membrane are critical cMUT design parameters.

In the next step, another layer of silicon nitride is deposited by LPCVD (**Figure 10(c)**) to form the membranes. The insulation layer and the membrane are of the same material. At this step, the membrane thickness is not finalized yet because another layer of silicon nitride will be added during the sealing step.

Small holes (etch holes) through the silicon nitride layer are opened through lithography and dry etch steps (**Figure 10(d)**). The etch holes are located on the etch channels. These holes provide access to the sacrificial layer for the etchant solution. When the wafer is immersed in KOH solution, KOH starts etching through these holes and etches its way into the cavity through the channels and releases the

membrane (**Figure 10(e)**). During the long KOH etch, the etch-stop layer protects the silicon wafer. The polysilicon etch rate of KOH depends on temperature. At room temperature, etching process may take several days, depending on the membrane size. This slow rate is due, in part, to the narrow and thin etch channels, which limit the diffusion of KOH. Near boiling point, the wet release may be as fast as several hours. However, because the etch selectivity worsens between silicon nitride and polysilicon at high temperatures, this fast release process is not always preferred. A moderate temperature is selected to give a reasonable release time while maintaining the high etch selectivity.

After the long release, the holes are sealed by another layer of LPCVD silicon nitride (**Figure 10(f)**). Because silicon nitride is deposited at low pressure, the cavity is considered to be vacuum-sealed for practical purposes. In immersion applications, including medical imaging, the cavity must be sealed; otherwise, fluid fills the cavity and hinders proper cMUT operation. If the cavity is not sealed, squeeze film damping will load the membrane vibrations, making the transducer inefficient. For the same reason (but with less effect), it is also undesirable to leave air inside the cavity while sealing.

The rest of the process establishes the electrical connections from the bond pads to the top and bottom electrodes. A lithography and an etch step open connections to the ground plane through the silicon nitride layers. Subsequently, aluminum is sputtered over the whole wafer and is patterned by lithography and wet etch (**Figure 10(g)**). Because it provides conformal coverage, sputtering is preferred for this step. This final lithography and etch step define the top electrode coverage over the membrane. After depositing the metal, it is best to anneal the wafer

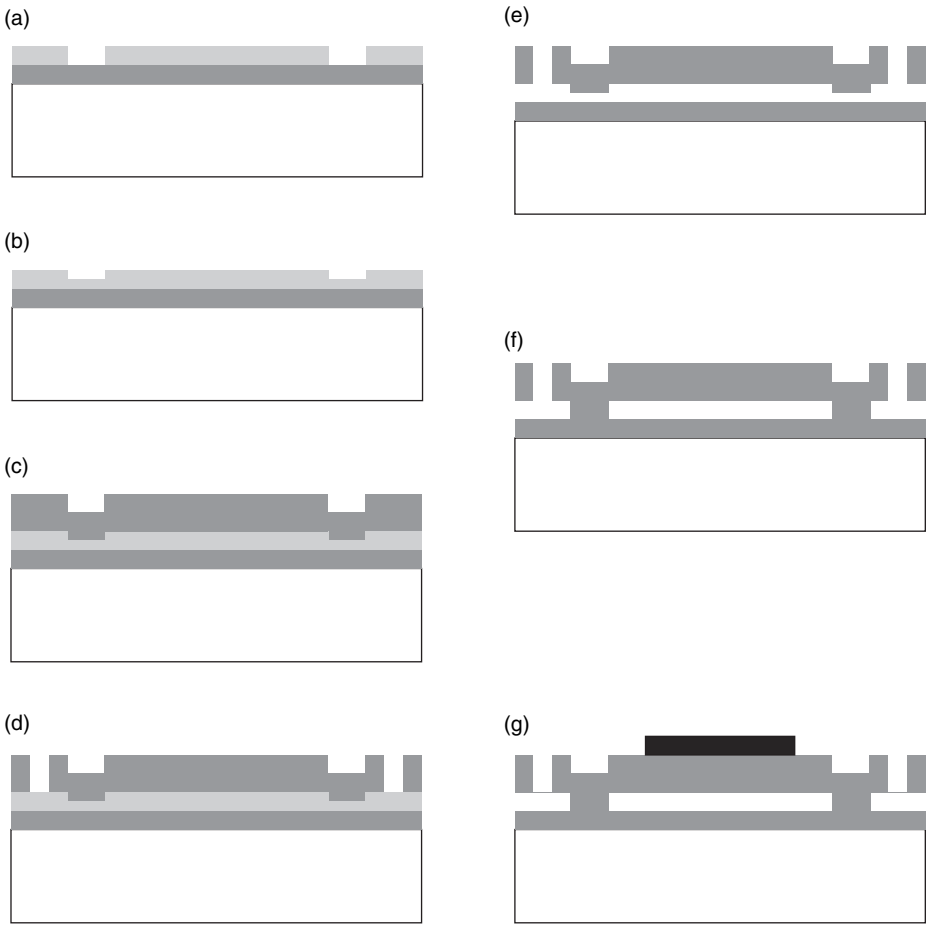


Figure 10 Sacrificial release process: (a) substrate doping, insulation layer deposition, sacrificial layer deposition, and patterning; (b) second sacrificial layer deposition; (c) membrane deposition; (d) etch holes; (e) KOH etch; (f) sealing; and (g) metal deposition and patterning.

to establish good ohmic contact with the ground. However, annealing generates high tensile stress in the aluminum electrode (around 500 MPa), which alters both the static deflection under atmospheric pressure and the mechanical resonant frequency of the membranes. Because a sufficiently good ohmic contact may still be obtained on a highly doped substrate, it is possible to skip the annealing step. Otherwise, high tensile stress must be accounted for in the cMUT design. **Figure 11** shows various membranes fabricated using the sacrificial release process.

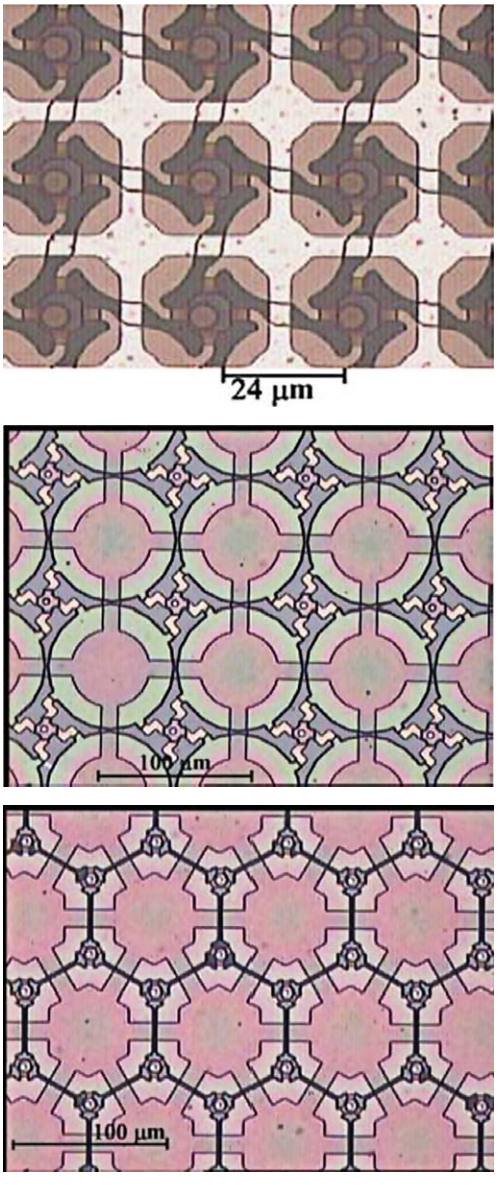


Figure 11 Various membranes fabricated using sacrificial release process.

There are several variations of sacrificial etch processing. For example, the fabrication may start with an insulating substrate such as quartz (Knight and Degertekin 2003). In this case, the bottom electrode should be formed by another deposited conducting layer. Instead of LPCVD nitride, one can use plasma enhanced chemical vapor deposition (PECVD) nitride (Caliano *et al.* 2000), which allows low-temperature processing. Another variation is the use of dry etching rather than wet etching to remove the sacrificial layer.

2.14.3.2 Bulk Micromachining – Wafer Bonding Process

In this technique, cavities are first defined on a silicon wafer, which is subsequently bonded to a silicon-on-insulator (SOI) wafer. The silicon layer of the SOI wafer forms the membranes after removing the handle part and the buried oxide (BOX) layer of the SOI wafer.

To define the gap height and the membrane shape, the prime quality silicon wafer is thermally oxidized to a predetermined thickness, followed by a photolithography step to define the cavity shape. The thermally grown silicon dioxide layer is etched with hydrofluoric acid solution (or dry etched) through the photoresist pattern all the way to the silicon, as shown in **Figure 12(a)**. After the photoresist is removed, another layer of silicon dioxide is thermally grown (**Figure 12(b)**). Because the membrane is silicon, which is not an insulator, the second oxidation isolates the conductive silicon substrate from the top electrode to avoid shorting.

Following the second oxidation, the SOI wafer and the prime wafer (after being RCA-cleaned and surface-activated) are brought together in vacuum, as shown in **Figure 12(c)**. As soon as the two wafers come in close contact, short-range van der Waals forces attract the two wafers, and weak hydrogen bonds develop between them. The wafers are immediately annealed/oxidized at 1100°C to form strong covalent bonds. After bonding and annealing, the handle of the SOI wafer (including the BOX layer) is removed to release the membranes (**Figure 12(d)**), either by wet etching or by dry etching. Wet etching of silicon, a well-known process, can be done with a solution of either KOH or tetramethylammonium hydroxide (TMAH). It is critical to ensure that the silicon dioxide layer on the back of the prime wafer holds off the wet etchant during this step. A better method is to grind/polish the handle portion of the

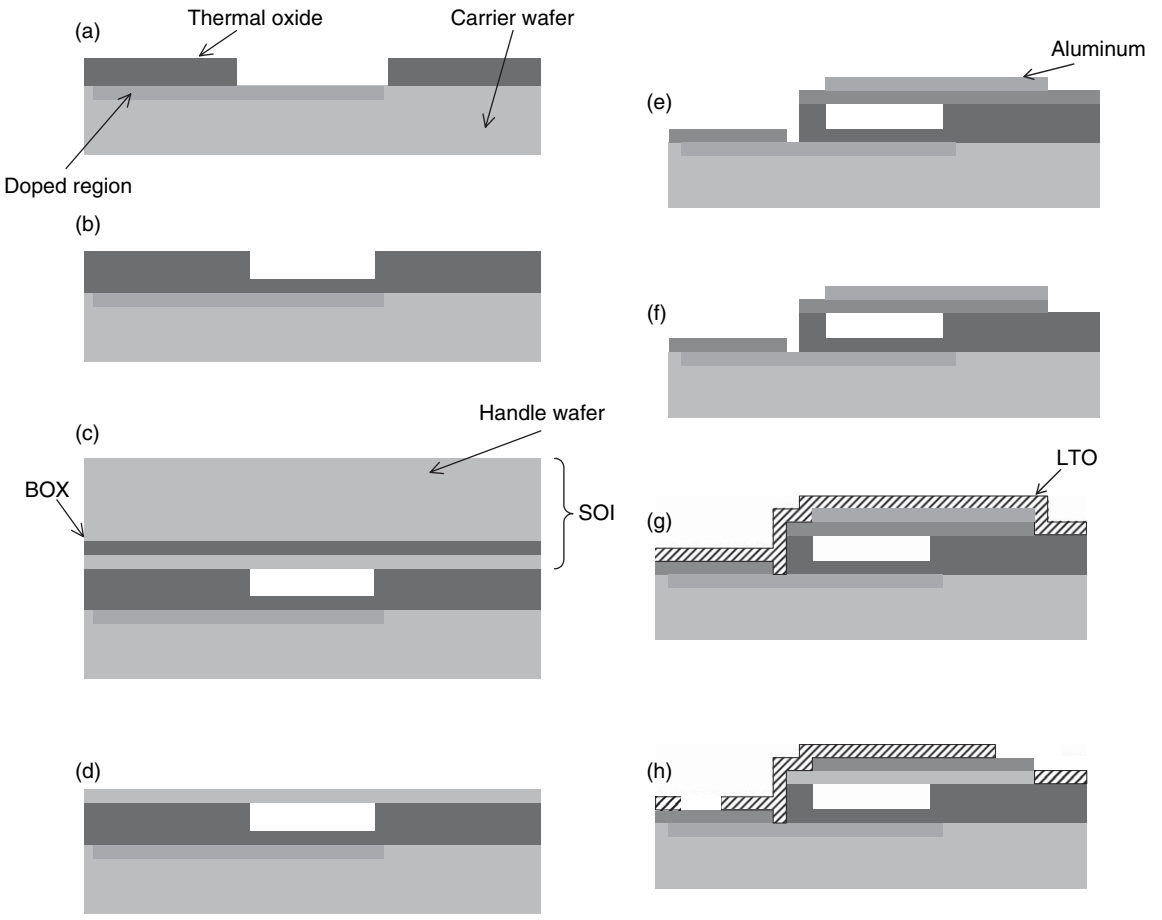


Figure 12 Wafer bonding fabrication steps: (a) cavity definition, (b) second oxidation, (c) wafer bonding, (d) removing handle part and buried oxide (BOX) layer, (e) metal deposition, (f) element definition, (g) low-temperature oxide (LTO) deposition, and (h) patterning of LTO for wire bonding.

SOI wafer down to 50 µm before the wet etch step, which relaxes the constraint on the thickness of the silicon dioxide layer on the back of the prime wafer. Because both KOH and TMAH have very slow etch rates for silicon dioxide, once the silicon etch is complete, the wet etch process will in effect stop at the BOX layer, which can then be removed with hydrofluoric acid. In this case, the silicon membrane acts as the etch stop. If the handle wafer is dry etched with deep reactive ion etching (DRIE), then the BOX layer serves as the etch-stop layer. Using recently developed DRIE equipment, this etch step may be 25 times faster than the wet etch, eliminating concerns about the serial nature (one wafer at a time) of the DRIE process.

The remaining steps of the wafer-bonded cMUT process are very similar to the surface micromachined cMUT process, with the exception of element definition. Openings through the silicon and silicon dioxide layers are defined with photolithography and dry etch steps to access the bottom silicon layer to make the ground connection. The top electrode is sputtered and patterned by another photolithography and wet etch step (**Figure 12(e)**). Because the wafer-bonded cMUTs have no surface topology at this step to cause discontinuity, the top electrode can also be evaporated. Finally, the elements are defined by etching isolating trenches, as shown in **Figure 12(f)**. Because the active silicon layer is not a good insulator, it must be etched all the way to the oxide layer to electrically isolate the elements. The trenches are defined with photolithography and then dry etched into the silicon. After defining elements, one can use low-temperature oxide (LTO) to passivate the metal electrode to provide an insulating surface (**Figure 12(g)** and **12(h)**).

2.14.4 Integration of Ultrasonic Transducer Arrays with Electronic Circuits

Modern ultrasonic imaging systems combine the use of multielement transducer arrays with digital circuitry and signal processing techniques to provide high-quality images at a relatively low cost. These systems also present real-time images at high frame rates. A typical commercial state-of-the-art ultrasonic imaging system is a massive instrument. Although this kind of a system can be moved around in a hospital environment, it does not quite fit to the modern definition of portable set by communication devices. The bulkiness of the ultrasonic imaging equipment is a result of the necessity to simultaneously process data coming from multiple transducer channels. A block diagram of the imaging system is shown in Figure 13. In a typical ultrasonic imaging system, the transducer array is located in a probe, which is connected to the processing unit through a bundle of cables. The processing unit contains a pulser, a low-noise preamplifier, a time-gain controlled amplifier, a low-pass filter, and an analog-to-digital converter for each channel. Digital circuits for transmit/receive (T/R) beamforming are also located in the processing unit. The total number of printed circuit boards in the unit can be as many as 40, depending on the number of ultrasonic channels. Channel counts in modern phased array systems can be as many as 256. During conventional phased array operation, all channels are active both in transmit and in receive.

The cable losses and parasitic capacitances degrade the noise performance of the imaging system and limit the achievable bandwidth. The integration of the ultrasonic transducer array with supporting electronics minimizes the parasitics and consequently improves the sensitivity and preserves the transducer bandwidth. Furthermore, integration helps decrease the number of interconnects between the transducer array and the signal processing unit to a manageable level by multiplexing several channels. For applications such as intraoperative navigation and intravascular diagnostic imaging, the size of the probe and the number of cables connected to the probe are even more important. Integrating the electronics with the array enables the realization of a compact probe with a minimum number of external connections. For large 2D transducer arrays with thousands of elements, multiplexing in the probe is a necessity.

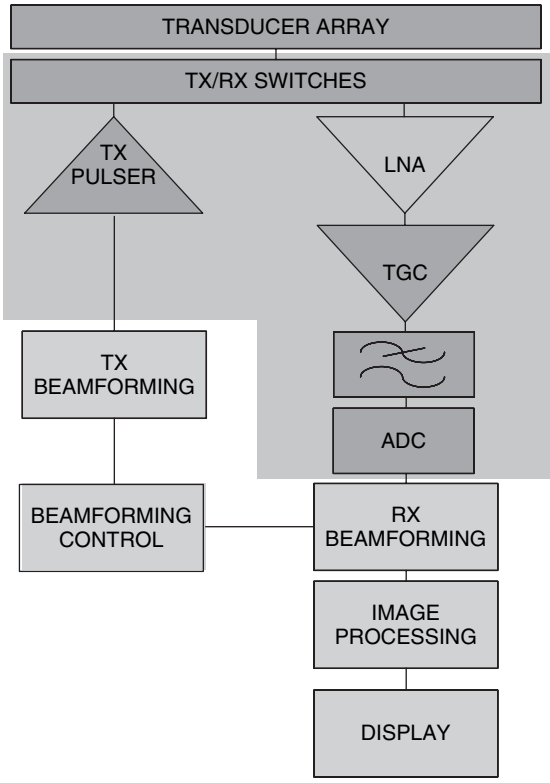


Figure 13 Block diagram of a medical ultrasonic imaging system. ADC, analog-to-digital converter; LNA, low-noise amplifier; TGC, time-gain control; TX, transmitter; RX, receiver.

Continuous scaling of microelectronics reduces the cost per function and makes it possible to pack more functionality on a single die. The increased computational power and the higher speed of microelectronic circuits can also ameliorate processing large amounts of data resulting from increased channel counts in next-generation ultrasound imaging systems.

In this section, different methods for integrating ultrasonic transducer arrays with electronic circuits are reviewed; important aspects of ultrasound transceiver design are also outlined.

2.14.4.1 Monolithic and Multichip Approaches

To integrate transducer arrays with associated T/R electronics, several interconnection schemes have been proposed. These schemes can be classified in two major groups: monolithic and multichip approaches.

Monolithic approaches integrate ultrasonic transducer arrays with supporting electronic circuits on the same substrate. This kind of integration can be achieved by post- or preprocessing standard silicon complementary metal oxide semiconductor (CMOS) integrated circuits (ICs). Monolithic integration of microelectromechanical systems (MEMS) with electronic circuits usually results in a compromise of the performance of one or both components. It has been demonstrated that a modified CMOS or bipolar CMOS (BiCMOS) process could be used to fabricate ultrasonic transducers on the same silicon substrate with electronic circuits (Eccardt *et al.* 1997, Niederer *et al.* 1999). Although this method provides a cost-effective means of integration, it is not suitable for 2D arrays because the usable active area is compromised either by the electronics, if the element area is shared with the electronics, or by the interconnect lines, if the electronics is placed on the periphery of the array and connected to the array elements through routing lines. In this approach, the number of available design parameters is also reduced due to the use of a standard process. An alternate method uses postprocessing to fabricate cMUT elements over the electronics chip, which is built using a standard CMOS process (Daft *et al.* 2004, Noble *et al.* 2001, 2002). The wafer with the electronics goes through passivation and chemical–mechanical polishing (CMP) processes before the cMUT fabrication process. This integration method, probably being only slightly costlier than the former, is suitable for 2D array integration because of its better area utilization. Because the preprocessed electronics limit the temperature of the postprocesses, a low-temperature process must be used to fabricate cMUTs over the electronics. Although a low-temperature cMUT process using PECVD has been tailored to achieve good control over the stress in the membrane and reasonable uniformity, a high-temperature cMUT process using LPCVD still results in better on-wafer and wafer-to-wafer uniformity and a higher yield.

The multichip approach requires that the electronics and the transducers be fabricated on separate wafers and that they be integrated in the final step. For arrays in linear configuration, the integration can be accomplished by wire bonding, which is viable if a larger footprint is acceptable and the level of tolerable parasitic capacitance is high. Flip-chip bonding is required if the transducer elements are densely packed, such as in a 2D array, or if the elements are very small so that a large parasitic capacitance would significantly degrade the performance. In the multichip approach, high-temperature processes can be used to fabricate high-density and high-performance cMUT arrays with high yield and uniformity. The integration of the cMUT arrays with the electronics comes at the expense of additional steps to the cMUT process, which are the fabrication of the through-wafer interconnects to bring the electrical connections to the backside of the silicon wafer and flip-chip bonding. Both of these processes are CMOS compatible. Because the electronics and the transducer arrays are fabricated on separate wafers, the overall yield of the manufacturing process is enhanced and the turn-around time is reduced. The cost reduction associated with the higher yield and a faster turn-around time compensates (probably surpasses) the increase in the cost due to the increase in the number of processing steps. The integration of array elements with electronics on the probe not only improves the performance but also reduces the number of external leads to a manageable level.

We have developed two flip-chip bonding techniques to integrate 2D cMUT arrays with front-end ICs. For both techniques, a Ti/Cu/Au metal stack is evaporated onto backside pads of cMUT arrays to enhance electrical contact and to establish the necessary under bump metallurgy (UBM) for flip-chip bonding. The first method is based on anisotropic conducting film (ACF). For this technique, a wire bonder is used to place 25-μm-diameter, 25-μm-tall gold stud bumps on the 50 μm × 50 μm bond pads on a matching chip, which can be an IC or a fan-out board. The gold studs are coined to obtain a uniform bump height across the whole die. ACF (FP1708E, Sony Chemicals, Tokyo, Japan) is then laminated on the matching chip. A flip-chip bonder (Model M8, Research Devices Inc., Piscataway, NJ) is used to align and bond the two parts by applying a pressure of 30 g per bump and a peak temperature of 190°C for 20 s. The ACF is cured at the elevated temperature. This film conducts at the points where it is squeezed between the cMUT pads and the gold bumps. Results obtained using this flip-chip bonding technique are shown in **Figure 14**. For the second flip-chip bonding technique, Sn/Pb solder bumps are deposited on the bond pads of the matching chip. These pads are first coated with 5 μm of Ni/Au with an electroless plating process to provide the appropriate UBM for the solder bumps. Solder bumps with diameters of 80 μm are placed onto these bond pads using a solder jetting process (Kasulke *et al.* 1998). Bumps can be alternatively

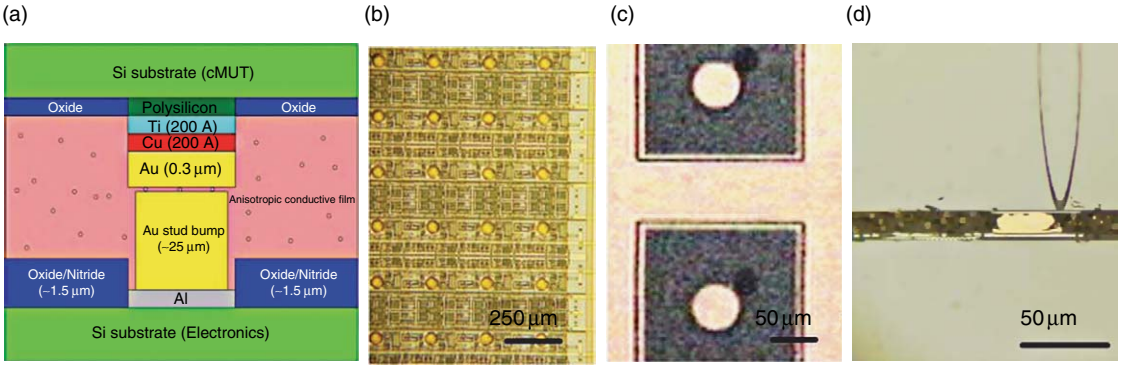


Figure 14 (a) Gold stud – anisotropic conducting film (ACF) flip-chip bonding technique. (b) Stud-bumped integrated circuit (IC) pads. (c) Backside bond pads on the capacitive micromachined ultrasonic transducer (cMUT) array. (d) Cross section of the flip-chip bonded sample.

placed on the backside pads of the cMUT array. The cMUT and IC are aligned and heated to 150°C with 4 g per bump pressure. Following the alignment, the samples are placed into an inert atmosphere oven and a reflow at 200°C is performed. The results from this second technique are shown in Figure 15.

2.14.4.2 Ultrasound Transceivers

Although digital circuits process data in the back-end for image reconstruction and enhancement, the signals transmitted and received by the transducer are fundamentally analog. Therefore, the analog front-end is critical as it sets limits on system performance

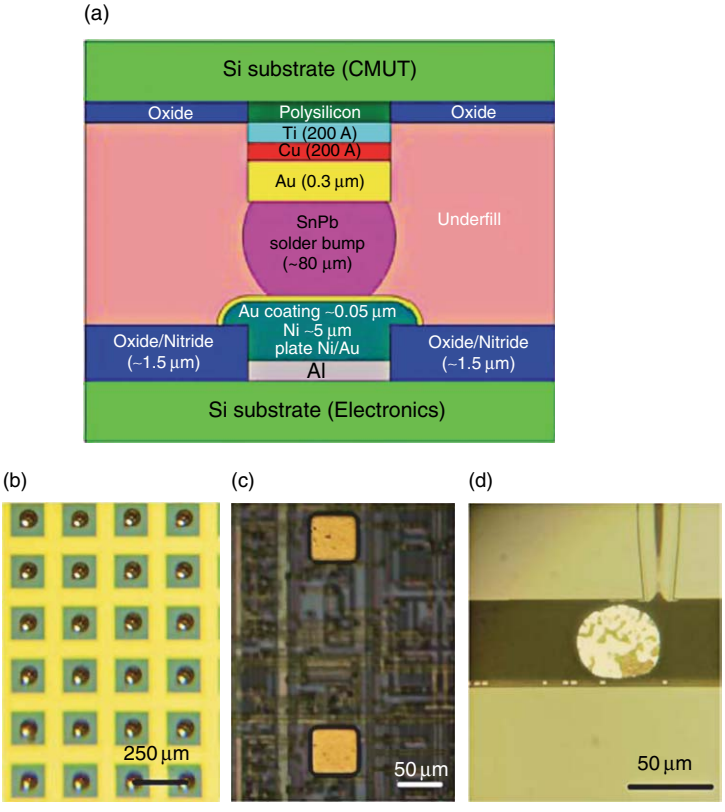


Figure 15 (a) Solder bump – Ni/Au plating flip-chip bonding technique. (b) Solder-bumped backside bond pads on the capacitive micromachined ultrasonic transducer (cMUT) array. (c) Electroless Ni/Au-plated IC pads. (d) Cross section of the flip-chip bonded sample.

(Schafer and Lewin 1984). In this section, the major trade-offs in the design of the analog front-end circuits are analyzed. The design considerations for ultrasound transceivers can be divided into three categories: the transmitter, the receiver, and the switching between the two.

The function of the transmitter is to excite the transducer to produce a short-duration pulse of acoustic energy. The transmitters in an array configuration are driven by pulses in a delay pattern determined by the transmit beamformer to produce a desired focal point. The pulses are amplified by the transmitter circuit driving the transducer. In more sophisticated systems the pulses might be generated by a digital-to-analog converter according to the timing information provided by the beamformer to shape the transmit pulses for better energy utilization or even for coding to improve the overall signal-to-noise ratio (SNR) of the system. The excitation voltages in existing systems can be as high as a few hundred volts to maximize the SNR. However, the excitation voltages cannot be arbitrarily high. There are limits on the total acoustic power set by the Food and Drug Administration (FDA), so that the acoustic radiation does not exceed the safe doses for the patient, e.g., $94\,\mathrm{m\,W\,cm^{-2}}$ maximum spatial-peak temporal-average intensity for fetal imaging (FDA 1997). It is difficult to measure the acoustic output power in different operating modes and to ensure compliance with FDA regulations (Patton *et al.* 1994). There are also other practical limits on the excitation voltages to control electromagnetic interference and to avoid a potential shock hazard to the patient. An additional practical limit for piezoceramic transducers is the gradual depoling caused by high-voltage pulses. For micromachined transducers, dielectric breakdown of thin films imposes an equivalent practical limit on DC bias and AC excitation voltages.

The major role of the receiver in an ultrasonic imaging system is to amplify the signals sensed by the transducer as a result of the echoes returning from the discontinuities within the tissue. The electronic circuits following the transducer in the receive signal chain should preserve the frequency characteristics of the received echo while adding minimum amount of noise. The overall bandwidth of the receiver should thus not exceed the signal bandwidth to prevent excess noise. The dynamic range of received signals in a typical ultrasound pulse echo system is quite large. The large dynamic range of signals is caused by two major physical mechanisms. First, there is a difference in amplitude between signals reflected from interfaces at different depths. This difference is due to the attenuation in the medium. Second, there is also a difference between the reflection coefficients of the reflectors at the same range. By assuming $1\,\mathrm{dB\,cm^{-1}\,MHz^{-1}}$ attenuation in tissue, a penetration depth of 5 cm results in a dynamic range of 100 dB at an operating frequency of 10 MHz. Adding another 60 dB accounting for the reflectivity variations results in a 160-dB total dynamic range (Brunner 2002). Such a dynamic range cannot be achieved by a practical receiver unless some sort of gain control is employed. Compared with radio receivers, ultrasound receivers fortunately are advantageous for implementing gain control. In an ultrasonic imaging system, part of the signal variation is due to the attenuation in the medium. Thus, changing the gain as a function of time compensates for this variation, leaving only the variation due to the reflectivity differences in the final image. The lower limit of the signal dynamic range is determined by the minimum detectable signal, which roughly corresponds to the integrated noise of the transducer-preamplifier combination. The upper limit of the signal dynamic range is set by the closest and strongest reflector in the field of view. The described gain control scheme is generally referred to as time-gain control or time-gain compensation (TGC). Other design considerations include low-phase distortion and high isolation of digital and analog signals if digital circuits are integrated on the same die with the analog circuits.

The T/R switch connects the transducer to the transmitter while isolating the receiver during the transmit cycle. Similarly, it isolates the transmitter and connects the transducer to the receiver during the receive cycle. The key features of the T/R switch are low loss and low distortion in the conducting state and good isolation in the nonconducting state. The typical implementation of the T/R switch is by using pn or pin junction diodes as shown in **Figure 16**.

As discussed earlier, exciting the transducer with a high-voltage pulse is advantageous to maximize the SNR of the received echoes. Generating high-voltage pulses is difficult in advanced process technologies that enable the increased processing power mentioned earlier. Increased processing power stems from the scaling of semiconductor devices; the continuous scaling of these devices enables faster operation, increased density and functionality, and lower power consumption. However, supply voltages need to be scaled as well; otherwise high electric field

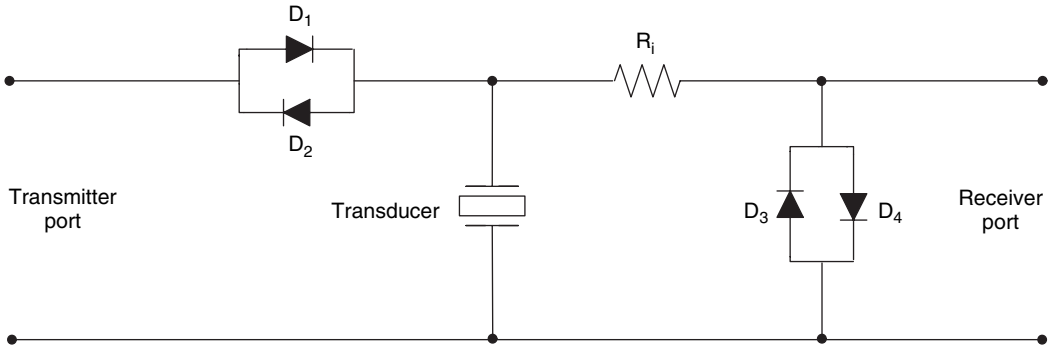


Figure 16 Conventional transmit/receive (T/R) switch.

intensities pose reliability problems. The major reliability-threatening mechanisms related to these fields are oxide breakdown, hot carrier degradation, and junction breakdown.

In systems involving both complex signal processing and high-voltage operation, it is usually preferable to perform most of the signal processing at low voltage and translate the result into high voltages as a last step. Although the integration of high-voltage devices and high-performance, low-power circuits is an attractive option, other system partitioning schemes that use multiple chips have also been proposed for ultrasonic imaging systems (Robinson and Mo 1992). However, for applications with a high channel count and a restricted area, these schemes are not feasible. The implementation of high-voltage circuits can be achieved by taking actions at the technology, layout, or circuit design levels.

One of the most straightforward approaches to implement high-voltage circuits is to use a process technology offering high-voltage devices along with low-voltage, high-speed transistors. These technologies require extra masks and processing steps and hence have a higher cost. These processes are also not as easily accessible as are mainstream standard CMOS processes. A more sophisticated and nonstandard approach for the implementation of high-voltage ICs is to design high-voltage devices with a standard, unmodified, low-voltage CMOS technology (Ballan and Declercq 1999, Maluf *et al.* 1996). In this approach, the designer uses the available layers in the given process but combines them in an unconventional way. For example, the n-well implant is very convenient to form lightly doped drain extensions in a high-voltage n-channel MOS transistor. In a standard process the well implant is followed by a diffusion step, which causes lateral diffusion that can significantly reduce the effective gate length. Although this approach does not require any process modification, the high-voltage devices need to be characterized and appropriate models should be generated to enable the use of these devices in practical circuits. A third approach to allow high-voltage operation is to employ special circuit structures in which individual devices operate at allowed voltage levels, but at some nodes high-voltage signals can be generated without overstressing the devices. Although the supply voltages in modern ICs constantly decrease, these ICs still need to communicate with older-generation ICs, resulting in some systems having mixed-voltage input/output (I/O) (Dabral and Maloney 1998). Due to this necessity, new circuit structures have been developed for mixed-voltage I/O that can tolerate voltages exceeding the nominal supply voltage. Some of these I/O circuits are also capable of driving the outside components at the same high-voltage levels. Cascoding or stacking is a commonly used technique in these circuits to accommodate high-voltage signals, typically twice as large as the nominal supply voltages without overstressing any of the transistors in the circuit (Annema *et al.* 2001, Singh and Salem 1999). A 7 V buffer implemented in a 2.5 V CMOS process has also been reported (Prodanov and Boccuzzi 2001).

Using the circuit design techniques mentioned above, we have demonstrated several generations of front-end ICs in standard CMOS and BiCMOS technologies to interface with several cMUT arrays for practical imaging applications (Wygant *et al.* 2004, 2005a, b). The basic building block of these circuits is shown in **Figure 17**. The images obtained using these ICs with cMUT arrays are presented in the following section.

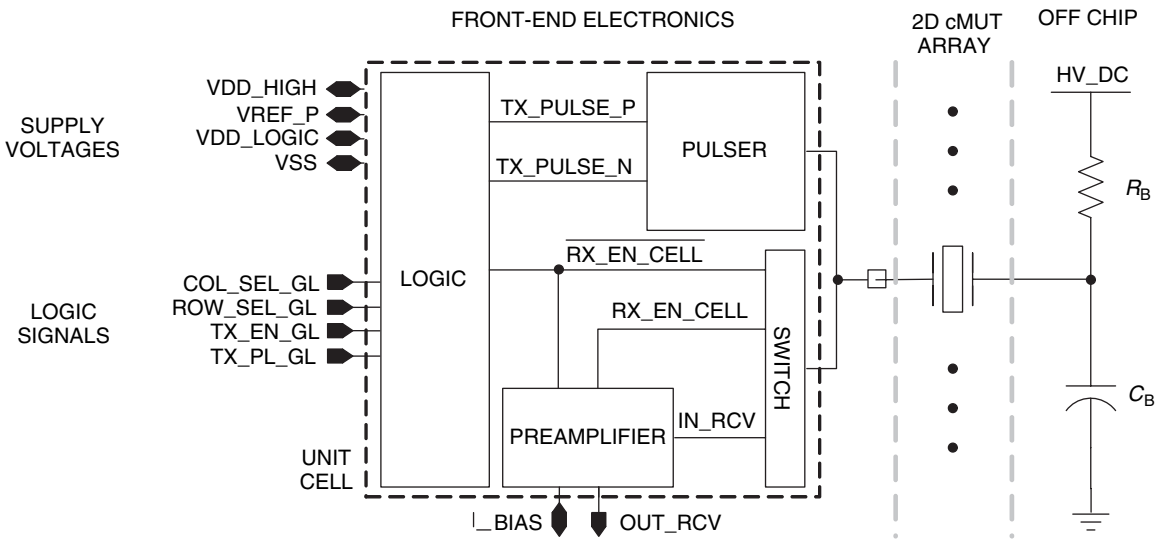


Figure 17 Basic building block of integrated front-end circuits for capacitive micromachined ultrasonic transducer (cMUTs).

2.14.5 Medical Imaging Applications

Although the medical applications of ultrasonic imaging followed the underwater and industrial applications chronologically, in terms of sophistication and impact, medical ultrasonic imaging has surpassed any other application. Ultrasonic imaging has several unique advantages compared with other medical imaging modalities. Ultrasound is a nonionizing radiation and is hence safe for diagnostic use even for obstetrical applications. The image presentation is in real time, allowing the study of moving internal structures. Ultrasonic imaging equipment is cost-effective and portable. All these features have made ultrasound a well-accepted and widely used imaging modality in medicine. Ultrasound is particularly useful in imaging cardiac structures, the vascular system, the fetus and uterus, abdominal organs such as the liver, kidneys and gall bladder, and the eye.

In this section, experimental imaging results from cMUT arrays are presented. These results demonstrate that micromachined transducer arrays integrated with supporting electronics enable many applications that have not been feasible with the existing piezoelectric transducer technology.

2.14.5.1 Conventional 2D Ultrasound Imaging

We have built 1D linear cMUT arrays of equivalent size, channel count, and operating frequency to state-of-the-art medical arrays employing piezoelectric transducers. These arrays contain 128 individually addressable elements, each 6 mm high by 200 μm wide. The element pitch is 250 μm, ideal for the chosen 3-MHz operating frequency. These arrays have been used along with a PC-based custom data acquisition system to demonstrate synthetic phased array cross-sectional images of a resolution test phantom (Oralkan *et al.* 2002). The resulting B-scan image of the resolution test target and the experimental and simulated axial and lateral point spread functions (PSFs) are shown in **Figure 18**.

2.14.5.2 Real-Time 3D Ultrasound Imaging

The implementation of real-time 3D ultrasound imaging systems relies on 2D transducer arrays. The small size of the elements in a 2D array causes degradation in acoustic output power and receive sensitivity. There are also severe difficulties in fabricating these densely populated 2D arrays and providing individual electrical connections to each element. Micromachined ultrasonic transducer technology is an ideal candidate to overcome these obstacles, preventing the realization of true 3D ultrasound imaging systems.

To demonstrate the potential of cMUTs for 3D imaging, we developed a system based on a 2D, 16×16 element, cMUT array. The target application for this system is endoscopic imaging to provide intraoperative guidance. Transducer arrays with operating frequencies ranging from 3 to 7.5 MHz

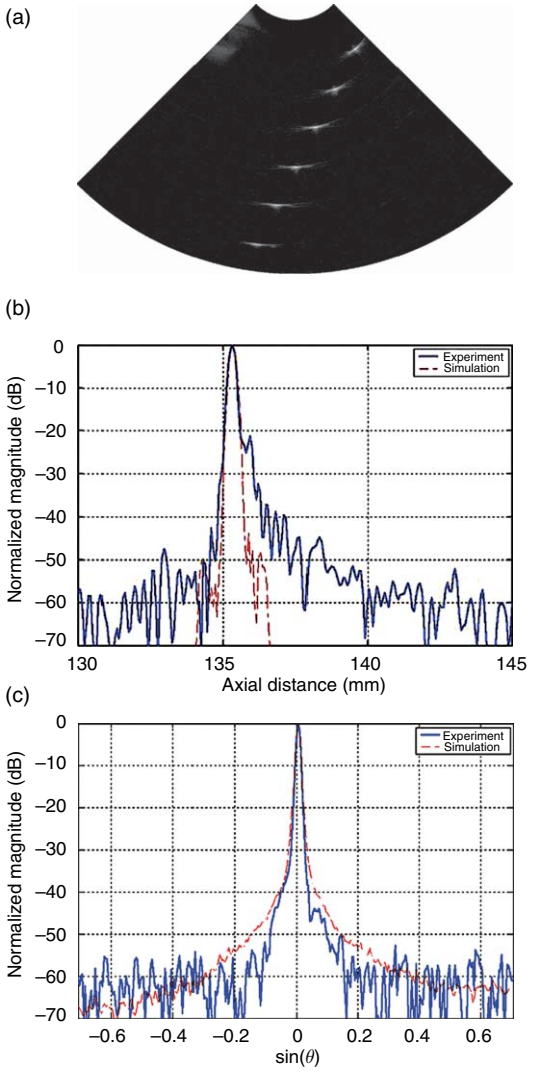


Figure 18 Two-dimensional imaging results. (a) Reconstructed B-scan sector images with a display dynamic range of 60 dB. (b) Experimental and simulated axial 1D point spread function (PSF). (c) Experimental and simulated lateral 1D PSF.

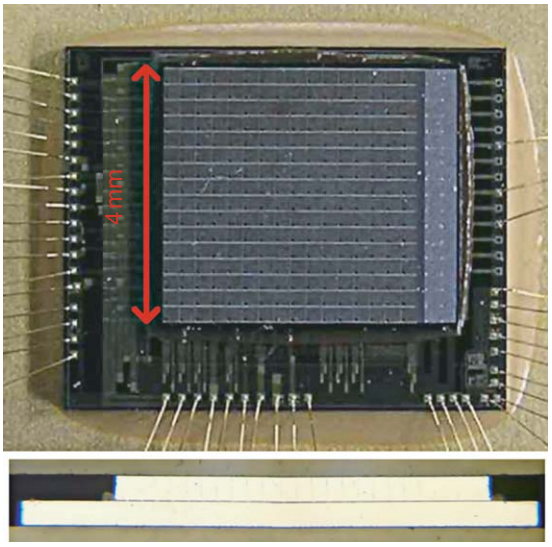


Figure 19 A 16 × 16 2D capacitive micromachined ultrasonic transducer (cMUT) array flip-chip bonded to a 16 × 16 front-end transmit/receive (T/R) integrated circuit (IC) (*top*). Cross-sectional view of the assembly (*bottom*).

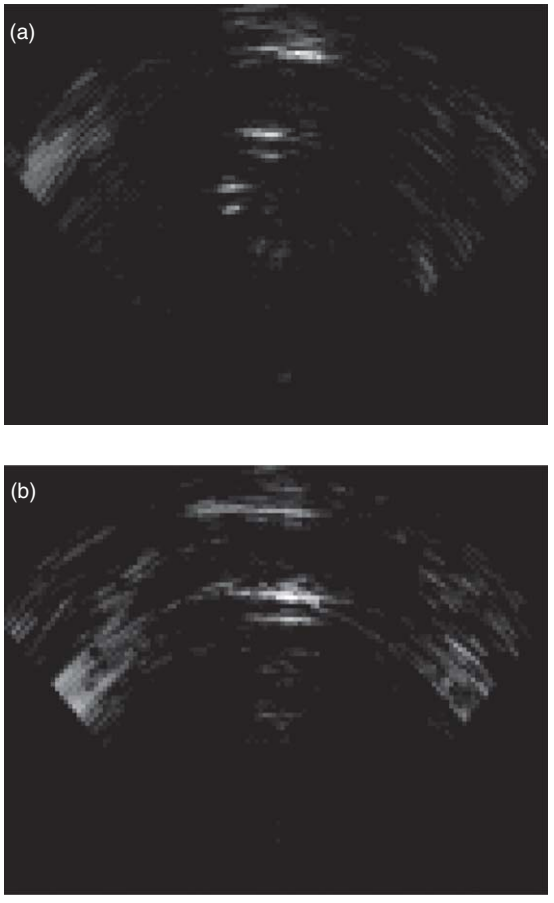


Figure 20 Cross-sectional images of the vessel phantom. (a) Central elevation plane and (b) central azimuth plane.

were fabricated for this system. The transducer array including DC bias pads measures 4 mm × 4.7 mm. The array elements are connected to flip-chip bond pads on the array back side with 400-µm-long through-wafer interconnects. The array pitch is 250 µm. The array is flip-chip bonded to a custom-designed IC that comprises the front-end electronics (**Figure 19**). Images of a vessel phantom using the described experimental synthetic aperture imaging system are shown in **Figure 20** (Wygant *et al.* 2004, 2005a, b). The system described here meets the challenges of catheter-based imaging. This prototype

system proves the feasibility of a compact probe employing a 2D array with integrated electronics.

2.14.5.3 Intravascular and Intracardiac Ultrasound Imaging

Forward-viewing ultrasound volume images are desired for many intravascular and intracardiac applications such as guiding treatment of chronic total occlusion, helping stent deployment, and monitoring ablation procedures in the heart. Because of the requirements for a guide wire in catheters, an annular ring is the preferred geometry for transducer arrays. However, it is very challenging to implement this geometry in a very small scale (1–2 mm) by using the existing piezoelectric transducer technology (Wang *et al.* 2002). On the contrary, cMUT arrays can be made in any arbitrary geometry with very small dimensions using photolithographic techniques and standard microfabrication processes. We used a 64-element, 2-mm-diameter cMUT ring array wire bonded to 64 T/R channels in a bank of four custom-designed ICs to demonstrate the feasibility of forward-looking intravascular ultrasound (**Figure 21**). In this experiment, the cMUT array has been operated in both conventional and collapse operating regimes by setting the DC bias voltage to 30 and 100 V, respectively. The resulting operating frequencies in the conventional and collapse (center of the membrane in contact with the bottom electrode) regimes are 8.5 and 19 MHz, respectively. We have imaged several phantoms including a small metal spring and deployed and undeployed forms of a Palmaz–Schatz coronary stent (Yeh *et al.* 2006). The 3D rendered images next to photographs are shown in **Figure 22**.

2.14.5.4 Photoacoustic Imaging

Traditional ultrasound images are formed by first transmitting ultrasound to a medium of interest and then receiving the ultrasound signals resulting from the interaction of the transmitted signals with the medium. This kind of an image is usually a representation of the mechanical properties of the medium and provides structural or anatomical information. The interaction of the medium with other forms of energy can provide additional information about the functional differences even in a structurally indifferent, uniform medium. For instance, when a short laser pulse is transmitted into a tissue, the introduced light energy is absorbed and scattered in a different

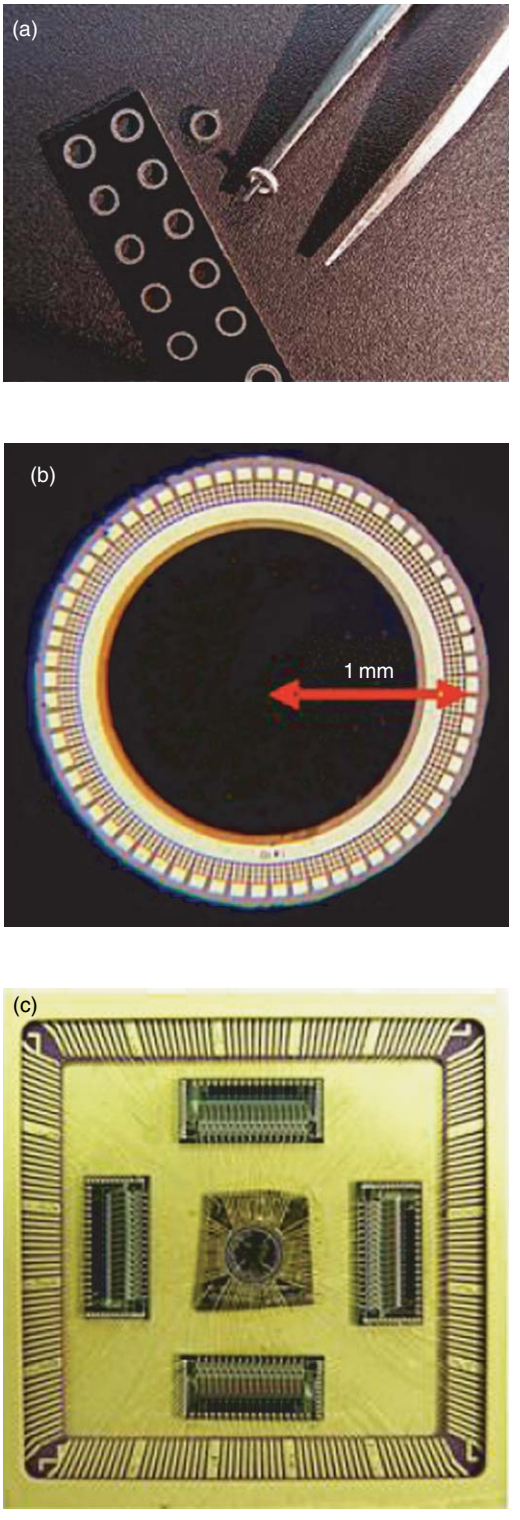


Figure 21 (a) Several capacitive micromachined ultrasonic transducer (cMUT) annular ring arrays. (b) Magnified view of a cMUT annular ring array. (c) cMUT annular ring array wire bonded to custom-designed front-end integrated circuits (ICs).

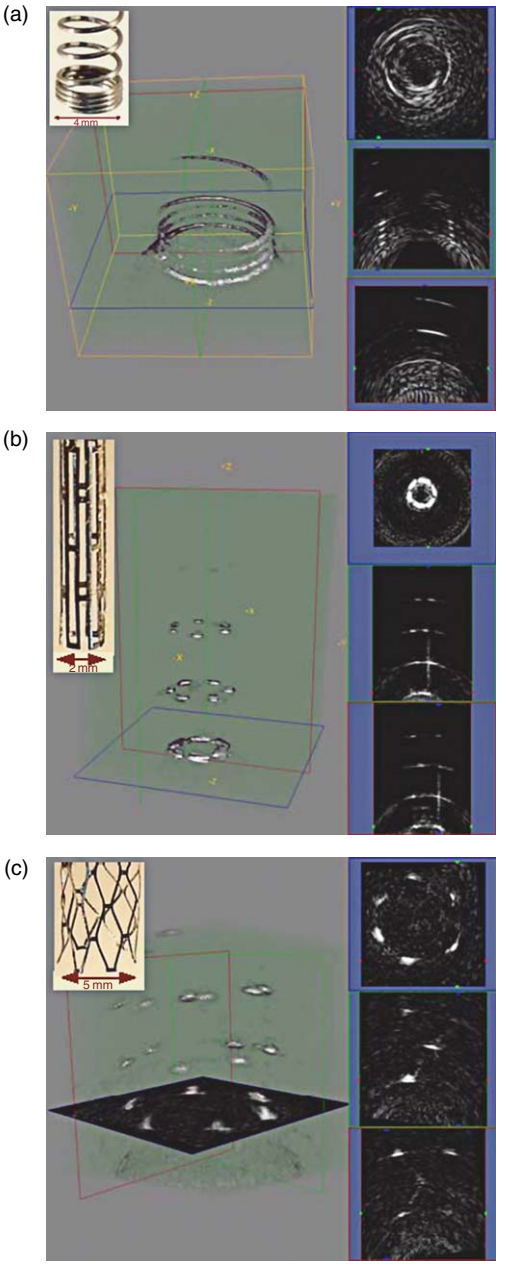


Figure 22 Three-dimensional rendered ultrasound images and some cross sections of (a) a small metal spring, (b) an undeployed Palmaz–Schatz stent, and (c) a deployed Palmaz–Schatz stent.

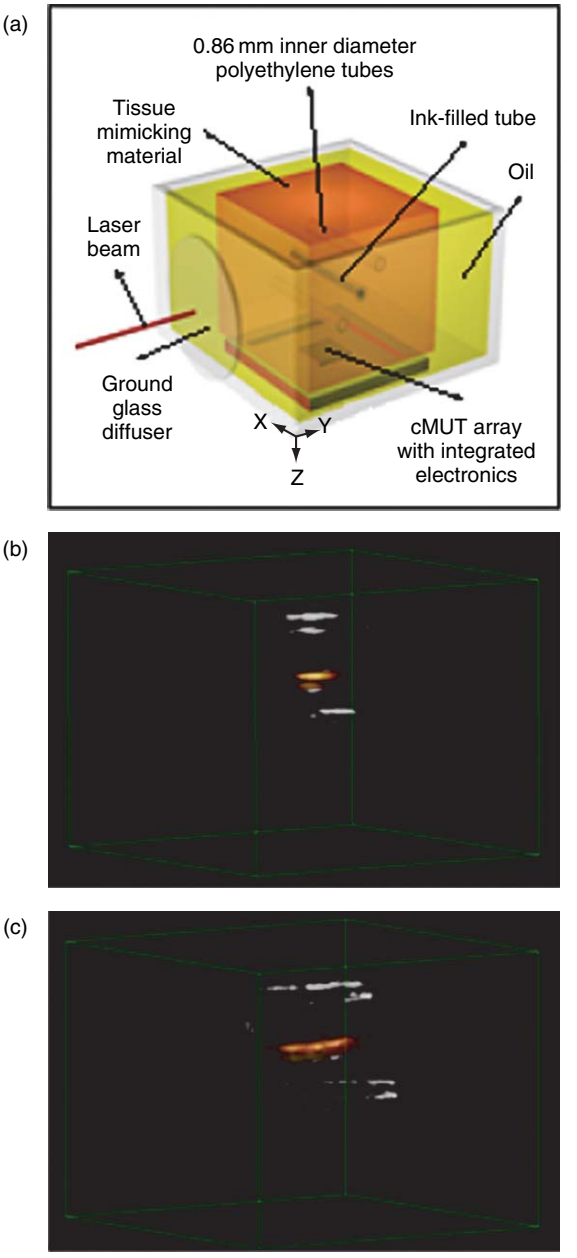


Figure 23 (a) Experimental photoacoustic imaging setup diagram; superimposed structural and functional images. Traditional pulse echo images are shown in grayscale. Photoacoustic images are shown in red. (b) Received by the 16×16 aperture. (c) Received by the 48×48 aperture.

manner by different parts of the tissue. The optical absorption depends on the wavelength of the light and the properties of the medium at the molecular or even the atomic level (Esenaliev *et al.* 1999). The regions with stronger absorption characteristics in a tissue generate stronger acoustic signals via the thermoelastic effect, which is simply the thermal expansion of the imaging regions resulting in a mechanical disturbance and hence an acoustic signal. By collecting these light-induced acoustic signals using a transducer or array of transducers, one can construct an image that is a representation of light

absorption characteristics of the sample. One example of this approach is to image the microvasculature in tissue by detecting blood oxygenation, which is usually a sign of angiogenesis indicating a cancerous lesion (Ku *et al.* 2005). In this example, the increased light absorption of the oxygenated blood is used to create a high-contrast image.

Photoacoustic imaging has typically relied on a single mechanically scanned focused piezoelectric transducer for detection of the laser-generated ultrasound. Using a cMUT array in place of a mechanically scanned element has a number of advantages. Three-dimensional images can be acquired in one shot using large 2D arrays that can be reliably fabricated using cMUT technology. We demonstrated the potential of cMUTs for photoacoustic imaging by using a 2D array (Vaithilingam *et al.* 2006, Wygant *et al.* 2005c). The experimental setup is shown in **Figure 23(a)**. Both structural reflection-mode ultrasound images and functional photoacoustic images are obtained using a 16×16-element and 48×48-element aperture by using mechanical scanning. The overlaid images in **Figure 23(b)** and **23(c)** show that the optical contrast of ink is only visible in the photoacoustic image (red); structural information about all the three tubes is captured by the traditional reflection-mode ultrasound image.

2.14.6 Conclusions

Since the introduction of cMUT technology in the early 1990s, significant milestones have been achieved in modeling, fabrication, and applications. Today, the analytical and numerical models can reliably be used to design transducer arrays in a frequency range from tens of kilohertz to tens of megahertz. The fabrication techniques are capable of producing repeatable and reliable devices. Integration with supporting electronic circuits using monolithic and multichip approaches has been demonstrated. The once challenging applications such as real-time 3D imaging and forward-looking intravascular imaging have proven feasible with this new technology.

Extensive research efforts on cMUTs spanning more than a decade have finally brought this technology to a point where large arrays can be manufactured reliably, and a consensus has been reached that these transducer arrays indeed provide superior image quality. However, underlying fabrication techniques make the cMUT still open for further performance and cost improvements.

The future research on cMUTs can be directed toward increased levels of integration, higher operating frequencies, lowering the cost, providing therapeutic tools, or combinations of these depending on the application of interest. All these different directions have one thing in common: All can benefit from the use of cMUT technology. The versatility of this technology makes it especially attractive for implementing future generations of ultrasonic imaging systems.

References

Annema A-J, Geelen G J G M, de Jong P C 2001 5.5-V I/O in a 2.5-V 0.25-μm CMOS technology. *J. Solid State Circuits* **36**, 528–38

Ballan H, Declercq M 1999 *High Voltage Devices and Circuits in Standard CMOS Technologies*. Kluwer Academic Publishers, Dordrecht, The Netherlands

Brunner E 2002 How ultrasound system considerations influence front-end component choice. *Analog Dialog* **36**, 1–4

Caliano G, Galanello F, Caronti A, Carotenuto R, Pappalardo M, Foglietti V, Lamberti N 2000 Micromachined ultrasonic transducers using silicon nitride membrane fabricated in PECVD technology. *Proc. IEEE Ultrasonics Symposium,* San Juan, Puerto Rico, pp. 963–8

Dabral S, Maloney T J 1998 *Basic ESD and I/O Design*. John Wiley & Sons, New York, NY

Daft C, Calmes S, da Graca D, Patel K, Wagner P, Ladabaum I 2004 Microfabricated ultrasonic transducers monolithically integrated with high voltage electronics. *Proc. IEEE Ultrasonics Symposium,* Montreal, Quebec, Canada, pp. 493–6

Eccardt P C, Niederer K, Fischer B 1996 Micromachined transducers for ultrasound applications. *Proc. IEEE Ultrasonics Symposium,* San Antonio, Texas, USA, pp. 1609–18

Eccardt P C, Niederer K, Fischer B 1997 Micromachined transducers for ultrasound applications. *Proc. IEEE Ultrasonics Symposium,* Toronto, Ontario, Canada, pp. 1609–18

Esenaliev R O, Karabutov A A, Oraevsky A A 1999 Sensitivity of laser opto-acoustic imaging in detection of small deeply embedded tumors. *IEEE J. Sel. Top. Quant. Electron.* **5**, 981–8

FDA 1997 *Information for Manufacturers Seeking Marketing Clearance of Diagnostic Ultrasound Systems and Transducers*. US Department of Health and Human Services Food and Drug Administration Center for Devices and Radiological Health

Haller M I, Khuri-Yakub B T 1994 A surface micromachined electrostatic ultrasonic air transducer. *Proc. IEEE Ultrasonics Symposium,* Cannes, France, pp. 1241–4

Hohm D, Hess G 1989 A subminiature condenser microphone with silicon nitride membrane and silicon backplate. *J. Acoust. Soc. Am.* **85**, 476–80

Kasulke P, Schmidt W, Titerle L, Bohnaker H, Oppert T, Zakel E 1998 Solder ball bumper SB^2-A flexible manufacturing toll for 3-dimensional sensor and microsystem packages. *Proc. 22nd International Electronics Manufacturing Technology Symposium,* Berlin, Germany: April 27–29, 1998

Khuri-Yakub B T, Cheng H, Degertekin F L, Ergun S, Hansen S, Jin X C, Oralkan O 2000 Silicon micromachined ultrasonic transducers. *Jpn. J. Appl. Phys.* **39**, 2883–7

Knight J G, Degertekin F L 2003 Fabrication and characterization of CMUTs for forward looking intravascular ultrasound imaging. *Proc. IEEE Ultrasonics Symposium,* Honolulu, Hawaii, USA, pp. 1175–8

Ku G, Wang X, Xie X, Stoica G, Wang L V 2005 Imaging of tumor angiogenesis in rat brains in vivo by photoacoustic tomography. *Appl. Opt.* **44(5)**, 770–5

Ladabaum I, Jin X, Soh H T, Atalar A, Khuri-Yakub B T 1998 Surface micromachined capacitive ultrasonic transducers. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **45**, 678–90

Lohfink A, Eccardt P-C, Benecke W, Meixner H 2003 Derivation of a 1D CMUT model from FEM results for linear and nonlinear equivalent circuit simulation. *Proc. IEEE Ultrasonics Symposium*, Vol.1, Honolulu, Hawaii, USA, pp. 465–8

Maluf N I, Reay R J, Kovacs G T A 1996 High-voltage devices and circuits fabricated using foundry CMOS for use with electrostatic MEM actuators. *Sens. Actuators A Phys.* **52**, 187–92

Mason W P 1948 *Electromechanical Transducers and Wave Filters*. D Van Nostrand Company, London

Niederer K, Eccardt P-C, Meixner H, Lerch R 1999 Micromachined transducer design for minimized generation of surface waves. *Proc. IEEE Ultrasonics Symposium,* Lake Tahoe, Nevada, USA, pp. 1137–9

Noble R A, Davies R R, Day M M, Koker L, King D O, Brunson K M, Jones A R D, McIntosh J S, Hutchins D A, Robertson T J, Saul P 2001 A cost-effective and manufacturable route to the fabrication of high-density 2D micromachined ultrasonic transducer arrays and (CMOS) signal conditioning electronics on the same silicon substrate. *Proc. IEEE Ultrasonics Symposium,* Atlanta, Georgia, USA, pp. 941–4

Noble R A, Davies R R, King D O, Day M M, Jones A R D, McIntosh J S, Hutchins D A, Saul P 2002 Low-temperature micromachined cMUTs with fully-integrated analogue front-end electronics. *Proc. IEEE Ultrasonics Symposium,* Munich, Germany, pp. 1045–50

Oralkan Ö, Ergun A S, Johnson J A, Karaman M, Demirci U, Kaviani K, Lee T H, Khuri-Yakub B T 2002 Capacitive micromachined ultrasonic transducer arrays: Next generation arrays for acoustic imaging? *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **49(11)**, 1596–610

Patton C A, Harris G R, Phillips R A 1994 Output levels and bioeffects indices from diagnostic ultrasound exposure data reported to the FDA. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **41**, 353–9

Prodanov V, Boccuzzi V 2001 7 V tristate-capable output buffer implemented in standard 2.5 V CMOS process. *Proc. Custom Integrated Circuits Conference,* San Diego, California, USA, pp. 497–500

Rafiq M, Wykes C 1991 The performance of capacitive ultrasonic transducers using v-grooved backplates. *Meas. Sci. Technol.* **2**, 168–74

Robinson A L, Mo J-H 1992 Applications of microelectronics and microfabrication to ultrasound imaging systems. *Proc. IEEE Ultrasonics Symposium,* Tuczon, Arizona, USA, pp. 681–91

Schafer M E, Lewin P A 1984 The influence of front-end hardware on digital ultrasonic imaging. *IEEE Trans. Sonics and Ultrasonics* **SU-31**, 295–306

Schindel D W, Hutchins D A 1995 The design and characterization of micromachined air-coupled capacitance transducers. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **UFFC-42**, 42–50

Singh G P, Salem R B 1999 High-voltage-tolerant I/O buffers with low voltage CMOS process. *J. Solid State Circuits* **34**, 1512–25

Suzuki K, Higuchi K, Tanigawa H 1989 A silicon electrostatic ultrasonic transducer. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **UFFC-36**, 620–7

Vaithilingam S, Wygant I O, Kuo P S, Zhuang X, Oralkan Ö, Olcott P D, Khuri-Yakub B T 2006 Capacitive micromachined ultrasonic transducers (CMUTs) for photoacoustic imaging. *Proc. SPIE Photons Plus Ultrasound: Imaging and Sensing 2005: Sixth Conf. on Biomedical Thermoacoustics, Optoacoustics, and Acousto-Optics,* San Jose, California, USA, Vol. 6086. pp. 1–11

Wang Y, Stephens D N, O'Donnell M 2002 Optimizing the beam pattern of a forward-viewing ring-annular ultrasound array for intravascular imaging. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **49**, 1652–64

Wygant I O, Zhuang X, Yeh D T, Nikoozadeh A, Oralkan Ö, Ergun A S, Karaman M, Khuri-Yakub B T 2004 Integrated ultrasonic imaging systems based on cMUT arrays: Recent progress. *Proc. IEEE Ultrasonics Symposium,* Montreal, Quebec, Canada, pp. 391–4

Wygant I O, Yeh D T, Zhuang X, Nikoozadeh A, Oralkan Ö, Ergun A S, Karaman M, Khuri-Yakub B T 2005a A miniature real-time volumetric ultrasonic imaging system. *Proc. SPIE Medical Imaging Conference,* San Diego, California, USA, pp. 26–36

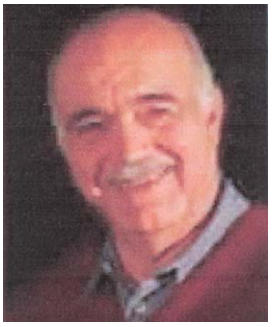
Wygant I O, Zhuang X, Yeh D T, Vaithilingam S, Nikoozadeh A, Oralkan Ö, Ergun A S, Karaman M, Khuri-Yakub B T 2005b An endoscopic imaging system based on a two-dimensional CMUT array: Real-time imaging results. *Proc. IEEE Ultrasonics Symposium,* Rotterdam, The Netherlands, pp. 792–5

Wygant I O, Zhuang X, Kuo P S, Yeh D T, Oralkan Ö, Khuri-Yakub B T 2005c Photoacoustic imaging using a two-dimensional CMUT array. *Proc. IEEE Ultrasonics Symposium,* Rotterdam, The Netherlands, pp. 1921–4

Yaralioglu G G, Badi M H, Ergun A S, Khuri-Yakub B T 2003 Improved equivalent circuit and finite element method modeling of capacitive micromachined ultrasonic transducers. *Proc. IEEE Ultrasonics Symposium,* Honolulu, Hawaii, USA, pp. 469–72

Yeh D T, Oralkan Ö, Wygant I O, O'Donnell M, Khuri-Yakub B T 2006 3-D ultrasound imaging using a forward-looking CMUT ring array for intravascular/intracardiac applications. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **53(6)**, 1202–11

Biographies



Butrus (Pierre) T. Khuri-Yakub was born in Beirut, Lebanon. He received a B.S. degree in 1970 from the American University of Beirut, an M.S. degree in 1972 from Dartmouth College, and a Ph.D. degree in 1975 from Stanford University, all in Electrical Engineering. He joined the research staff at the E.L. Ginzton Laboratory of Stanford University in 1976 as a research associate. He was promoted to Senior Research Associate in 1978 and to Professor of Electrical Engineering in 1982. He has served on many university committees such as Graduate Admissions and Undergraduate Academic Council of the School of Engineering.

Professor Khuri-Yakub has been teaching at both the graduate and undergraduate levels for over 15 years, and his current research interests include *in situ* acoustic sensors (temperature, film thickness, resist cure, etc.) for the monitoring and control of integrated circuit manufacturing processes, micromachining silicon to make acoustic materials and devices such as airborne and water immersion ultrasonic transducers and arrays, and fluid ejectors, and in the field of ultrasonic nondestructive evaluation and acoustic imaging and microscopy.

Professor Khuri-Yakub is a fellow of the IEEE, a senior member of the Acoustical Society of America, and a member of Tau Beta Pi. He is an associate editor of *Research in Nondestructive Evaluation*, a journal of the American Society for Nondestructive Testing, as well as a member of AdCom of the IEEE Group on Ultrasonics Ferroelectrics and Frequency Control (January 1, 1994–January 1, 1997).

Professor Khuri-Yakub has authored over 450 publications and has been principal inventor or co-inventor on over 72 patents. He received the American University of Beirut School of Engineering Distinguished Alumni Award, May 2005, Stanford University Outstanding Inventor Award, June 2004, IEEE UFFC Distinguished Lecturer Award, 1999, Stanford University School of Engineering Distinguished Advisor Award, June 1987, and the Medal of the City of Bordeaux for contributions to NDE, 1983.

Arif Sanli Ergun was born in Ankara, Turkey, in 1969. He received his B.Sc., M.Sc., and Ph.D. degrees in 1991, 1994, and 1999, respectively, all in electrical and electronics engineering, from Bilkent University, Ankara, Turkey. He was a research assistant in Bilkent University between 1991 and 1999, and an engineering research associate at E. L. Ginzton Laboratory, Stanford University, Stanford, CA, between 2000 and 2006. He is currently a research scientist at Siemens Corporate Research.

His research interests are microwave electronics, ultrasound, ultrasound imaging, and MEMS/NEMS, with an emphasis on capacitive micromachined ultrasonic transducers (cMUTs). He is a member of the IEEE and the Electron Devices Society.

Ömer Oralkan received the BS degree from Bilkent University, Ankara, Turkey, in 1995, the MS degree from Clemson University, Clemson, SC, in 1997, and the Ph.D. degree from Stanford University, Stanford, CA, in 2004, all in electrical engineering.

He joined the research staff at the E. L. Ginzton Laboratory of Stanford University in 2004 as an engineering research associate. He was promoted to the rank of Senior Research Engineer in 2007. His past and present research interests include analog and digital circuit design, semiconductor device physics and fabrication, micromachined sensors and actuators, and medical imaging. His current research focuses on the design and implementation of integrated systems for catheter-based medical imaging applications, photoacoustic imaging, and chemical and biological sensor arrays.

Dr. Oralkan has authored and coauthored over 60 publications and received the 2002 Outstanding Paper Award of the IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society. He is a member of the IEEE, SPIE, and AIUM.

Goksen Goksenin Yaralioglu was born in Akhisar, Turkey, in 1970. He received his BS, MS and Ph.D. degrees from Bilkent University, Ankara, Turkey, in 1992, 1994, and 1999, respectively, all in electrical engineering. From 2000 to 2006 he was an engineering research associate in the E. L. Ginzton Laboratory, Stanford University. Currently, he is a senior MEMS engineer at Invensense, Inc., Santa Clara, CA. His past and present research interests include design, modeling, and applications of micromachined transducers, atomic force microscopy at ultrasonic frequencies, and acoustic microscopy.

2.15 Tissue Engineering

Jeffrey T. Borenstein, Draper Laboratory, Cambridge, MA, USA

2.15.1 Introduction

The field of tissue engineering has emerged in the 1980s in response to an ever-increasing shortage of donor tissues and organs available to patients suffering from end-stage tissue and organ loss (Langer and Vacanti 1993, Vacanti and Langer 1999). Advances in surgical techniques and transplantation medicine have enabled clinicians to save the lives of hundreds of thousands of patients suffering from failure of vital organs such as the heart, the lungs, or the liver. An enormous number of people suffer from the loss of other tissues such as skin, bone, and cartilage that results from injuries such as burns, fractures, or overuse or is caused by illnesses such as diabetes. For these patients, donor tissues can provide a permanent solution but may generate additional complications depending on the source of the replacement tissue. It is the goal of the field of tissue engineering to provide an unlimited source of replacement tissues and organs for transplantation, thereby easing the growing crisis facing the nation's health care system. In this chapter, the emerging role of microsystems technology as a platform for tissue engineering is described.

2.15.1.1 Clinical Need for Engineered Tissues

Pioneering advances in the 20th century established transplantation of cadaver organs as a viable and often lifesaving treatment for patients suffering from end-stage organ failure. Kidney, heart, liver, and lung transplants, for instance, are now common and often lead to long-term survival of the recipient. However, these patients have to endure increasingly long waiting periods for a suitable donor organ and often do not survive till then. In addition, the use of allograft tissues and organs necessitates the use of immunosuppressive drugs in order to prevent donor graft rejection; these compounds are both expensive and responsible for severe side effects. The numbers of US patients waiting for a donor organ to become available for specific organ types are summarized in **Table 1**.

In parallel with advances in the transplantation of donor organs, a revolution in the development and use of mechanical organ assist and organ replacement devices is under way. For those patients suffering from kidney failure, renal dialysis provides a means for removing toxins from the blood and is capable of

Table 1 June 29, 2007 data on size of waiting list in United States for waiting list candidates for vital organ transplants. Source the Organ Procurement and Transplantation Network (OPTN)

	All organs	*Kidney*	*Liver*	*Pancreas*	*Kidney/pancreas*	*Heart*	*Lung*	*Heart/lung*	*Intestine*
Registrations	96,583	72,372	16,884	1,684	2,329	2,697	2,757	117	234

sustaining patients for many years. Such extracorporeal approaches for liver and lung assist devices may provide a bridge to transplant but do not represent viable long-term strategies. The use of extracorporeal membrane oxygenators (ECMOs) and more recently intravascular oxygenators has provided cardiopulmonary support to neonates and to pediatric and adult patients suffering from acute lung or heart failure. Despite extraordinary progress in mechanical assist devices such as the totally implantable mechanical heart, these mechanical organ replacement devices still face many difficult challenges in providing safe and effective short-term therapy. Ultimately, such mechanical devices may never represent a long-term viable strategy for organ replacement because of foreign body reactions and logistical challenges related to the size, power requirements, and operational incompatibility of the devices with human physiology.

Approaches extending beyond cadaveric organ transplantation and mechanical devices are also available for many organs and tissues. For example, split liver transplants have enabled genetically related individuals to provide a part of their organ to a recipient, by taking advantage of the regenerative capacity of the liver so as to ensure long-term liver function in the donor. However, such partial organ transplants remain rare and present the unfortunate possibility of morbidity or mortality for the healthy donor. In addition, a patient with a burn injury, a diabetic ulcer, or a damaged blood vessel, for instance, may provide an autograft for transplant from a different site in their body. These approaches, though common, do present a significant problem for the patient in terms of donor site morbidity. Another avenue for addressing the shortage of donor organs involves the investigation of animal organs as a potential source for transplants (Seebach *et al.* 2001). Known as xenotransplantation, this field focuses principally on the use of pig (porcine) organs, because of phenotypic and physiological similarities between porcine and human tissue. The principal obstacle to the use of xenotransplantation involves genetic differences between humans and pigs and the possibility of passing genetic viruses between species.

The challenges facing the various routes to addressing the shortage of donor tissues and organs have spawned the field of tissue engineering (Berthiaume and Yarmush 1995, Kaplan *et al.* 2005, Lysaght and Reyes 2001), where cell-based grafts and devices are generated in the laboratory and transplanted into the recipient. Most of the advances in tissue engineering in the 1990s have involved simpler tissue systems such as skin and cartilage. These structures are simpler than vital organs for several reasons: they involve fewer cell types, possess less complex microstructures and microenvironments, and often do not require an intrinsic vascular supply. Tubular structures such as the intestine, ureter, pulmonary artery, and even the bladder have been constructed in the laboratory using tissue engineering (Mooney *et al.* 1994, Niklason *et al.* 1999) and have been transplanted successfully into patients. These examples provide a glimpse of what will one day be possible as the field of tissue engineering advances, enabled in part by microsystems technology as will be described in later sections.

2.15.1.1.1 Shortage of tissues and organs for transplant medicine

Tissue loss and failure represent enormous challenges in terms of impact on quality of life and economic burden (Langer and Vacanti 1993, Palsson and Bhatia 2004). While organ failure is more directly associated with patient mortality, for elderly patients even seemingly non-life-threatening events such as a hip fracture are associated with high death rates within 2 years of transplantation. For younger patients, the loss of skin, bone, connective tissues, or muscle can present devastating consequences such as immobility and its long-term complications. Over 1 million burn patients in the United States would benefit annually from an engineered tissue solution to heal the damaged site, according to the American Burn Association. Pressure sores or pressure ulcers in bedridden patients, vascular disorders, and diabetes are responsible for another 1.25 million US patients per year. Nerve and spinal cord defects present another challenging and tragic situation amenable to an engineered tissue solution. Nearly 1.5 million patients in the United States suffer from ailments requiring bone grafts, hip and other joint replacements, or internal fixation. In terms of connective tissues, enormous numbers of patients undergo knee surgeries involving the patella or meniscus, as well as tendon and ligament surgeries. At the national level, millions of patients suffer from the loss of teeth and dental tissues, potentially resulting in nutritional deficiencies and other complications. In all of these cases, the ability to generate healthy replacement tissue in the laboratory would revolutionize patient care.

An ever-increasing shortage of donor organs has led to a crisis in the nation's health care system. Over

100 000 patients are now listed on the UNOS list of those waiting for an organ transplant. Each year, although 30 000 patients with liver failure caused by diseases such as cirrhosis and hepatitis await organ transplants, only 3000 livers are transplanted. Even for those patients fortunate enough to receive a donor organ, the potential for immune rejection often leads to a lifetime of expensive and potentially dangerous immunosuppressive drugs. Similar crises exist for patients with lung and heart failure. Over 1 million patients suffer from cardiac and blood vessel ailments for which engineered tissue replacements would provide critically needed long-term solutions. Tubular structures such as the intestine, the urethra, and the bladder are the subjects of intensive efforts to develop engineered tissue solutions.

The explosion of type II diabetes, as a result of the aging population and the surge in obesity and metabolic disorders, has further contributed to the critical shortage of tissues and organs. This is manifested directly by the need for healthy pancreatic islet cells as well as by diabetic complications that damage the heart, blood vessels, and kidneys. More than 500 000 patients suffer from kidney diseases requiring long-term support or transplantation. As described in Section 2.15.1.1.2, renal dialysis provides a chronic approach for organ assist but has many drawbacks in terms of quality of life and long-term prognosis. Finally, neurodegenerative diseases such as Alzheimer's disease and Parkinson's disease present opportunities for cell-based therapies that may not comprise conventional engineered tissues such as skin or organ replacement but may also benefit from microsystems-based approaches for cell transplantation that are synergistic with the microtechnology platforms described herein.

2.15.1.1.2 Organ assist devices

The most prevalent organ assist device in use is the kidney dialyzer. A typical therapeutic regimen involves three 4-h treatments per week, in a clinical setting such as a dialysis center. The requirement for such expensive and frequent treatments to occur in the clinic presents enormous challenges for both the patient and the health care provider. Furthermore, the long-term prognosis for kidney dialysis patients suffering from end-stage renal failure is grim, with a 20% 5-year survival rate and staggering complications because of the cyclic nature of toxin levels in the blood as well as inflammatory responses. Pioneering work by Kolff, Berk and colleagues (Kolff, Berk *et al.* 1944) resulted in the first kidney dialysis machines, which were bulky and complex, but demonstrated the ability to sustain life in patients. These devices typically operate by passing the patient's blood through a series of nanoporous vessels, in the manner of an external microcirculation. Small molecules such as urea and creatinine, which are toxic at high levels, are removed from the blood, as in a healthy kidney. In addition, other small molecules such as electrolytes and glucose, as well as large amounts of water, are also removed. In a healthy kidney, these blood components are reabsorbed via active transport to maintain homeostasis and the fluid–electrolyte and nutritive balances. In kidney dialysis, the cartridge serves only as a passive filter, and therefore, essential blood components must be manually restored to the patient.

In the past several decades, several generations and versions of kidney dialyzers have been developed and demonstrated, using configurations involving hollow fibers, flat plates, spirals, and rolled tubes. In each case, the dimensions of the smallest structures within the cartridge are much larger than those found within the nephrons of the kidney, principally because of limitations in the technology used for constructing the dialysis cartridge. For instance, the small vessels in the glomerulus are in the range of 10 μm in diameter (Guyton and Hall 2000), whereas hollow fibers have diameters of several hundred microns. Therefore, the clearance efficiency of small molecules is dramatically reduced because of longer diffusion distances in conventional dialysis, presenting an opportunity for the development of a microelectromechanical systems (MEMS)-based renal dialysis system using microsystems technology (Kaazempur-Mofrad *et al.* 2004). Efforts toward a tissue-engineered kidney (Humes 1996) for acute renal failure (ARF) are also progressing rapidly.

Liver assist devices are less well developed than renal assist systems, principally because renal function can be substantially replaced by passive filtration and manual replacement of essential blood components, whereas liver function cannot (Allen and Bhatia 2002). In the case of the liver, protein and drug metabolism represent very complex functions that are only a fraction of the full spectrum of the key roles played by the liver. The complexity of these functions necessitates the use of viable liver cells in most systems; however, some liver assist systems do not contain living cells or tissues. In either case, liver assist devices have been demonstrated to be efficacious only in short-term applications such as recovery from acute liver failure or as a bridge to

transplantation. Microsystems technologies have played a key role in the understanding of liver cell function in *ex vivo* and laboratory systems (Allen and Bhatia 2002), as will be described later.

Cardiac assist devices are plentiful and have been highly successful. The totally implantable artificial heart represents the most challenging and complex of these assist devices (Dowling *et al.* 2000). Other systems include pacemakers, implantable cardioverter defibrillators (ICDs), and ventricular assist devices. In each case, these devices require sources of power, electronics and mechanical components, and the capability to replace one or more aspects of normal physiological cardiac function. Microsystems technologies have already enabled dramatic advances in cardiac assist devices, such as in the case of the ICD, where a MEMS accelerometer is used to detect the patient's motions as part of the feedback circuit that controls the activation of electrical stimulation.

Artificial lung devices have evolved over the past several decades and implantable versions are now being developed (Lund *et al.* 2002). These systems provide cardiopulmonary support by supplying oxygen from the blood through a membrane oxygenator. Microsystems technologies are now being explored to provide smaller versions of the membrane oxygenators, with smaller dimensions for the circulatory channels, as well as microtechnology-based approaches to provide novel oxygen sources for the device. In the cases of heart, lung, liver, and kidney assist devices, as well as for other applications, the central goal is to replace organ function in a manner that is as physiologically compatible and realistic as possible. Microsystems technologies are playing an increasingly important role because of their ability to miniaturize structural components and overall systems, as well as because of their potential for smart systems containing sensors for autonomous operation and real-time control.

2.15.1.1.3 Platforms for drug and vaccine discovery

Cell-based systems for transplantation as a therapeutic avenue for end-stage tissue or organ failure represent the largest set of potential applications for tissue engineering, but by no means the only one. In addition to generating devices for surgical transplantation, tissue engineering is capable of providing laboratory-based systems that mimic the behavior of human tissue *in vitro*. Initially, these systems were used as basic research and discovery tools, but there is increasing interest in the use of engineered human tissues as a platform for testing of drug safety and efficacy, and as a tool for high-throughput assays for small molecules, polymers, and genomes. Vaccine discovery is another opportunity for engineered tissues based on human cells, because of the well-known problems with efforts to develop vaccines based on safety and efficacy information derived from animal studies. Progress toward an engineered thymic organoid (Poznansky *et al.* 2000), for instance, relies heavily on the ability of a microstructured scaffold to replicate the microenvironment in which B cells and T cells gain their targeting abilities against specific antigens. Microfabrication technology provides an avenue for producing a microenvironment that mimics that found in human physiology, because of the relevant dimensions of cells and organ microstructures at size scales of roughly 10 μm. In addition, the batch processing aspect of microfabrication techniques enables low-cost replication of microfluidic cell-based assays with fluorescent readouts, thereby providing a path for high-throughput screening.

2.15.1.2 Convergence of Microsystems and Biology

Early efforts in the field of microsystems utilized silicon-based microsensors comprising mechanical and electrical components on a chip that were used for measuring pressure, flow, or inertial motion. As the field progressed, applications in the life sciences began to emerge, but even here the first products were still silicon-based microelectromechanical sensors for blood pressure or other basic measurements. However, since the late 1990s, microsystems technologies have become an integral part of life sciences and biomedical research (Khademhosseini *et al.* 2006), because of their ability to interact with and control biological materials such as cells and proteins, and because microfabrication techniques capable of constructing 3D biological systems have now been developed. Before the advent of microsystems technologies and microfabrication techniques, it was not possible to engineer the microenvironments responsible for biological processes such as development and reproduction, injury, disease, and regeneration. The resolution and scalability of microfabrication technology, combined with its flexibility with regard to materials and chemistry, make it the ideal toolset for exploring biological systems.

Environmental cues associated with triggering cellular processes such as adhesion, migration, proliferation, differentiation, and apoptosis have been

difficult to reproduce *in vitro* because of the timescales and spatial dimensions involved (Liu and Chen 2005). Microfabrication-based tools such as replica molding, microcontact printing, and photodefinable matrices have been successfully applied to the investigation of spatiotemporal cues for cellular processes and ultimately have been used to construct engineered tissues and organ replacement devices by mimicking physiological systems at the macroscale and at the microscale. These environmental cues take several forms, including those associated with the matrix and the substrate, those associated with soluble factors, cell–cell interactions, and dynamic effects related to fluid stresses and other time-varying phenomena.

2.15.1.3 Investigations of Interactions of Microstructures and Cells

Topographic control of cell behavior has been investigated for several decades, first at the microscale but more recently at the submicron level, as investigators have discovered that nanoscale phenomena in numerous tissues are critical in determining cell behavior and tissue development (Flemming *et al.* 1999). Other interactions between substrates or matrices and cells have been probed using microfabrication techniques. The most typical means of controlling cell–substrate interactions has been through the use of chemical deposition or modification to create patterns of adhesive and nonadhesive regions (Bhatia *et al.* 1997, 1998a, b). Modulation of the size and distribution of adhesive regions has been shown to control cell-specific functions and proliferation (Chen *et al.* 1997). These pathways appear to be switched on and off by varying the spatial distribution of surface adhesion molecules, often through the use of self-assembled monolayers (SAMs). Limitations in the flexibility of SAM processes with regard to surface chemistry have led to the emergence of several other microsystems-based methods for patterning and engineering surfaces. For instance, microcontact printing with proteins has been exercised in concert with polyethylene glycol (PEG) surface coatings to generate adhesive and nonadhesive regions (Albrecht *et al.* 2006). Optical tweezers-based methods have been used to trap cells and form organized cellular arrays, as have been electric- and magnetic field-based approaches such as dielectrophoresis (Voldman *et al.* 2002).

Interactions between cells are known as homotypic interactions if the cells are of the same phenotype and as heterotypic interactions if the interactions are between different phenotypes; these interactions are a second critical aspect of engineered tissue formation. One of the seminal examples of the use of microfabrication technology to investigate heterotypic interactions for tissue engineering applications was the work of Bhatia *et al.* (1998b) involving interactions between hepatocytes and fibroblasts. Micropatterns of collagen were applied to substrates in order to provide a template for hepatocyte adherence, followed by culturing with fibroblasts to fill in the remaining spaces on the 2D substrate with fibroblasts (Bhatia *et al.* 1999). These studies probed the influence of hepatocyte/fibroblast spacing and densities on liver-specific functions and ultimately investigated the degree to which these heterotypic interactions are mediated by direct contact versus communication through exchange of soluble chemokines. Other studies have utilized microfabrication techniques to probe mechanisms involved in the downregulation of cell proliferation (Nelson and Chen 2003). In a fashion analogous to contact guidance phenomena between cells and a substrate, contact inhibition governs cellular proliferation via signals generated by contact between neighboring cells. Microfabrication techniques have been used to separate cell density effects from contact inhibition effects, highlighting the significance of specific cell–cell contacts (cadherins) in vasculature.

Three-dimensional platforms for organizing and probing cell–cell interactions *in vitro* have been developed by Khademhosseini *et al.* (2005). In this approach, surface adhesion molecules are not necessary to form aggregates of multiphenotype arrays. Rather, cells are introduced into microwells via microfluidic channels that can be assembled and reassembled because of the reconfigurable nature of the polymer matrix. Subsequent channel assembly can be conducted along different axes to provide lanes of cells of each type introduced into the microwells, thereby forming a highly arrayed platform for high-throughput screening and other discovery applications.

Fluid mechanical effects, both through direct fluid mechanical forces and through dynamic generation and control of chemical gradients, are another avenue for investigating cellular interactions at the microscale. Laminar flow streams at very low Reynolds numbers can be used to control chemical concentrations locally, and a microfluidic gradient mixer has been developed to produce controlled gradients of surface-bound molecules

for cell and tissue engineering applications (Dertinger *et al.* 2002). Direct interactions modulated by fluid mechanical forces have been central to advancements in understanding endothelial cell behavior and its role in atherogenesis. Wall shear stress modulation has been shown to be a controlling factor in activating endothelial cells (Shen *et al.* 1992) and in the development of atherosclerotic plaques (Malek and Izumo 1996). Shear forces in other systems, such as liver bioreactors (Allen and Bhatia 2003), and in the renal proximal tubules of nephrons (Essig and Friedlander 2003) are believed to exert control over function and to be associated with disease processes.

Forces at the subcellular level may also be probed using microscale and nanoscale techniques (Raghavan and Chen 2004). Contact guidance phenomena represent one such approach for studying interaction between cytoskeletal structures and substrates. Other approaches involve measurements of surface buckling as a function of the subcellular forces exerted on an elastomeric surface. More recently, Tan *et al.* (2003) have pioneered the use of microstructures such as arrays of microneedles as a monitor for subcellular forces, using needle deflection as a sensing technique. The top of the microneedles is coated with surface adhesion molecules, whereas the floor of the microneedle array is coated with PEG or another chemical that inhibits cell adhesion. In this embodiment, surface chemistry can be independently controlled and varied relative to the mechanical parameters of the microneedles. Modulation of the mechanical sensing element can be achieved simply by varying the height and the diameter of the microneedles. **Figure 1** is an image of a microneedle array with the tips of the posts coated with fibronectin and smooth muscle cells (SMCs) attached to the microneedles (Tan *et al.* 2003). The full potential of microneedle arrays and other microfabrication-based tools for probing cell behavior is only beginning to be realized.

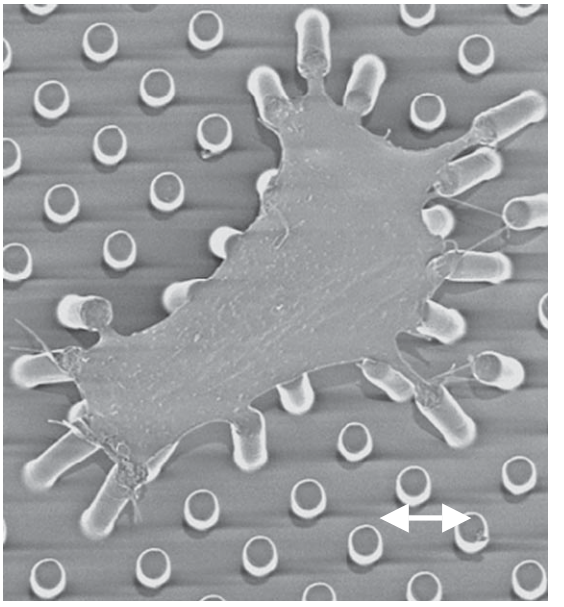


Figure 1 Smooth muscle cell attached to the tips of fibronectin-coated microneedles in poly(dimethylsiloxane) (PDMS). Bar, 10 μm. (Source: Tan J L, Tien J, Pirone D, Gray D S, Chen C S 2003 Cells lying on a bed of microneedles: An approach to isolate mechanical force. *Proc. Natl. Acad. Sci. U S A* **100**, 1484–9.)

2.15.2 Microstructures and Cellular Control

Engineered tissues rely on the control of a large number of experimental parameters in order to produce a construct that mimics physiological properties. These parameters include factors related to the cells that make up the engineered tissue, the biochemical environment and biochemical properties of the materials used to make up the substrate of the engineered tissue, and the structural properties of the scaffolding that comprises the backbone of the construct. The advantages of utilizing microfabrication and nanofabrication techniques to build scaffolds for tissue engineering relate principally to the ability of these fabrication methods to produce structures that resemble the mechanical microenvironment of living tissues and organs. The four components of the microenvironment that influence cell behavior in engineered tissues, spatial modes of control, topographic modes of control, the fluid mechanical environment, and the biochemical and surface chemical environment, will be covered in Sections 2.15.2.1–2.15.2.4.

2.15.2.1 Spatial Control of Cell Behavior

The spatial environment of cells in the various tissues of the body represents a critical element of control of cell behavior and the structure and function of tissues and organs. For instance, endothelial cells in the glomerulus of the kidney and the retina in the eye form fenestrations in the vessel walls that provide critical

functions. Liver sinusoids represent specific spatial microenvironments that are critical to the proper functioning of hepatocytes. In many instances, spatial control involves the formation of structures suitable for coculture of multiple cell types. The orientation and polarization of specific cells in the liver must be controlled to engender proper function; this is also true in the retina and the cardiovascular system.

Numerous microfabrication techniques that are capable of producing 3D spatial microenvironments using layering methods have been developed (Borenstein *et al.* 2002). As will be described later, many of these layering techniques involve the use of 3D microfluidic structures to produce microvascular networks, hepatocyte chambers, and other engineered cellular assemblies, as will be described in greater detail in a later section. Another approach, pioneered by Khademhosseini *et al.* (2005), utilizes the reconfigurable nature of poly(dimethylsiloxane) (PDMS) scaffolds to produce 2D assemblies of microwells for coculture of multiphenotype cell arrays.

Microfluidic and micropatterning techniques developed for spatial control of cell behavior in engineered tissues typically utilize surface chemistry to modulate cell function (Kane *et al.* 1999). These methods involve photolithography, either applied directly to the substrate or applied through an indirect class of methods known as soft lithography. Direct methods can involve photocrosslinkable materials (Nuttelman *et al.* 2002) or may use UV immobilization of proteins that modulate cell adhesion. These techniques are fairly limited in their applicability due to restrictions on the type of substrate and surface adhesion molecules that can be employed. Photocrosslinkable hydrogels and other materials are being developed to enable direct construction of scaffolds with patterned surfaces to enable the arrangement of multiple cell types and to localize growth factors within the matrix. Properties such as adhesion, migration, and differentiation have been influenced by spatially controlling either the arrangement of surface adhesion molecules or the degree of crosslinking within scaffolds.

There have been significant advances regarding the microfluidic methods for controlling the spatial environment of tissue engineering scaffolds. The simplest approaches utilize a high concentration source that forms a diffusion-controlled gradient. These approaches do not produce stable or controllable means for generating gradients necessary for precisely arranged cellular constructs. However, a new technique for producing precise gradients using microfluidic mixers has revolutionized the field, typically using chemotaxis as a means to attract particular cell types to specific locations within an engineered construct. Growth factors modulated within these gradient systems are being used to produce engineered neural and vascular tissues.

Micropatterning techniques employ soft lithography to modify the surfaces of both degradable and nondegradable polymer substrates. These materials include polylactic acid (PLA), poly(lactic co-glycolic acid (PLGA), and chitosan, with surface modification molecules such as RGD (arginine–glycine–aspartic acid) or PEG. In an early pioneering study, Chen *et al.* (1997) utilized soft lithography to produce a variety of patterns for cell adhesion to a substrate and determined that cell life and death can be controlled by the specific spatial patterns of surface adhesion molecules on a substrate. Seminal studies involving neurons and liver cocultures of hepatocytes and fibroblasts have been conducted using surface adhesion molecules patterned using soft lithographic stamps. In the case of liver coculture, micropatterns have provided a means to determine whether direct cell–cell contact or communication via circulating biomolecules is required for physiological cell function. Ultimately, engineered tissue formation will be enabled by 3D scaffolds combining microfluidics and micropatterning to organize cell cocultures.

2.15.2.2 Topographic Control of Cell Behavior

Signaling associated with the mechanical properties of substrates is now recognized as an extremely powerful modality for controlling cell behavior in tissues and organs. A revolutionary concept known as cellular tensegrity (Ingber 1993) models cells as mechanical structures in which the cytoskeleton and intracellular organelles are treated as the framework of the cell driven principally by mechanical interactions with the environment. A very large body of work associated with the investigation of topographic effects on cell behavior has emerged, with applications including the use of topography to prevent cell adhesion on implant surfaces, promote cell migration and differentiation in engineered tissues, and align cells for applications in which cell orientation is critically related to function such as in blood vessels and liver.

2.15.2.2.1 Contact guidance

The basement membrane, which makes up the underlying surface of most tissues and serves as the matrix on which cells reside, is a complex and rich mixture of components such as collagen, laminin, and fibronectin. The chemistry of these materials has a profound effect on many aspects of cell behavior, including adhesion strength, migration, and proliferation (Mooney and Hansen 1992). Numerous studies of special receptors known as integrin receptors have been conducted, leading to a greater understanding of the manner in which specific ligands present surfaces that enhance or inhibit various cell behavior. While these receptor-modulated interactions are largely chemical in nature, they are only a part of the spectrum of interactions that control cell behavior in physiological and engineered systems.

In addition to providing knowledge on surface chemical interactions, recent work has led to the discovery of an equally complex and significant family of interactions that are best characterized as types of mechanotransduction. Mechanical interactions between cells and the underlying substrate can take several forms, including topographic interactions as well as surface tensile forces. The advent of micromachining techniques for tissue engineering scaffolds has opened the door for topographic control of cell behavior at the microscale and at the nanoscale. These interactions are generally classified as contact guidance, in which the structural characteristics of a substrate influence the behavior of cells interacting with the surface. Cells that have been investigated include fibroblasts, epithelial cells, endothelial cells, astrocytes, and oligodendrocytes. Numerous techniques for producing micro- and nanoscale topography have been used, as will be described in the following sections.

First observed by Harrison in 1914, the influence of microscale topography on cell migration was termed as contact guidance by Weiss (1945), in a series of experiments utilizing available microtopographic substrates to characterize the influence of microscale surface architecture on cell motion. Microfilaments within the cytoskeletal structure are believed to play an important role in contact guidance, because substrate topography can limit the bundling or bending of microfilaments that normally occurs in cell motion. Contact points between cells and substrates are typically in the range of 1 μm in size, and therefore, microstructural features of that order are expected to have powerful effects on cell adhesion, motion, and function. In the 1980s, Brunette *et al.* (1983) used photolithography to pattern materials such as titanium to demonstrate the influence of microscale topography on the migration of fibroblasts and other cells. Brunette was also the first to use polymeric substrates for studies of topographic interactions with cells. Brunette later demonstrated that integration of dental tissue-engineered implants was improved when surface microtexture was applied to the scaffold surface. Others such as Den Braber demonstrated that the inflammatory response to an implant could be modulated by surface microtopography (Den Braber *et al.* 1998). Curtis and Clark (Clark *et al.* 1991) found that surface microtopography with specific properties could completely inhibit the adhesion of various cells to the surface, an effect that is totally independent of surface chemistry. Many of these studies have been conducted as a function of specific topographic features and have investigated the strength of the contact guidance interaction as a function of the lateral and vertical dimensions of the surface topography. In general, the strength of the contact guidance interaction for microtopographic surfaces increases as the depth of the feature increases and decreases as the width of the feature increases (Van Kooten *et al.* 1998). Interactions are enhanced, and cell behavior approaches that of the *in vivo* environment more closely, when the surface microtopography approximates cellular dimensions more closely. Bettinger *et al.* (2006a) have generated microscale patterns on the biodegradable elastomeric substrate polyglycerol sebacate (PGS) and have shown that bovine aortic endothelial cells (BAECs) cultured on these substrates respond strongly to the underlying topographic features. Cell alignment and elongation on micron-scale grids and gradients are dramatically enhanced relative to flat control substrates, as shown in **Figures 2** and **3** (Bettinger *et al.* 2006a). In **Figure 2**, the alignment of BAECs on gratings as a function of the grating parameters is graphed, and it is clear that the endothelial cells interact strongly with the surface texture in a way that aligns them very tightly around 0°; the pattern also elongates them severalfold when compared to an endothelial cell bound to a flat surface. In **Figure 3**, patterns of BAECs on a grid are shown.

The first studies of nanoscale topographic control over cell behavior were reported by Rosenberg (1962). More recently, numerous studies of the actual dimensions and features of basement membranes have been conducted. In addition to the typical

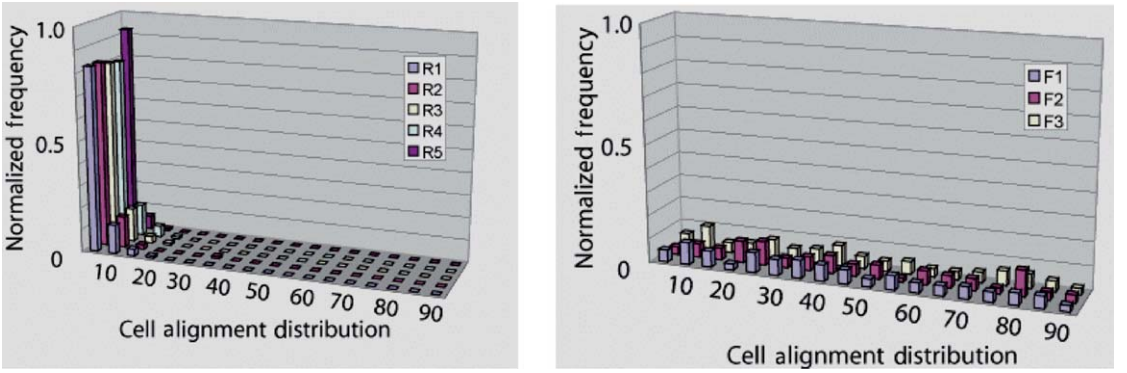


Figure 2 Alignment of bovine aortic endothelial cells (BAECs) for several experiments (R1–R5) for an ordered pattern of ridges and grooves (*left*) and for a flat substrate (*right*). (Source: Bettinger C J, Orrick B, Langer R, Borenstein J 2006a Tunable contact guidance using feature geometry in biodegradable substrates. *Biomaterials* **27**, 2558–65.)

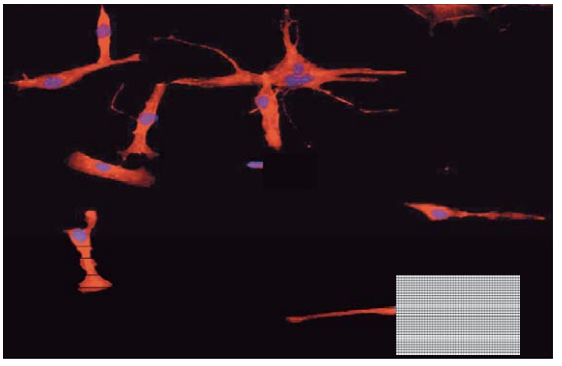


Figure 3 Image of bovine aortic endothelial cells (BAECs) cultured on polyglycerol sebacate (PGS) patterned with a grid of 0.9 μm width lines. (Source: Bettinger C J, Orrick B, Langer R, Borenstein J 2006a Tunable contact guidance using feature geometry in biodegradable substrates. *Biomaterials* **27**, 2558–65.)

appearance of pore-like structures, ridges and valleys in the basement membrane surface have been observed in experiments involving microscopic analysis of the membrane itself or a cast thereof. One of the most studied surfaces is the corneal epithelial basement membrane, which is characteristic in the macaque monkey and in humans (Abrams *et al.* 1997). In these investigations, characteristic dimensions in the range of 70–190 μm were identified for the height and width of specific features. The basement membrane of the glomerulus in the rat and the bovine kidney has also been studied by several groups, leading to the identification of features in the range of 10 nm. These and other findings have led to a significant concern related to the large body of literature associated with microtopographic control over cell behavior. Namely, this concern relates to the fact that most of the studies conducted on microfabricated substrates and contact guidance of cells have been done at dimensions of 1 μm or larger, while it is believed that the actual topography of physiological basement membranes occurs predominantly at the scale of 100 nm or less. Therefore, recent investigations have begun to focus on the sub-500-nm range for investigations of contact guidance.

Numerous methods have been used to produce features in the range of 500 nm or less. Clearly, most etching techniques are easily capable of producing vertical groove depths much less than 500 nm, most utilizing standard photolithography and reactive ion etching. Clark *et al.* (1991) used X-ray lithography to produce grooves with widths near 100 nm, while other researchers utilized techniques that are less controlled. For instance, Goodman utilized casting of polymethylmethacrylate (PMMA) polymer films into an actual basement membrane to create a master mold for subsequent polyurethane (PU) casting of substrates with dimensions near 100 nm (Goodman *et al.* 1996). Sandblasting and sanding have been used to produce features as small as 100 nm or less. And as mentioned earlier, Curtis has patented a technique for preventing cell adhesion by using a specific nanotopographic pattern.

2.15.2.3 Fluid Mechanical Control of Cell Behavior

Beginning with the pioneering work of Gimbrone *et al.* (1974) and Jaffe *et al.* (1973), the influence of fluid mechanical forces on endothelial cells

(Garcia-Cardena *et al.* 2001) has been the subject of intense investigation. Flow-induced alignment of endothelial cells at high flow rates is a well-studied effect, but more complex phenomena that are associated with shear stress have been the subject of intense investigation since the early 1990s. The influence of shear stress on endothelial cell function has been associated with the geometrically focal nature of atherosclerosis, in which atherosclerotic plaques form preferentially at the outer walls of blood vessel bifurcations. It has been shown that these regions along the outer walls of bifurcations more often experience low, time-varying levels of shear stress, whereas most arterial blood vessel walls experience shear in the range of 10–15 dynes cm^{-2}, a much more physiologically desirable level of shear. Endothelial cells experiencing physiological levels of shear are quiescent and exhibit a gene expression profile that is protection against the formation of atherosclerotic plaques. In regions of low oscillatory shear, endothelial cells become activated, exhibiting gene expression profiles that are atherogenic in nature. At very high levels of shear, in the range of 60–70 dynes cm^{-2}, endothelial cells can be physically damaged and may be mechanically sheared from vessel walls.

Most of the seminal studies conducted in the field of fluid mechanics and mechanotransduction on endothelial cell function have utilized some type of cone–plate viscometer arrangement that produces controlled levels of shear as a function of distance from the vertex of the cone. Cells are cultured on a flat plate positioned adjacent to a small-angle cone that provides controlled levels of shear stress. Microfluidic approaches to the analysis of shear are emerging, however (Weinberg *et al.* 2004). Recent computational fluid dynamic models of vascular networks for tissue engineering have accounted for wall shear stress effects in microfluidic channels, enabling the study of endothelial cell phenomena in microchannels with physiological features such as vessel distensibility (Krenz and Dawson 2002). These analyses treat shear stress as a control variable in the microfluidic system, enabling studies of endothelial cell function and development of functional blood vessels in microchannel networks. **Figure 4** illustrates a section of a PLGA-micromolded microvascular network generated by this computational model that results in a controlled, constant level of wall shear stress throughout the pattern of bifurcated channels (Weinberg *et al.* 2004).

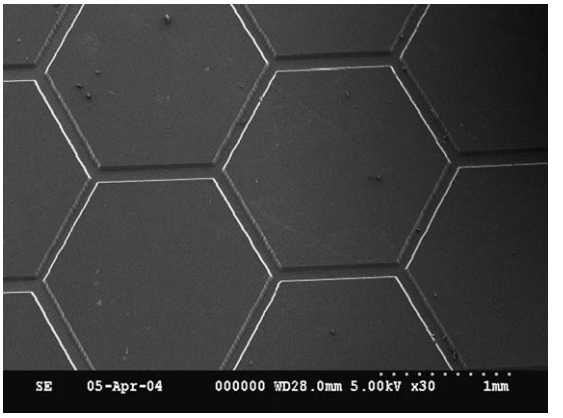


Figure 4 Section of a micromolded poly(lactic co-glycolic acid) (PLGA) scaffold machined with a microvascular network pattern generated to control the wall shear stress to a constant level through all channels. (Source: Weinberg E J, Borenstein J T, Kaazempur-Mofrad M R, Orrick B, Vacanti J P 2004 Design and fabrication of a constant shear microfluidic network for tissue engineering. *MRS Symp. Proc.*, Vol. 820, MRS Press, Warrendale, PA, pp. 121–6.)

2.15.2.4 Chemical Control of Cell Behavior

Microfabrication and microsystems technology are capable of generating controlled microenvironments and environmental cues based on engineered surface chemistry and circulating chemical factors. Chemical control can be accomplished using one of the following two principal approaches:

(1) Modification of an existing surface using physical or chemical means
(2) Deposition of a material with specific properties onto the substrate

One of the most common approaches for chemical control has been the application of micropatterned regions with adhesive and nonadhesive regions. Micropatterning typically utilizes photolithographic techniques, either by directly patterning the surface or by generating a stamp for microcontact printing of the surface adhesion molecules. In addition to the application of specific cell adhesion ligands, other types of surface modification can produce specific effects such as anticoagulation or prevention of biofouling with proteins. These modification techniques are critical for the development of devices such as liver assist devices and renal dialysis cartridges. PEG can be used to prevent protein adhesion and can be immobilized on surfaces using a number of microfabrication-based techniques. Heparin binding can be

accomplished in order to promote hemocompatibility and to reduce the incidence of clotting.

2.15.2.4.1 Circulating chemokines

Microtechnology can be used to control the spatial distribution of circulating chemokines, by leveraging specific fluid mechanical properties of microchannels. Typically, flow in microchannels is laminar, and therefore, streamlines can be maintained at channel bifurcations due to the absence of convection-driven mixing. Diffusive transport is typically very limited, and therefore, cells can be positioned at specific regions within merged streams in order to generate a particular microenvironment. Multiple flow streams can be processed in chambers that merge or split the flow, as well as in precision flow mixing chambers, thereby creating a gradient of soluble factors. A microfluidic gradient mixer, to be described in Section 2.15.5.1.2, generates a gradient of soluble molecules such as growth factors, but these soluble factors are typically bound to the substrate before the introduction of cells for culture and the development of engineered tissue constructs.

2.15.2.4.2 Surface-bound molecules

Tissue-engineered systems and artificial implantable devices must be designed so that contact with blood and other tissues does not result in adverse inflammatory and immune responses. The surface of an implant presents itself to the body only temporarily; proteins are rapidly adsorbed onto the artificial implant surface, and therefore, cells and body fluids react with the adsorbed layers (Wilson *et al.* 2005, Yang *et al.* 2001) rather than the bare implant surface. Integration of implanted devices depends critically on the cell–substrate with the adsorbed surface proteins, and therefore, the nature of this adsorption highlights the need to develop methods to present a controlled surface to the cells and blood. Materials such as fibronectin and vitrogen have been shown to promote the adhesion and migration of cells such as osteoblasts; as will be described in Section 2.15.3, tissue engineering substrates must be selected partly on their ability to adsorb proteins. Other applications, such as renal dialysis devices, utilize materials such as polyethersulfone (PES) that are selected because of their ability to impede protein adsorption, so that procedures such as hemofiltration can proceed without hindrance by the rapid buildup of proteins on the filtration membrane. As mentioned in Section 2.15.5.1.2, haptotactic signals, in which substrate adhesion mediates the motion of cells along surfaces, is another important mechanism by which control of the substrate presented to the cells is used to control cell behavior in tissue engineering applications.

In addition to adhesion and differentiation, the proliferation and differentiation of cells is strongly influenced by surface chemistry. Mechanotransduction via surface adhesion molecules influences cell apoptosis, as seen in the pioneering work of Chen *et al.* (1997), in which cell life or death is controlled by the spacing and geometry of microcontact-printed surface molecules. Trade-offs between surface adhesion and differentiation have been observed in many studies, and therefore, the entire spectrum of cellular activities must be considered when engineering the surface chemistry of tissue-engineered devices and implants.

Surface energy and surface charges, often associated with the level of hydration and wettability of surfaces, are other important factors to consider in engineering substrates for engineered tissues. Many biomaterials such as titanium are selected because of their hydrophilicity, an important factor in the success of hip implants. However, numerous biodegradable polymers are typically hydrophobic, and therefore, surface modification must take place in order to prepare them for seeding with cells. This is also true with nondegradable materials such as PDMS, where oxygen plasma treatment generates a transient hydrophilic nature that enables layer-to-layer bonding. Cell adhesion on positively charged surfaces is strong and distributed, whereas neutral or negatively charged substrates allow cells to adhere at specific points of contact. In addition to having direct effects on cells, surface charges influence the subsequent adsorption of proteins such as fibronectin and vitrogen.

2.15.3 Biomaterials and Scaffolding for Tissue-Engineered Microsystems

Most applications of biomaterials in tissue engineering follow the paradigm set forth by the seminal work of Vacanti and Langer (1999) in which a biomaterial scaffold is seeded with cells, the resulting device is cultured *in vitro*, and is then implanted to augment or replace a failing tissue or organ. These biomaterials span a range from completely synthetic materials such as silicon, metals, glasses, and ceramics and synthetic polymers such as polyesters and silicone materials to biologically based materials such as collagen, chitosan, and hyaluronic acid. Bioactive scaffolds that impart biological function such as

growth factors, drug delivery to minimize infection and inflammation, and other smart materials are emerging as part of the portfolio of candidate biomaterials for tissue engineering. The role of the biomaterial is to provide signals and cues that may be mechanical, chemical, or biological in nature and to provide a structure for population with cells that will control mass transport, cell adhesion and cell motion, and the deposition of extracellular matrix (ECM). Biodegradable materials that provide temporary chemical and structural functions are a critical element of this portfolio, from the standpoint of the need for an initial structural platform on which the new tissue will begin the process of formation and will ultimately replace.

2.15.3.1 Inorganic Biomaterials

In the field of biomaterials for tissue engineering and implantable devices, inorganic materials such as silicon, metals, glasses, and ceramics played a very important role until fairly recently. Metals such as titanium are still the standard approach for implants such as a hip prosthesis that requires high strength and minimal corrosion over decades. However, these inorganic materials possess mechanical, chemical, and biological properties, very different from native tissue, and therefore, their applicability in cell and tissue engineering is very limited (Yang *et al.* 2001). The mechanical robustness of these high-strength, rigid materials is paramount in certain implantable applications, but interactions between inorganic biomaterials and cells, both in the body and in engineered constructs, are not optimal, and therefore, a shift toward biomimetic materials such as engineered biopolymers has occurred for most tissue engineering systems.

2.15.3.1.1 Silicon

Silicon substrates are an ideal platform for investigations of cell behavior on microfabricated surfaces and in microdevices (Kaihara *et al.* 2000), because of the broad range of silicon processing techniques developed for microelectronics and MEMS applications. However, engineered tissues for clinical applications rarely consider the use of silicon as a substrate material because of its rigidity, brittleness, and relatively poor biocompatibility. Therefore, micromachined silicon structures have been utilized as a discovery tool rather than as a clinical end point in most tissue engineering research. Many cell culture studies have utilized silicon substrates, either simply as a surface for microcontact printing of surface adhesion molecules or as a substrate for etching microchannels or other features. One novel use of silicon as a scaffolding material for tissue engineering utilizes porous silicon as a mold for polymeric scaffolds (Chin *et al.* 2001). While crystalline silicon is mostly used as a substrate for *in vitro* experimentation, porous silicon shows evidence of biodegradation and other biocompatibility-related features that have drawn increasing attention for tissue engineering applications.

2.15.3.1.2 Metals

The biocompatibility of metals such as titanium and tantalum is well known, and numerous medical devices have been based on these excellent properties. Other metals used in implants include stainless steel and cobalt-based alloys. The hip prosthesis represents an excellent example of a long-life (20 years) implant that has provided long-term mechanical strength while minimizing undesirable responses such as inflammation, foreign body response, and the formation of bacterial biofilms. However, metals are not biodegradable and have a very limited flexibility in terms of processing.

While metals are not typically utilized for microfabricated scaffolds in tissue engineering, they have played a critical role in the establishment of fundamental mechanisms in cell behavior leading to the development of engineered tissues. Early work by Brunette and coworkers led to the observation that epithelial cell migration can be dramatically influenced by microscale topography on titanium-coated surfaces. These studies are important because of the role of epithelial cell migration in the loss of teeth and dental implants, and in the development of tissue-engineered approaches for dental (Young *et al.* 2002), craniofacial (Warren *et al.* 2003), and other applications. Initial studies utilized grooved silicon substrates with groove dimensions and periodicities in the range of 10–150 μm, covered with a thin layer of titanium (Brunette *et al.* 1983). Enhanced motion of epithelial cells associated with particular groove geometries suggested pathways for the prevention of cell migration using specified surface architectures.

2.15.3.1.3 Glasses and ceramics

As in the case of silicon, most of the tissue engineering research conducted with glasses has aimed at the establishment of fundamental mechanisms in cell culture and coculture. Glasses are an ideal substrate for microfabricated cell culture studies because of their transparency, ease of fabrication,

controllable surface properties, and relative biocompatibility. Micropatterning of surface adhesion molecules on glass substrates has led to seminal discoveries regarding hepatocyte culture, in the work of Bhatia *et al.* (1998a). In one study, borosilicate wafers, a low-cost amorphous glass substrate typically used for anodic bonding processes in MEMS devices, were used as the platform for micropatterned coculture of rat hepatocytes and 3T3 fibroblasts. Standard photolithography, which works well with borosilicate wafers as long as surface humidity is controlled, was used to pattern the surfaces. Surface modification via silane coupling was used to pattern the glass surface with surface adhesion moieties. Thiol-based coupling can also be used for surface modification, typically in concert with PEG in the non-adhesive-patterned regions in order to prevent cell adhesion in these areas. Cocultures of hepatocytes with other cell types such as biliary epithelial cells, stellate cells, and other cells have been investigated, but in this investigation, primary rat hepatocytes have been cocultured with 3T3 fibroblasts in order to investigate the critical elements of heterotypic interactions between different cell types in a microfabricated system.

Microfabricated ceramic scaffolds for cell culture have been applied to bone tissue engineering, using solid freeform fabrication (SFF) techniques that are described in Section 2.15.4.1. In one study, Chu *et al.* (2002) constructed orthopedic implant devices using a ceramic suspension that can be cured with UV light. These devices typically do not achieve a high resolution because of limitations in the fabrication technique. Strenuous efforts have been made to improve the resolution of stereolithography (SLA) and other SFF techniques, but these improvements are not typically translatable across all biomaterials, and therefore, the field of tissue engineering has not benefited as much as was anticipated from these technologies. Other materials such as hydroxyapatite are often used in bone tissue engineering in combination with SLA and other SFF techniques.

2.15.3.2 Biocompatible Polymers

Polymeric materials have emerged as the dominant scaffolding material for most tissue engineering applications, because of their superior biological and mechanical properties, low cost, and flexible processing capabilities. Biopolymers for tissue engineering can be natural, such as collagen, chitosan, glycosaminoglycans, and starch, and these materials have been successful in applications to skin, bone, and cartilage. Interactions between these materials and cells are far more optimal than for most of the inorganic scaffolding materials mentioned earlier, and bulk transport of nutrients and growth factors across these materials is also quite rapid. However, most of these natural materials are difficult to process and to control, and their mechanical properties are not optimal. Therefore, synthetic biodegradable polymers have emerged as the basis for most applications in tissue engineering.

2.15.3.2.1 Silicone elastomers

PDMS has served as the predominant structural material for the development of microfluidic devices and soft lithographic applications due to a combination of desirable properties. These include its high mechanical strength and flexibility, transparency, compatibility with sterilization, low cost, and ease of fabrication. The biocompatibility of PDMS, a silicone rubber material typically produced from a two-part mixture of curing agent and base, is still the subject of intensive investigations, although silicone materials are used in numerous clinical applications including catheters and membrane oxygenators. High oxygen diffusivity through the bulk of PDMS is an important factor when considering bioreactor applications, particularly with highly metabolic cells such as hepatocytes. Its ultimate utility in tissue engineering applications will likely be limited to laboratory devices, extracorporeal systems, and short-term implantable devices because it is not biodegradable and is ultimately prone to thrombosis due to nonoptimal hemocompatibility.

2.15.3.2.2 PMMA, PU, and other polymers

Microstructures built using a variety of polymeric materials have been investigated for cell culture and tissue engineering applications. One comprehensive study of various nondegradable polymeric materials microfabricated for culturing SMCs (Li *et al.* 2000) investigated PMMA, SU-8 epoxy resin, and PDMS. The SU-8 microstructures were fabricated using standard photolithographic techniques as described in Section 2.15.4.2.1, while the PDMS structures were formed directly from SU-8-negative master molds. For PMMA, the microstructures were produced using X-ray lithography to create high aspect ratio structures. For each substrate, a range of microstructural geometries were evaluated, including columnar structures, holes, and curved surfaces. In general,

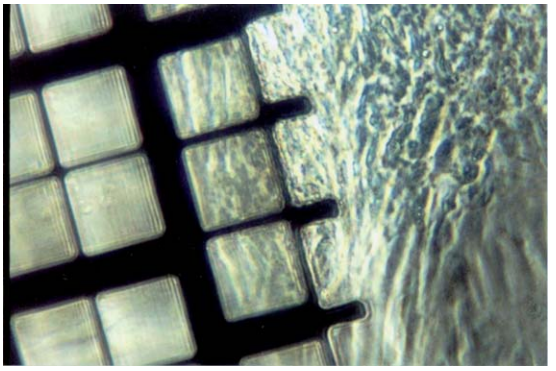


Figure 5 Smooth muscle cells cultured on polymethylmethacrylate (PMMA) columns, showing a significant barrier to cell spreading from the ~15-μm gaps. (Source: Li M, Glawe J D, Green H, Mills D K, McShane M J, Gale B K 2000 Effect of high-aspect-ratio microstructures on cell growth and attachment. *Proc.1st Annu. Int. IEEE-EMBS Spec. Top. Conf. Microtechnologies in Medicine and Biology*, Lyon, France, Vol. 66, Poster, 531–6.)

SMCs preferred curved surfaces, with much higher cell densities in culture than were observed in microstructures with sharp angles and rectangular geometries. This is commonly observed in cell culture on microfabricated substrates, likely because of the difficulty in establishing continuous layers of cells across sharp boundaries. **Figure 5** illustrates SMCs cultured on micromachined PMMA columns; note that small gaps between the columns limit the spreading of the SMCs (Li *et al.* 2000).

PU is a well-studied material with a long track record of use in biomedical devices. However, the mechanical and chemical properties of PU have limited its utility in certain tissue engineering applications, such as myocardial tissue. In one such application, McDevitt *et al.* (2003) have reported on micropatterning of a new class of biodegradable elastomeric PU in order to culture cardiomyocyte sheets. These novel materials have tunable degradation rates through the introduction of polycaprolactone (PCL) and a chain extender. Lines of laminin in the 10-μm range of width and spacing were microcontact-printed onto the biodegradable PU substrates, and the resulting robustness and stability of the cardiomyocyte lanes were enhanced.

2.15.3.3 Biodegradable Polymers

Biodegradable polymers represent the basis for most of the advances in tissue engineering applications for which cells are seeded onto a scaffold and implanted permanently. Resorption of the scaffold results in a fully natural replacement tissue without excess materials remaining in the body, clearly the ideal scenario for most applications. Most of these bioresorbable materials degrade via hydrolytic reactions rather than enzyme-driven processes, and therefore, the degradation process is highly controlled and relatively resistant to environmental variations between patients and between implantation sites.

The excellent performance of surgical sutures based on copolymers of polyglycolic acid (PGA) and PLA in the past has led to the use of these materials for numerous tissue engineering applications (Agrawal *et al.* 1997). These copolymers have highly tunable mechanical and degradation properties and therefore would seem to be an ideal substrate. However, the emergence of mechanotransduction as a significant factor in cell–substrate interactions (Ingber 2004, McKnight and Frangos 2003), and the fact that most basement membranes in the body are elastomeric in nature, has resulted in the pursuit of biodegradable elastomers for numerous tissue engineering applications. In addition, new classes of biopolymers have been developed in an effort to reduce the inflammatory response resulting from the degradation products of conventional PGA/PLA copolymers.

2.15.3.3.1 Biodegradable nonelastomeric materials

A number of biodegradable materials have been microfabricated for tissue engineering applications, including PLGA copolymers and PCL. These materials are not elastomeric and therefore exhibit a fairly high degree of rigidity that influences the ability to form constructs mimicking physiological tissues. Properties such as molecular weight, crystallinity, and the monomer ratios in biopolymers such as PGA/PLA copolymers can be utilized to tune the mechanical strength and degradation rate. Typically, high-molecular-weight PGA and PLA are generated using ring-opening polymerization (ROP) of the glycolide and lactide components (Yang *et al.* 2001). Each of these constituents provides specific properties to the resulting copolymer. The glycolides are highly crystalline, with glass transition temperatures near 50°C and melting temperatures near 200°C. PGA is highly hydrophilic, and its resorption time is typically in the range of 3 months. The mechanical strength of PGA drops sharply during the early stages of the degradation process, a major problem for many applications that require better midterm

stability until the engineered tissue itself becomes sufficiently robust.

PLA is not as highly crystalline as PGA and has a similar glass transition temperature but a slightly lower melting point (Yang *et al.* 2001). There are three important isomers of PLA, resulting from the chirality associated with a methyl group on the α-carbon unit. These three isomers are referred to as L-lactide, D-lactide, and D,L-lactide. The degradation rate and the mechanical properties of these three stereoisomers vary substantially, and limits on the thermal stability of the various monomers place restrictions on the processing conditions that can be used with each material. Plasticizing action within the monomers can alter mechanical properties in unpredictable ways, and therefore, careful attention must be paid during processing of the PLA class of materials.

Combinations of PLA and PGA to form the copolymer PLGA are frequently utilized in tissue engineering applications. The well-known minimum degradation time for this system occurs at specific ratios of the two monomers. Several groups have reported successful microfabrication of PLGA for tissue engineering applications, including the work of King *et al.* (2002, 2004) in which 3D constructs were compression-micromolded and the work of Vozzi *et al.* (2003) in which PLGA was deposited by a microsyringe technique to produce multilayer cell-seeded constructs. In a more recent development, Ryu *et al.* (2006) have fabricated devices from PLGA and polydioxanone (PDO) into 3D scaffolds suitable for engineered tissues; in **Figure 6**, a 3D PDO scaffold with feature sizes in the range of 20 μm is shown.

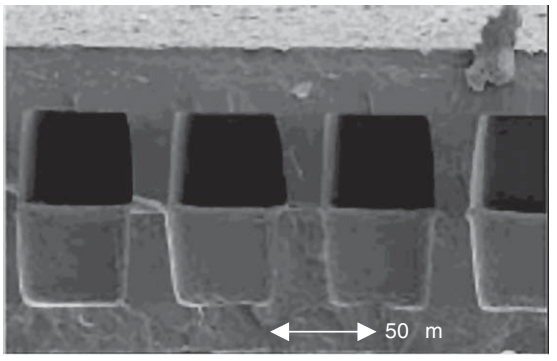


Figure 6 Scanning electron microscope (SEM) cross section of 20-μm structures micromolded into a 3D scaffold constructed from polydioxanone (PDO). (Source: Ryu W, Min S, Fasching R J, Prinz F B 2006 3D multi-layered microfabricated tissue scaffolds of biodegradable polymers. *Soc. Biomater. Conf.*, Pittsburgh, PA, USA.)

Another significant biodegradable nonelastomeric material is PCL, also produced by ROP to produce a polymer with a very low melting point near 60°C, roughly the same as the glass transition temperature. The degradability of PCL is because of an ester linkage that is subject to hydrolysis. Biocompatibility is considered to be very high, but the degradation time is so long (2 years) that numerous techniques have been pursued to speed resorption. Copolymers with PLA have been used to produce more rapidly degrading substrates. Like PGA/PLA, PCL is a relatively stiff material and therefore is not suitable for tissue engineering applications such as for blood vessels (Ziegler and Nerem 1994), heart muscle, and other soft tissues. Another limitation relates to the difficulty in functionalizing materials such as PLGA, although recent work by Keegan *et al.* (2003, 2004) has resulted in the demonstration of biomimetic scaffolding for tissue engineering and drug delivery.

2.15.3.3.2 Biodegradable elastomers

As described in the previous section, biomaterials such as PCL and polylactic co-glycolic acid (PLGA) have served as cornerstones in the field of biodegradable scaffolding for tissue engineering, as demonstrated by the numerous instances in which these materials have been implanted in animals and in humans for tissue engineering applications. In the case of PLGA, degradation times can be tuned by varying the lactic/glycolic ratio, as well as other variables in the copolymer preparation process. For PCL, degradation times are typically much longer, in the range of 1–2 years, but can also be modified to a lesser degree. One of the principal drawbacks of these and other thermoplastic materials is the fact that mechanically they do not resemble ECM, the proteinaceous substructure underlying cells and forming the mechanical scaffolding for tissues. In most cases, ECM is very strong yet highly elastic, resembling vulcanized rubber much more closely than the rigid polystyrene-like mechanics of PLGA and PCL.

Several efforts to develop tissue engineering scaffolding materials that resemble natural ECM more closely have been reported. Hydrogels with soft, rubbery mechanics have been developed, based on alginates as well as PEG chemistries. Pioneering work by David Tirrell (Petka *et al.* 1998) has led to the development of a periodic table of artificial proteins with elastomeric properties. A class of materials

based on polyhydroxyalkanoates has been applied to tissue engineering applications by Sodian *et al.* (2002). In each of these cases, the combination of distensibility and mechanical toughness does not approach that of natural ECM, and therefore, the applicability of these materials for scaffolding is limited. A new biodegradable elastomer known as PGS, or biorubber, has been reported by Wang *et al.* (2002). This new material is less expensive, tougher, and more distensible than the aforementioned biodegradable elastomers. In addition, PGS appears to be more biocompatible in implantable applications, producing less inflammatory response in early *in vivo* studies with small animals. The degradation of PGS occurs through hydrolysis through an ester chemical bond, with a low density of crosslinking to preserve the elastomeric properties. The monomers that comprise PGS are nontoxic and have been approved by the FDA for numerous applications.

2.15.3.3.3 Hydrogels

Elastomeric hydrogels referenced in the previous section are based on alginates as well as on PEG chemistries. Recent progress in calcium alginate hydrogel processing has led to a microfluidic biomaterial (Cabodi *et al.* 2005a). This material has been utilized in tissue engineering applications because of its low cytotoxicity and relatively high mechanical stability. Another powerful advantage of this material when compared to many of the previously mentioned biopolymers is the high degree of diffusive mass transport through the bulk of the scaffold. Microsystems-based scaffolds such as PDMS, PMMA, and PGS do not exhibit significant levels of mass transport of nutrients through the bulk of the material, other than the high oxygen diffusivity through PDMS. Therefore, transport of small molecule nutrients must take place within microchannels or through semipermeable membranes within the constructs. In the case of a calcium alginate hydrogel, by contrast, small molecule transport occurs readily within the bulk of the scaffold, as demonstrated by experiments using fluorescein- and rhodamine-labeled dextran as simulants for nutrients. The scaffold networks are assembled using soft lithographic preparation of master molds and casting of calcium alginate slabs with microchannel networks. Adjacent slab layers of the calcium alginate can be readily bonded. **Figure 7** illustrates a microfluidic network, constructed from micromolded calcium alginate, which has been infused with rhodamine-labeled fluid (Cabodi *et al.* 2005a).

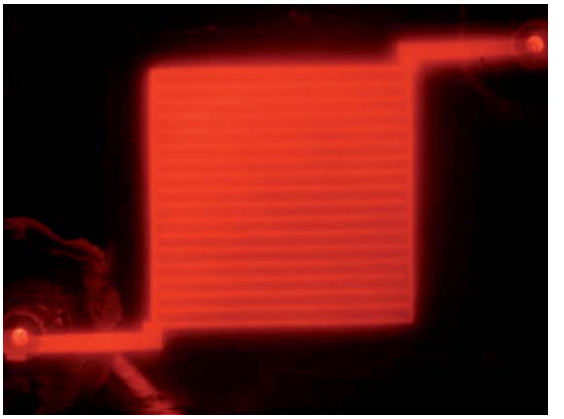


Figure 7 Rhodamine B isothiocyanate (RITC)-labeled dextran infused in micromolded channel network in calcium alginate gel. (Source: Cabodi M, Choi N W, Gleghorn J P, Lee C S D, Bonassar L J, Stroock A D 2005a A microfluidic biomaterial. *J. Am. Chem. Soc.* **127**, 13788–9.)

2.15.3.4 Biological Materials

Synthetic materials such as polystyrene, PDMS, and biodegradable materials such as PLGA, PGS, and PCL form the basis of most microscale technologies in tissue engineering (Armani and Liu 2000). The ability to tune these materials, including mechanical, chemical, and degradation properties, as well as surface modification techniques such as protein adsorption or chemical modification, represents powerful means to achieve specific behaviors that promote the development of healthy tissue with physiological properties. However, biological materials such as collagen, laminin, and fibronectin benefit from the fact that their interactions with cells are known to result in structure and function that enables long-term stable tissues and organs. These materials are often used to coat synthetic microfabricated scaffolds in order to render the scaffold surfaces more biocompatible. For instance, culturing of endothelial cells on uncoated PDMS substrates results in poor adhesion and scant coverage (Borenstein *et al.* 2002); cell culture on microfabricated PDMS scaffolds coated with fibronectin, for instance, results in a much more extensive coverage with endothelium. As will be described in a later section, micropatterning of surfaces with laminin produces axon formation and polarization. Therefore, these materials represent powerful tools for eliciting specific cellular responses in tissue engineering and may be used either as patterned surface coatings or as substrates themselves for engineered tissues.

2.15.3.4.1 Collagen

Collagen is the most prevalent protein in the connective tissues and is found in more than two dozen types, including type I collagen, found in bone and tendons, and type II collagen, found in articular cartilage. Mechanically, collagen is strong and nonelastomeric and forms the mechanical support structure for skin, cartilage, tendons, ligaments, and bone. Collagen is formed from subunits in the range of a few nanometers in length and is in the range of a few nanometers; these strands are twisted into a helical structure that provides the functionality of the material. Collagen has been used in skin substitutes for burn patients and is commonly used in cosmetic surgery. As will be described in a later section, many investigators utilize collagen as a surface adhesion molecule for micropatterning studies for tissue engineering applications.

Numerous methods for microfabricating collagen scaffolds for tissue engineering applications have been developed. In one approach utilizing excitation by light, the Campagnola group (Basu *et al.* 2005) has reported a technique known as multiphoton excitation to produce 3D ECMs with a 3D structure. This approach utilizes multiphoton confinement within a narrow focal plane for nanoscale architectural control and is capable of reproducing the topographic as well as the biochemical cues that are critical in defining the properties of skin and other tissues. In another study, Downing *et al.* (2005) utilized micromolding of collagen to produce an analog of the basal lamina. Silicon master molds were fabricated with lines, ridges, and other features in the 50–200 μm range of depth and width, as inspired by observations regarding the actual geometry of the basal lamina of skin. Keratinocytes cultured on microtextured collagen substrates showed that stratification and differentiation of the resulting epidermal layer increased as the interaction between keratinocytes and the microtexturing was increased, by increasing the depth of the features and by reducing their width. These studies suggest that microfabrication of biological materials such as collagen and laminin represents an important component of scaffold development for engineered tissues.

2.15.3.4.2 Silk

The superior mechanical and biological properties of silk fibroin make it an excellent candidate for scaffolding for tissue engineering applications. Silk has been FDA-approved for numerous applications and is extraordinarily strong and flexible, with highly tunable properties. Spider silk can be processed to form porous matrices, films, fibers, and hydrogels, and can be functionalized with surface moieties such as bone morphogenetic protein (BMP) for bone tissue engineering applications. There have been recent investigations on the combination of spider silk with biosilica (Foo *et al.* 2006), a material found in the structure of single-celled organisms such as diatoms, insect exoskeletons, and many higher-order animals. New silk–silica nanocomposites present an opportunity for both *in vitro* tissue engineering and *in vivo* regeneration, and these materials may also be compatible with microfabrication and nanofabrication techniques to produce engineered microenvironments for the replication of tissue structure and function.

2.15.4 Microfabrication Processes for Engineered Tissues

Microfabrication processes for tissue engineering include lithography-based processes such as micromachining and micromolding, as well as SFF methods in which scaffolds are formed directly by curing or sintering (Hutmacher 2001, Hutmacher *et al.* 2004, Leong *et al.* 2003, Tsang and Bhatia 2004, Yang *et al.* 2002). The principal advantage of SFF methods is that they are inherently 3D in nature, while the advantage of conventional lithography-based methods is superior spatial resolution. Each set of methods has been successfully applied to biodegradable polymer scaffold fabrication in various tissue engineering applications.

2.15.4.1 Solid Freeform Fabrication

SFF techniques include SLA, selective laser sintering (SLS), and other methods in which a laser is used to cure precise regions of a scaffold through photopolymerization or sintering. These processes use computer-aided design (CAD) to generate a 3D recipe for the scaffold, which is organized into 2D slices typically a few hundred microns in thickness. Scaffolds are produced by mechanically controlling the laser within the horizontal plane, forming the pattern by polymerizing the monomer in specified locations according to CAD. The vertical and lateral resolution of these techniques depends on numerous factors; lateral resolution is typically in the range of 50–250 μm. The laser spot size is a principal driver of resolution; the other issue is the manufacturing time

for scaffolds, which increases dramatically as the spatial resolution is improved. Vertical resolution is determined by the heat-affected zone depth in the photopolymerization process and by the stepper motor controlling the step size in the vertical dimension.

2.15.4.1.1 Stereolithography

SLA and the advanced version known as microstereolithography are mainstays of numerous industries and have been applied to several tissue engineering applications as well. Recent advances by Cooke *et al.* (2003) have enabled the scale-up of SLA toward sizes suitable for bone tissue engineering applications, overcoming a previous barrier to the achievement of structures of sufficient size for engineered tissues. These workers reported on bone defect repair with a matrix of poly(propylene fumarate) and diethyl fumarate. Biodegradable materials have been utilized in SLA scaffolds.

The close cousin of SLA, SLS, utilizes a laser to bond microparticles in a powder bed (Yang *et al.* 2002). The porosity of SLS scaffolds is typically much higher than that of SLA scaffolds. Scaffolding materials used in SLS include natural materials such as calcium phosphate, as well as biodegradable materials such as PCL and PLGA. Each of these techniques is a direct write method that requires substantial processing time in order to produce a scaffold of sufficient scale and resolution for complex tissue engineering applications. However, as will be described in Section 2.15.4.1.2, these techniques possess fundamentally 3D aspects that lend them to engineering large and complex tissues and organs.

2.15.4.1.2 Three-dimensional printing

Three-dimensional printing (3DP) (Kim *et al.* 1998) utilizes a different technique to solidify a powder bed to form a scaffold for tissue engineering. In a manner similar to that used in SLA and SLS, 3DP utilizes a scanning system that directs a writer toward specific positions in a 2D plane and then steps down in the vertical dimension to write the next slice. However, 3DP utilizes a print head similar to an ink-jet printer cartridge to apply a jet of chemical binder toward a powder bed. Subsequently, fresh layers of powder are deposited, and adjacent layers are joined by the binder. Once the part is completed, powder that is not sintered by the binder is removed, and the scaffold is removed.

One of the principal advantages of 3DP is that it can be conducted under ambient conditions. Unbound powder may be readily removed, but powder that resides within internal recesses of the scaffold is more difficult to extract. Therefore, the porosity of 3DP scaffolds is a challenge. Resolution challenges relate to the accuracy of the servo controlling x–y positioning of the ink-jet printer head, as well as the nozzle size. Another challenge with 3DP relates to the requirement for bulk scaffolding materials to be provided in a specific powder format. Sintering of the powder via binder deposition represents another challenge, in that each scaffolding material possesses its own limitations regarding agglomeration of powder. Surface properties such as roughness and microporosity are additional aspects of 3DP technology that require attention.

Numerous biodegradable scaffolding materials have been fabricated using 3DP, including PLGA, PCL, and polyethylene oxide (PEO). There exists a nearly unlimited array of scaffolding materials that can be fabricated using 3DP. Feature resolution remains a major challenge for 3DP; features of 200–1000 μm represent the minimum achievable geometries for most applications. Cytotoxicity of 3DP scaffolding materials represents a major challenge, and therefore, the impact of 3DP on tissue engineering has been limited to a fairly narrow group of specific applications. Natural materials such as water mixed with carbohydrates have been developed as binders, addressing the cytotoxicity problem but presenting other challenges regarding postpreparation processing.

2.15.4.1.2.(i) Other SFF techniques Other techniques such as fused deposition modeling (FDM) and ballistic particle manufacturing (BPM) have also been applied to tissue engineering applications with varying degrees of success. FDM technology utilizes an x–y plotter with vertical stepping capability (Zein *et al.* 2002). Molten scaffolding materials such as polymers or ceramics are ejected from a nozzle onto a surface containing FDM layers. Again, polymers such as PCL and high-density polyethylene (HDPE) have been used to produce FDM scaffolds, as have been materials such as calcium phosphate. Porosity is generally not as high as with some of the other techniques, and control over pore sizes and the tortuosity of the pore network are also difficult to control. The continuous deposition of molten material onto the solidified surface leads to challenges in vertical resolution.

Newer techniques have been developed to address some of the shortcomings of FDM and

BPM. One of these techniques, microsyringe deposition (MD), has been established by the group of Bhatia as a means to produce 3D high-precision cell-seeded scaffolds from PLGA. The syringe needle is roughly 20 μm in diameter, thereby leading to horizontal resolution far superior to all other SFF techniques. Another significant advantage of MD is the ability to integrate biological materials such as proteins into the scaffold as it is being constructed, due to the low processing temperatures. These low temperatures also contribute to an excellent degree of control over the vertical resolution. **Figure 8** is a schematic of the fabrication process for PLGA molding, using a PDMS transfer mold and a spinning process to produce a PLGA scaffold (Vozzi *et al.* 2003).

In general, SFF techniques represent an excellent way to produce tissue engineering scaffolds from a wide range of biomaterials, including polymers, metals, and ceramics. Large-scale, highly porous scaffolds have been produced using SFF techniques with tuned bulk mechanical and bulk transport properties, two of the major requirements for all tissue engineering applications. The CAD-driven nature of SFF lends itself to the development of customized scaffolds that can be produced in response to the specific requirements of a given patient, such as replacing the lost skin, cartilage, muscle, nerve, or bone tissue from a particular patient injury. Ultimately, the ability to produce scaffolding of a given size (e.g., for a pediatric patient) will enable made-to-order replacement organs.

The principal limitations of SFF techniques relate to their inability to reproduce the cellular microenvironment at the same size scale as that of physiological systems. For instance, it is known that

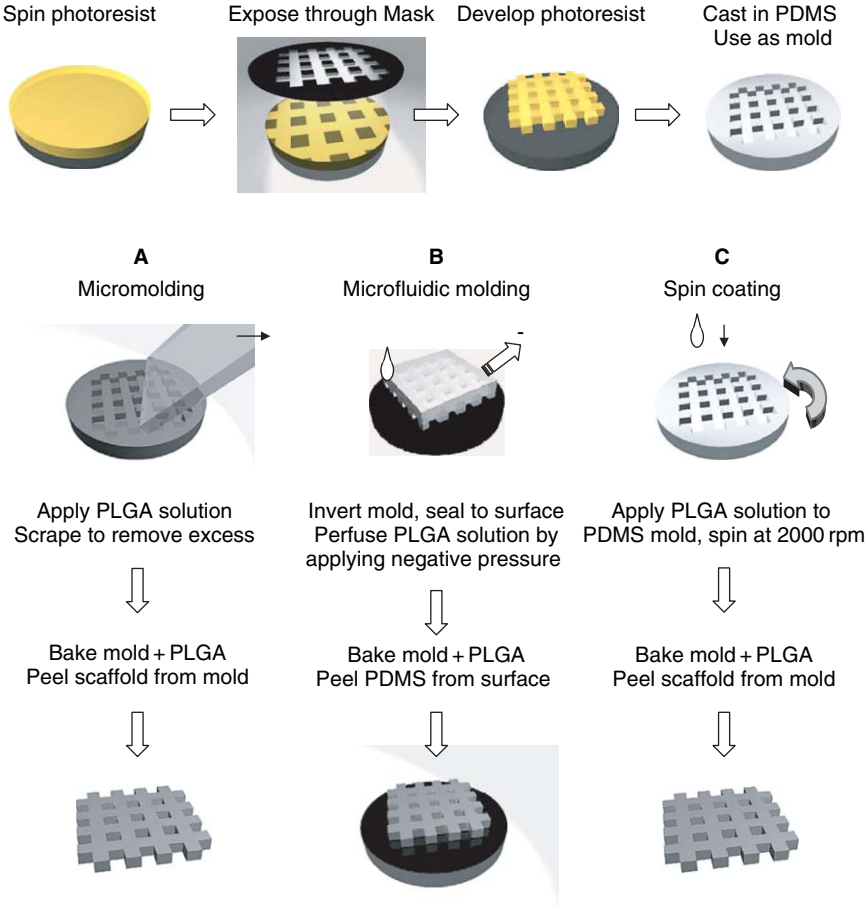


Figure 8 Schematic of microsyringe deposition process illustrating technique for producing 3D microfabricated scaffolds from the biodegradable scaffold poly(lactic co-glycolic acid) (PLGA). PDMS, poly(dimethylsiloxane). (Source: Vozzi G, Flaim C, Ahluwalia A, Bhatia S 2003 Fabrication of PLGA scaffolds using soft lithography and microsyringe deposition. *Biomaterials* **24**, 2533–40.)

specific arrangements of complex cocultures are required to reproduce functions in blood vessels, liver, and other tissues. These cocultures are not random in nature but rather involve precise collocation of multiple cell types in a manner that cannot be architecturally controlled by an SFF scaffold. While it is possible that the resolution of SFF techniques may be improved, this would undoubtedly be accompanied by a significant reduction in throughput. Therefore, it seems necessary that in order to produce scaffolding capable of replicating the cellular microenvironment, methods arising from MEMS fabrication techniques will be required. In Section 2.15.4.2, a brief overview of these techniques is provided.

2.15.4.2 Lithography-Based Technologies for Scaffolding

Photolithography, the central process for fabrication of integrated circuits and MEMS devices, has become a critical element in the development of engineered tissues. Many of the fundamental studies in the field of tissue engineering have utilized lithographic techniques to pattern biomolecules on surfaces to control cell–cell and cell–substrate interactions. These methods are often referred to as soft lithography and have been used to study parameters controlling cell fate, adhesion, migration, proliferation, and differentiation. As described in earlier sections, lithographic processes at the microscale and at the nanoscale have been used to demonstrate cell behaviors such as contact guidance and contact inhibition. Coculture experiments involving heterotypic cell–cell interactions have also utilized lithographic processes to define specific spatial relationships between various cell types and to answer questions such as whether cell functionality is driven by direct cell–cell contact or by communication via circulating biomolecules. In Sections 2.15.4.2.1–2.15.4.2.6, various methods used to fabricate structures for the purposes of probing cell behavior and assembling engineered tissues will be described in brief.

2.15.4.2.1 Conventional photolithography and etching

Lithographic and etching techniques for applications in microsystems have been described in detail in other chapters of this book and in other references. In this section, aspects of these processes specific to the development of engineered tissues are described in brief. Photolithography can be used to pattern substrates for many purposes, including the formation of topographic features, spatial features such as microchannels, or adhesive and nonadhesive regions. Conventional photolithography comprising the application of thin (~2 μm) layers of photoresist followed by plasma etching is the most common method for producing topographic or spatial features on silicon substrates, as described in numerous references addressing contact guidance phenomena and the formation of organoid structures. For micron-scale features, there are no special aspects of the lithographic process other than as a template for reactive ion etching or wet etching. For nanoscale features, advanced lithographic processes such as conformable contact lithography (CCL) (Goodberlet and Kavak 2002) or other methods capable of forming ~100-nm-scale features on the surface are used, particularly when investigating contact guidance phenomena requiring such high-resolution patterning.

Deep reactive ion etching (DRIE) is often used to transfer patterns from lithography into the bulk of a silicon substrate, either for direct cell culture onto the micromachined surface or for the formation of a master mold for replica molding of polymer films. The advantage of DRIE is the ability to maintain surface features with high accuracy, but a disadvantage is the nonphysiological nature of the etched surface, with vertical walls and sharp 90° angles. Isotropic plasma etching provides a more U-shaped sidewall, which is more compatible with cell adhesion but causes the etched via to become larger than the lithographic pattern as the etch depth increases. Wet etching techniques such as hydrofluoric/nitric/acetic acid (HNA) mixtures or crystallographic etchants (e.g., KOH) can also be used to form specific 3D structures within silicon wafers for cell culture of replica molding of polymers.

2.15.4.2.2 High aspect ratio photolithography

Many researchers in the field of tissue engineering do not have access to fabrication facilities with DRIE or other plasma etching tools, and therefore, high aspect ratio photolithography has become a method of choice for producing deep structures on the surface of silicon wafers for direct cell experimentation or for polymer molding. Before the development of SU-8 epoxy resin, now a standard bearer in the formation of deep structures on silicon wafers, researchers were forced to resort to LIGA (an acronym from German

words for lithography, electroplating, and molding) related techniques, in which high-energy beams in synchrotron facilities were used to expose thick polymeric films to obtain the result as is now realized simply and straightforwardly using SU-8. Layers of thicknesses from 25 μm to several hundred microns can be deposited and patterned with a single spin, and multiple spins of SU-8 can produce thicker layers or more complex 3D structures, including closed channel networks. The biocompatibility of SU-8 has been investigated in several studies and has been shown to be acceptable for cell culture applications. However, the principal utility of SU-8 is as a master for replica molding a host of polymeric scaffolding materials for tissue engineering.

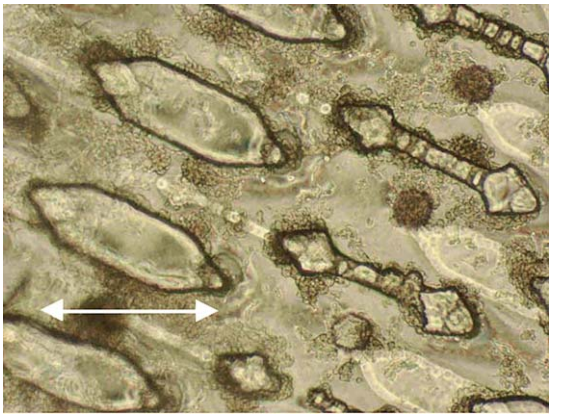


Figure 9 Hepatoma cells cultured in 3D poly(dimethylsiloxane) (PDMS) bioreactor. Bar, 10 μm. (Source: Leclerc E, Sakai Y, Fujii T 2003 A multi-layer PDMS microfluidic device for tissue engineering applications. *Proc. 2003 Transducers Conf.* pp. 415–18.)

2.15.4.2.3 Micromolding and embossing

Early studies in tissue engineering utilized polymeric scaffolds fashioned using SFF techniques, porogen leaching, electrospinning, and other techniques that do not produce micron-scale resolution constructs but rather much lower precision structures with the tuning of bulk properties such as porosity and biomolecular permeability. Scaffold preparation techniques were developed to produce foams and gels, and physiological features such as capillaries and high-precision organ-specific entities such as sinusoids could be produced only by using directed self-assembly from appropriate growth factors and other chemical cues. With the advent of polymer replica molding roughly in the late 1990s, led by Whitesides', Quake's, and Beebe's groups (Jo and Beebe 1999, Quake and Scherer 2000, Whitesides and Stroock 2001), a new avenue for producing high-resolution structures in a host of polymeric substrates has been established. These techniques are based on the principle that polymeric structures can be generated with micron-scale precision by casting, micromolding, or embossing the features from a micromachined silicon master. Predominantly, this technique has utilized the substrate material PDMS, an elastomeric silicone rubber that is transparent and highly biocompatible. Structures with minimum dimensions of 100 nm or less can be readily reproduced from micromachined silicon masters using solvent casting. Cell culture with endothelial cells, hepatocytes, embryonic stem cells, and numerous other cell types has been conducted in PDMS bioreactors. Multiple layers of PDMS micromachined films can be assembled into 3D structures using oxygen plasma bonding, which forms hydroxyl groups on the surface that result in permanent, high-strength bonds between layers. High oxygen diffusivity through PDMS enables sufficient oxygenation of cells within PDMS bioreactors even in the absence of oxygenated media, such as in the case of hepatocyte culture in 3D microfluidic devices. **Figure 9** shows a 3D microfluidic bioreactor for hepatocyte culture developed by Leclerc *et al.* (1993). The Takayama group at Michigan (Gu *et al.*, 2004) has reported a microfluidic cell culture system in PDMS using a refreshable Braille display; an image of the PDMS perfusion culture device is shown in **Figure 10**.

For applications in tissue engineering, a biodegradable scaffolding material is most often desired, and therefore, methods for molding or embossing a host of biodegradable polymers have been developed. As described in Section 2.15.3, these biodegradable polymers include PLGA, an FDA-approved polymer used for numerous applications such as vicryl sutures. Advantages of PLGA include its relative degree of biocompatibility, the tunability of the degradation time as a function of the lactic/glycolic ratio in the copolymer, and its use as a scaffold in the field of tissue engineering that is evidenced by a vast number of studies. Disadvantages include mechanical rigidity, sudden loss of mechanical strength and mass during the course of the degradation process, and acidic by-products of the hydrolysis reaction through which the material degrades.

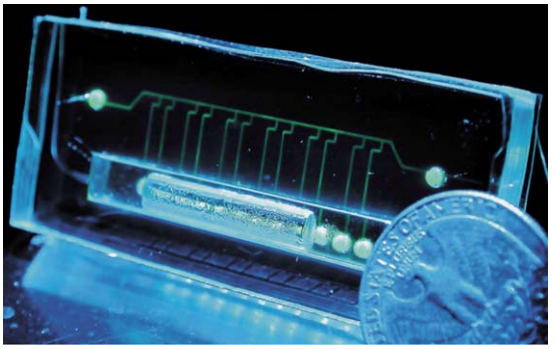


Figure 10 A microfluidic cell culture chip with fluorescein-loaded channels next to a US quarter. The channels are arranged to align to pin actuators on refreshable Braille displays, a tool typically used by the blind to read computer text. When the pin movement is interfaced appropriately with the microfluidic chip, one obtains reprogrammable microfluidic cell culture systems where the pins act as reconfigurable valves and pumps. (Source: Gu W, Zhu X, Futai N, Cho B S, Takayama S 2004 Computerized microfluidic cell culture using elastomeric channels and Braille displays. *Proc. Natl. Acad. Sci. U S A* **101**, 15861–6.)

Numerous techniques for constructing micromachined scaffolds from PLGA have been demonstrated. In one implementation, scaffolds with a minimum feature size of roughly 500 μm were constructed using a molding process, and multiple layers were sealed together by depositing metal films and alloying them together. In another embodiment, multilayer PLGA scaffolds were assembled without the use of any adhesive materials, by fusing the layers together at elevated temperature and pressure. The PLGA layers were constructed by a double-transfer mold process in which SU-8 lithographically patterned silicon master molds with the positive image of an interconnected microchannel network were used to produce negative molds in PDMS. The PDMS transfer mold is necessary because PLGA is rigid and cannot be separated from the silicon master mold. This double-transfer mold process preserves high-resolution features and is a robust means of producing fully degradable scaffold structures. A cross section of a multilayer microfluidic device constructed from PLGA is illustrated in **Figure 11** (King *et al.* 2004). Other embodiments assemble multiple layers by spinning films of PLGA successively to form 3D scaffolds, sometimes with cells and other biological materials interspersed within the layers.

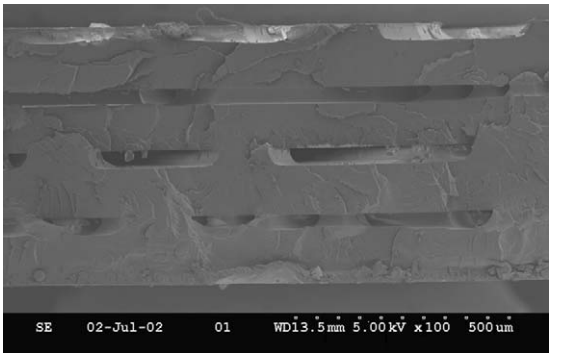


Figure 11 Cross section of 3D scaffold constructed from compression-molded poly(lactic co-glycolic acid) (PLGA).

2.15.4.2.4 Electroplating

Micromolding techniques for PDMS and biodegradable polymers based on photolithographically defined silicon master molds suffer from some limitations involving the ability to define true 3D structures. For instance, in the case of microvascular networks, simple lithographic patterning and etching of master molds results in microchannels with variable width but a fixed depth, a highly nonphysiological situation. The fluidic resistance of such channels is much larger than that of circular channels with a similar hydraulic diameter, and other properties such as the vessel wall mechanics, clotting cascades, and oxygen and nutrient transport capabilities are also widely divergent from natural vasculature. In order to achieve a closer approximation to physiological systems, multiple lithographic steps with precise alignment of layers must be utilized, to produce a single layout with varying-depth structures.

In order to obviate the need for multiple lithographic and/or etching steps, molds produced by electroplating processes may be used to produce complex multidepth structures. The process utilizes patterns of conducting islands on a nonconducting substrate patterned by a single photolithographic step. As electrodeposition progresses, islands of conducting material are joined to form larger and hence thicker structures, thereby enabling the formation of features of aspect ratios near 1:1. These electroplated structures may be used as master molds for solvent casting of PDMS and other biopolymer scaffolds with dimensions approximating physiological systems. **Figure 12** illustrates a mold generated by electroplating a pattern with features of varying height and width on a silicon wafer (LaVan *et al.* 2003).

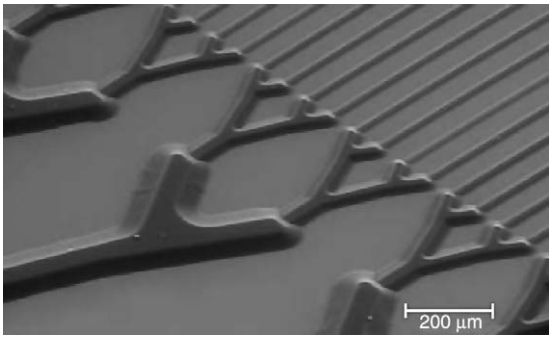


Figure 12 SEM of electroplated silicon mold for fabrication of polymer micromolded vascular networks, showing the capability of varying the height as well as the width of the features. (Source: LaVan D A, George P M, Langer R 2003 Simple, three-dimensional microfabrication of electrodeposited structures. *Angew. Chem. Int. Ed. Engl.* **42**, 1262–5.)

Figure 13 Multiphenotype array of embryonic stem cells, AML12 hepatocytes, and NIH-3T3 fibroblasts cocultured in reconfigurable poly(dimethylsiloxane) (PDMS) bioreactor. (Source: Khademhosseini A, Yeh J, Eng G, Karp J M, Kaji H, Borenstein J, Langer R 2005 Cell docking inside microwells within reversibly sealed microfluidic channels for multiphenotype cellular arrays. *Lab Chip* **5**, 1380–6.)

2.15.4.2.5 Assembly techniques

For applications in tissue engineering, microsystems-based polymeric scaffolds must take on three-dimensionality and be capable of reproducing complex structures. Therefore, methods to assemble large, complex 3D structures must be developed. Lamination techniques are required for stacking and integration of multiple microfluidic layers of polymeric scaffold materials. In the case of PDMS polymer layers, assembly may be accomplished simply by exposing layers to oxygen plasma and placing the layers in contact. Treatment with an oxygen plasma results in an extremely strong bond between layers, after which the interlayer bonds are stronger than the bulk material. Without plasma bonding, PDMS layers may be reversibly sealed together to form closed lumens capable of supporting fluid flow, but layers may be separated from one another to enable additional processing steps. Khademhosseini *et al.* (2005) have developed a technique for the culturing of multiphenotype arrays, in which microwells are micromolded into PDMS films, and then a top layer of PDMS with microchannels is placed over the microwell array to provide a means to fluidically address the microwells. Once these microwells are seeded with a given cell type, the microchannel layer is lifted off the top of the microwell array and additional microfluidic processing can occur. An image of multiple cell types cultured in such a reconfigurable microfluidic device is shown in **Figure 13** (Khademhosseini *et al.* 2005). This reconfigurable nature of PDMS has many benefits in the development of microfabricated cell culture systems.

Many of the biodegradable polymer systems have different and potentially more complex means of assembly into multilayer constructs. In the case of PLGA, thermal fusion bonding joins multiple layers together. Pressure and elevated temperature are used to establish a strong, irreversible bond between adjacent layers, as described by King *et al.* (2004); the process window for this treatment requires that the temperature and pressure excursion be sufficient to fully seal the layers without causing them to deform or collapse. In the case of PGS, sealing of layers is accomplished by joining partially wetted layers together to form large stacks. Crosslinking using multiple polymerization steps can be used to bond PGS layers together (Bettinger *et al.* 2006b). Conjugation of moieties using specific chemical routes can be used to bond layers, but there are significant limitations related to the chemistries and temperatures involved.

In each of the aforemetioned cases, integration of the multilayer stacks must be achieved and connectors placed to enable the introduction of cells and media. Devices for laboratory cell culture studies have different constraints than do systems that will ultimately function as extracorporeal or implantable organ assist or replacement devices. Contact with blood and limitations in the total fluid capacity of the latter systems require that numerous design rules be followed, because of concerns related to hemocompatibility and the potential presence of large

dead volumes or inadequate mixing. In general, such 3D microfluidic devices are assembled in such a way as to minimize dead volume and transients within the system due to distensible connectors and to avoid sharp corners that promote turbulence and thrombosis. Conventional tubing with large bore diameters may be avoided because of the large volumes of fluid in the connectors relative to the microdevice and because of potential incompatibility between the connector material and blood or other biological fluids involved. Attachment between the external tubing connectors and the microdevice may be done with microfabricated and molded connectors rather than by manual techniques involving macroscopic components. The trend is toward integration of the entire microfluidic network into a microfabricated platform with a minimum of connectors that may potentially serve as a nidus for a clot or an infection.

2.15.4.2.6 Nanostructured matrices

Engineered scaffolds and matrices with microscale dimensions provide spatial, topographic, and fluidic guidance at size scales similar to those of the cells populating these devices. However, many cell–matrix interactions occur at far smaller size scales, for reasons described earlier that involve interactions with the cytoskeletal components as well as through other nanoscale mechanisms. Matrices that promote the development of structural and functional tissues through the control of growth, differentiation, and development comprise a critical element of engineered tissues, and nanostructured matrices with highly developed properties have been applied toward cardiac, nerve, vascular, and other tissues. First reported by Zhang *et al.* (1993), these nanostructured matrices are formed via self-assembly of oligopeptides into hydrogels with high water content (Holmes *et al.* 2000, Kisiday *et al.* 2002). The nanostructured peptides consist of alternating units of positively and negatively charged amino acid groups. In one study (Holmes *et al.* 2000), the development of cultured neuronal cells in these nanostructured matrices was compared with prior studies in which neurons were cultured in other matrices. Proteins such as laminin, fibronectin, and collagen have also been used as matrices for neurite outgrowth, and these proteins are better as attachment motifs when compared to untreated tissue culture plastic, glass, and other standard substrates. Nanostructured peptides result in excellent neurite outgrowth with functional synapse formation, and *in vivo* studies indicate that nerve repair is possible using these novel materials. **Figure 14** illustrates the neurite outgrowth resulting from culturing with a nanostructured peptide (Holmes *et al.* 2000).

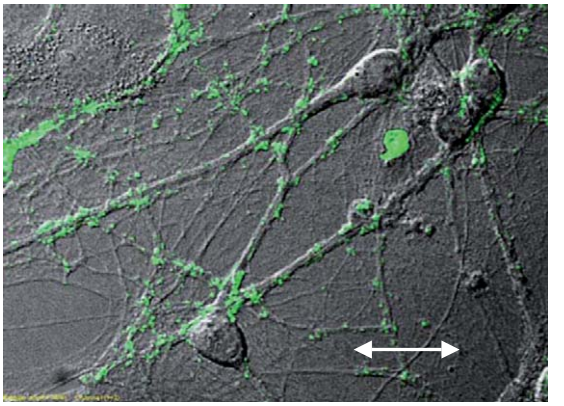


Figure 14 Primary rat hippocampal neurons form active synapses on the designer self-assembling peptide nanofiber scaffolds. The neurons were cultured on the peptide scaffold for ~2 weeks before active synapses were assessed. The confocal images show bright discrete green labeling indicative of synaptically active membranes following FM1-43 incubation of neurons. Active synapses on the peptide surface. The active synapses on these scaffold materials with high density indicate that the peptide scaffold is a permissible substrate for synapse formation. The same scaffold was later used in hamster mouse brain damage repair to bridge the optical nerve cap. Bar, 10 μm. (Source: Holmes *et al*. *Proc. Natl. Acad. Sci. U S A* 2000.)

2.15.4.3 Cell Printing Technologies

One of the most novel and exciting new methods for producing engineered tissues with high precision is the deposition of cells directly from a nozzle moving over a surface (Mironov *et al.* 2003). In essence, these techniques do not even use a scaffold to organize cells into agglomerates for tissue engineering applications but rather deposit cells and other materials directly onto the engineered construct. In one variant of this method, microsystems technology is applied to tissue engineering through the implementation of a microfabricated element (the ink-jet printer head) that deposits cells in a 3D matrix. Several methods for depositing the cells have been demonstrated, including spraying, ink-jet printing, and laser-guided deposition (optical tweezers). In the spray deposition technique, a liquid suspension of cells is deposited on the surface using an airbrush-type spray head. In one successful demonstration of this technique, NIH 3T3 fibroblasts and primary hepatocytes were cocultured

in a collagen sandwich in which collagen gel layers were deposited, interspersed by layers of sprayed cells (De Silva *et al.* 2005). Subsequent functional assays demonstrated that the cells were not damaged by the process; in fact, liver-specific function as monitored by cytochrome *P*450 activity was better than in control samples. An alternative to spraying, ink-jet printing of cells has been shown to maintain high levels of cell viability despite the high degree of stress associated with the deposition process (Xu *et al.* 2005). This system has been applied to Chinese hamster ovary (CHO) cells with excellent results. These systems essentially utilize a commercial ink-jet printer with a control program that draws the engineered cellular assembly. Higher-precision, noncontact direct deposition of cells has been achieved using optical tweezer systems. Cells are essentially steered toward the substrate by a laser/lens system that directs the flow of cells onto the surface. Extraordinarily high-resolution assemblies, essentially achieving 1-μm precision, can be achieved using this technique. Overall, however, each of these cell printing methods has as a drawback a slow rate of deposition, which results in lengthy processing times for scaffolds of relevant size scales.

2.15.5 Microsystems for Cell and Tissue Engineering

Microsystems for engineering cells and tissues in the laboratory are based on a combination of microfabrication techniques, biomaterial substrates and matrices, and cells and biochemical factors. Since the mid-1990s a multitude of experimental tools and devices based on microsystems technologies have been developed toward the goal of cell and tissue engineering, as will be described in this section. These systems may be inherently 2D in nature, as in the case of micropatterned surfaces for control of cell adhesion, migration, and function and chemical gradients to promote cell motion and the organization of cells into tissues. Three-dimensional systems comprise structures produced using the SFF techniques described earlier, as well as more classical microsystems-based approaches such as microfluidic bioreactors and microvascular networks. In addition, techniques such as polymer demixing and microscale templating have been employed to form a 3D architecture.

2.15.5.1 Two-Dimensional Microsystems

In Section 2.15.4, numerous techniques for forming 2D microfabricated structures for tissue engineering were described. These techniques can be broadly classified into those methods that form layers via photolithographic patterning of biomolecules on surfaces, techniques that generate chemical gradients using microfluidics, and laser patterning approaches. In Sections 2.15.5.1.1–2.15.5.1.3, these approaches will be described in greater detail, with an emphasis on how these approaches can be integrated into fully 3D systems in the subsequent discussion.

2.15.5.1.1 Micropatterning of surface adhesion molecules

Numerous surface modification moieties including PEG, RGD peptides, and biotin have been used to generate engineered cell and tissue constructs. These surface modifiers have been micropatterned using direct photolithographic techniques as well as microcontact printing. Scaffolding materials such as PDMS, PLGA, and chitosan have been modified using these techniques; however, these techniques are typically limited to 2D devices. The most common micropatterning approach is to pattern adhesive and nonadhesive regions on a tissue-engineered substrate, using techniques such as silane-based coupling agents or thiol-based systems. Surfaces with stable adhesive or nonadhesive properties, based on the degree of hydrophilicity, for instance, can be patterned by microcontact printing of surface adhesion molecules on specific regions, thereby creating a construct for patterned cell culture. Techniques other than surface adhesion molecule patterning can also be used; for instance, Curtis *et al.* (2001) have developed a technique in which specific topographic nanostructures can be used to prevent cell adhesion.

Specific ligand–receptor combinations such as RGD (arginine–glycine–aspartic acid) peptides and their integrins comprise a major system for modifying surfaces to control cell adhesion. This system can control cell growth and apoptosis, as well as differentiation and other critical functions. Covalent linkages of RGD have been used to modulate cell–surface interactions in order to mimic physiological interactions with the basement membrane of numerous tissues and with blood. Liver coculture systems have been studied using adhesive patterning of ECM analog collagen I. These studies have investigated the interactions between hepatocytes and fibroblasts with specific spatial relationships between the two cell

types. These studies utilized glass substrates patterned using photolithography, followed by collagen deposition and liftoff to remove specific regions of collagen. Hepatocytes were cultured on the surfaces, following the pattern of the lithographic mask, and then fibroblasts were added to form the coculture. These systems have been used to investigate and establish some of the fundamental principles involved in liver cell biology and underlying the critical elements of engineered liver assist devices.

2.15.5.1.2 Chemical gradient systems

The use of chemical gradients to control cell migration and to organize tissues is a critical element of the development of engineered tissues. Systems utilizing chemoattractive substances (chemotaxis), distributions of adhesion molecules (haptotaxis), or chemorepellent systems (fugetaxis) are a vital tool in the organization of cells within devices or matrices, in addition to fluid mechanical methods for distributing cells. Gradients of soluble factors have been the subject of numerous studies, but historically, gradients of substrate-bound molecules have been difficult to generate and maintain because of multiple experimental challenges. Pioneering work by the Whitesides Laboratory (Dertinger *et al.* 2002) has established a simple and highly novel technique for establishing and maintaining chemical gradients using microfluidic mixing chambers. Dertinger *et al.* reported the fabrication of a gradient mixer with long serpentine channels; the length of the channels must be sufficient to enable the substances to mix completely via diffusion. Networks are specifically designed to generate gradients suitable for particular applications; an image of the gradient mixer is shown in **Figure 15** (Dertinger *et al.* 2002). Gradients in substrate-bound moieties are achieved through adsorption of these soluble species from the liquid mixer. This gradient mixer that enabled the initial investigation of a smooth gradient, rather than of a sharp gradient, in ECM proteins is able to facilitate the early stages of neuronal development.

2.15.5.1.3 Laser-patterning techniques

The aforementioned techniques require the use of photolithographic equipment and photomasks, which are expensive and typically require a highly controlled environment for processing. Numerous types of lasers can be used to etch polymers by ablation processes that minimize damage to nonablated regions and maximize resolution by controlling the extent of the heat-affected zone and by sharpening

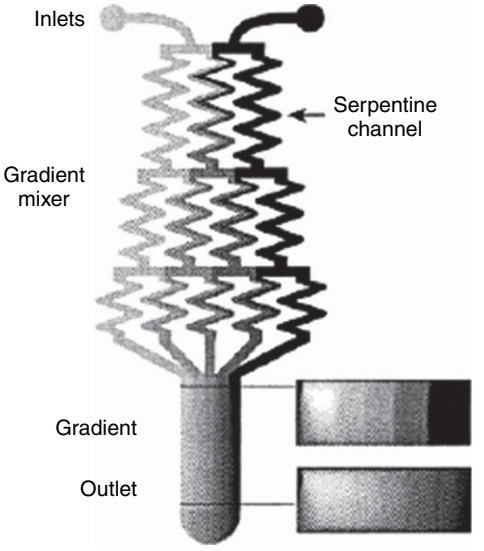


Figure 15 Schematic of microfluidic mixer using laminar flows in microchannels to generate substrate-bound chemical gradients. (Source: Dertinger S K W, Jiang X, Li Z, Murthy V N, Whitesides G M 2002 Gradients of substrate-bound laminin orient axonal specification of neurons. *Proc. Natl. Acad. Sci. U S A* **99**, 12542–7.)

the boundary between the laser-machined region and its surroundings. These techniques have also been used to fabricate microvalves and micropumps for other applications such as drug delivery (Mescher *et al.* 2003). Structures with minimum dimensions of 10 μm have been produced through the use of laser ablation. Lasers can also be used to modify surfaces through the incorporation of specific gas-phase chemistries under ambient conditions.

Another laser-based approach, known as nanosphere lithography, utilizes an organized array of nanoparticles to modify a laser beam irradiating the surface of a polymer, thereby creating nanoscale features on the surface of the substrate (Frey *et al.* 2000). Novel methods are used to generate the ordered array of nanoparticles, and numerous process parameters such as the properties of the nanoparticles, the laser energy, and the optical path can be used to tune the process. The principal limitation with nanosphere lithography is similar to that of certain other nanolithographic techniques in that the range of potential patterns that can be generated is bounded by the physics of the optical technique. However, this process is capable of generating nanoscale surface patterns that may be extremely useful in influencing cell behavior on biopolymer substrates at a cost much lower than that of comparable direct nanolithographic patterning techniques.

2.15.5.2 Three-Dimensional Microfluidic Systems

Three-dimensional cell culture systems and engineered devices and tissues are emerging as potential pathways toward novel therapies for patients suffering from end-stage tissue loss and organ failure. The 3D aspect of these devices and systems bears significance from multiple perspectives, including the behavior of 2D versus 3D cell culture, the requirement for a vasculature to deliver oxygen and nutrients, and the size and scale of devices necessary to augment or replace physiological function.

2.15.5.2.1 Biological basis for 2D vs. 3D culture

Recent evidence from the field of cancer biology has led to the inescapable conclusion that 3D cell culture is capable of replicating physiological function far more closely than 2D culture (Griffith and Swartz 2006). Cells grown on flat layers are deprived of the full array of environmental cues available from the 3D microenvironment of living tissue, including mechanical and biochemical cues originating from the ECM surrounding cells in most tissues. The pioneering work of the Bissell group (Weaver *et al.* 1997) first demonstrated that integrin receptors caused breast cancer cells to lose their cancerous properties, a finding that had never been observed in conventional 2D flat-plate culture. Three-dimensional cell culture has also demonstrated that metastasis of breast cancer cells occurs via mechanisms that require 3D systems in order to observe, namely, complex shape changes and migration enhancement that cannot be observed in 2D culture. Cell migration and proliferation in craniofacial tissue engineering studies have been dramatically enhanced in 3D culture as well.

Not just the spatial environment in the bioreactor but also the nature of the matrix itself has been shown to play a significant role in enhancing the performance of 3D systems. The paradigm of self-assembling biological systems for tissue engineering and other applications is emerging, with promising results in many areas (Zhang 2002). Amino acids that self-assemble into nanofibers, as reported by Stupp (Hartgerink *et al.* 2001), result in 3D properties that have led to advances in nerve tissue formation. Nanofibers of self-assembling peptides with 3D properties, as described in Section 2.15.3, have resulted in advances in cardiac and nerve tissue engineering, as well as in the formation of microvasculature. Microvascular network formation represents a significant challenge for which microfabrication techniques can make a major contribution.

2.15.5.2.2 Requirement for vascularization

Three-dimensional scaffolding and matrices have clearly been demonstrated to provide enhanced properties for tissue engineering in numerous applications. Another key observation that arises from the tissue engineering research over past several decades is the crucial role of vascularization in the development of engineered tissues and organs. Avascular tissues such as skin (Auger *et al.* 1995, Bell *et al.* 1981, Parenteau *et al.* 1996, Purdue *et al.* 1997) and cartilage (Athanasiou *et al.* 1998, Klien and Schumacher 2003) have been successfully reproduced in the laboratory using numerous approaches. However, thick, complex tissues that require a vascular supply have been far more difficult to produce in tissue engineering laboratories. Particularly for highly metabolic tissues such as cardiac or liver tissue, when the thickness of the growing construct exceeds a few hundred microns, cells along the inner regions become necrotic due to hypoxia. Nature addresses this problem by providing a rich, intrinsic vascular supply that reaches into these tissues such that no region is farther than 100–200 μm from a blood vessel. These organs possess a vasculature that scales from large arteries and veins down to capillary networks with diameters as small as 10 μm or less. These facets of complex tissues and organs have led to the conclusion that engineered tissues will require a means to develop an intrinsic vascular supply.

2.15.5.2.3 Microfluidic bioreactors for cell and tissue growth

Bioreactors for cell and tissue engineering have benefited dramatically from microfabrication and microsystems technologies, because of the ability to control the spatial, chemical, topographic, and fluid mechanical microenvironment in ways that mimic physiological systems more closely (Desai 2000). Microfluidic bioreactors for hepatic and renal support are emerging as viable means to treat acute or chronic failure of these organs, paving the way toward miniaturized and ultimately portable or even wearable or implantable systems. However, many enormous technical challenges remain, and therefore, the most important role of microfabrication

technology in many current studies is to elucidate more completely the critical elements of engineered cell-based systems that will provide long-term sustainable function. In the field of stem cells, bioreactors capable of promoting stem cell differentiation will ultimately provide crucial insights into the cell-sourcing issues that face many of these therapeutic avenues.

Hepatocyte bioreactors, a first step toward liver assist and liver replacement devices, have been constructed using a number of microfabrication-based approaches. Critical elements of hepatocyte bioreactors include fluid mechanical, structural, and biochemical properties that mimic the liver microenvironment, including cell–substrate interactions that simulate hepatocyte anchorage on the basement membrane as well as microvascular networks to provide oxygen and nutrients to the highly metabolic cells. The Fujii group (Leclerc *et al.* 2003) has reported the fabrication and testing of a multilayer PDMS bioreactor seeded with a human hepatocarcinoma cell line (HepG2) along with 3T3 mouse fibroblasts to provide the connective tissue element. Oxygen was delivered to the cells from the environment through the PDMS bioreactor walls, taking advantage of the extremely high diffusivity of molecular oxygen through the bulk polymer. Cells were observed to attach, consume glucose, and produce albumin over periods of days to weeks, confirming the basic viability of this type of device for culture of the robust hepatocyte cell line.

The work of the Fujii group leverages the significant foundation of microfluidic technology in the PDMS platform. For implantable applications, biodegradable substrates will ultimately be preferred, and Bettinger (Bettinger *et al.* 2006b) has reported on HepG2 cell culture in a PGS bioreactor consisting of multiple layers of microfluidic channels. In this case, oxygen diffusivity is markedly lower than in PDMS, and therefore, the cells must be supplied with oxygen through the media flowing through the microfluidic network. Hepatocyte viability was monitored by observation of long-term cell attachment to the constructs as well as through albumin production. In order to promote hepatocyte adhesion while minimizing the possibility of shear-related damage, the hepatocyte chamber was populated with a staggered array of posts surrounded by pressure-relief channels. This study confirmed for the first time that a large-scale biodegradable microfluidic bioreactor could be used in long-term culture of hepatocytes.

Culture of primary liver cells represents a more difficult challenge. Primary cells require specific substrate configurations, such as spheroids/aggregates, 3D matrices such as the nanostructured self-assembled peptides (sapeptides), or support cells such as fibroblasts in coculture (Bhatia *et al.* 1998b). Bioreactor configurations comprise single-unit systems in which all functions take place within one controlled microenvironment, as well as systems that divide the functions such as ammonia conversion and cytochrome *P*450 detoxification into separately controlled chambers. Bioreactor configurations are typically modeled based on physiological configurations, and therefore, radial flow parallel plate bioreactors are a common approach. Challenges with these approaches such as those of the cell density can be addressed by microfabrication, in that narrower channel heights result in higher cell densities per unit volume. The limitation in this approach is not the fabrication technique, rather it is the high shear stresses that develop in bioreactor chambers in which the channel height drops to 50 μm or less. The principal engineering limitation in all of the liver bioreactors is the requirement for high oxygen transport to feed the high oxygen consumption rate (OCR) of hepatocytes (Patzer 2004, Roy *et al.* 2001). Numerous efforts have been made to increase oxygen delivery to the cells, including hollow fiber-based approaches in which oxygen is directed through additional pathways toward the cells. Internal membrane oxygenation has shown very promising results in several studies and in multiple membrane configurations in which a second membrane is employed to reduce the quantity of toxins that reach the cells. Many of these efforts are aimed at reducing the ultimate cell volume that will be required to provide replacement liver function for patients suffering from liver failure.

Engineered microenvironments for stem cell control and regulation represent an extremely important application of microscale fabrication and microsystems technology. Cell sourcing represents one of the most significant barriers in the field of regenerative medicine, with traditional sources such as autografts, allografts, and xenografts experiencing limitations in availability and applicability. Both adult and embryonic stem cells are sources with enormous potential for tissue engineering. Microfabricated bioreactors containing mechanical and biochemical cues to regulate stem cells are now emerging. Embryonic stem cells can be encouraged to differentiate into a vast array of cell types based on the appropriate sequence

of environmental signals and cues. Adult stem cell microenvironments, such as those generated through nanostructured matrices described in a later section (Holmes *et al.* 2000), are capable of replicating the niches required to control their fate.

2.15.5.2.4 Large-scale microfluidic devices for cell-based systems

Cell-seeded scaffolds for tissue engineering can be produced using a wide variety of techniques, comprising a spectrum of approaches that rely to a varying extent on precision microtechnologies. The principal engineering limitation facing efforts to develop laboratory-grown complex tissues is the requirement for an intrinsic vascular supply for the growing tissue. Natural tissues solve mass transport challenges by providing a vast network of blood vessels organized according to principles that minimize the energy expended and maximize the effectiveness of oxygen and nutrient exchange. Therefore, these networks obey the principles that govern the structure, scaling, and integration of such vessels into an organization of channels that can be characterized by design rules involving vessel length and the average diameter of successive generations of vessels.

Great progress has been made in the development of engineered tissues using angiogenic methods to provide an intrinsic vascular supply. However, these approaches, which rely on the use of growth factors such as vascular endothelial growth factors (VEGF) and platelet-derived growth factors (PDGF) to produce vascularization (Peters *et al.* 2002, Richardson *et al.* 2001), still face challenges regarding long-term mechanical stability of the vasculature. Recent progress in multiculture systems (Levenberg *et al.* 2005) has resulted in successful generation of vascularized skeletal muscle tissue. For vital organs, the properties of the vascular supply are more challenging, however, and therefore, an engineering approach to this problem, through the use of microfabricated scaffolds containing vessel networks, has been developed. Such engineered vascular networks can be generated using polymer micromolding or embossing techniques followed by lamination of assembled micromachined layers (Borenstein *et al.* 2002, Lim *et al.* 2003, Wang *et al.* 2005), on a variety of scaffolding materials, as described in the next section.

Microvascular networks provide the basis for oxygen and nutrient transport in complex, 3D tissues and organs. Over the past several years, the use of photolithography and polymer replica molding has been applied toward the development of large-scale 3D vascular networks. Computational fluid dynamic models have been used to provide design rules for these networks, beginning with simple fluid mechanical principles but also encompassing a number of complex phenomena. These models are based on morphometric analysis of the microvasculature of vital organs, as obtained using casting techniques that employ waxes and other materials infused into the vascular bed, followed by etching or removal of the tissue, and subsequently detailed measurements of the cast replica of the vessels. Numerous morphometric studies have been reported (Escaned *et al.* 2001, Kassab and Fung 1994, 1995, Kassab *et al.* 1993), often accompanied by microfluidic scaling laws that provide a framework for microfabrication-based approaches to replicate the physiology of the network (Kaazempur-Mofrad *et al.* 2001). Laminar pipe flow as governed by the Hagen–Pouiseuille equation establishes the starting assumptions for network design. However, the mixed-phase nature of blood with a solid fraction of hematocrit and platelets, the distensibility and compressibility of vessel walls and erythrocytes, respectively, and small channel effects such as the Fahraeus–Lindquist effect all lead to far more complex analytical models. These networks have a fractal-like geometry, much like coral and many other structures found in nature. Computational fluid mechanical models incorporate these phenomena and these principles in order to provide a template for the construction of scaffolds comprising microvascular networks.

Once computational models have established a design target for the microvascular network, microfabrication technology is used to produce polymeric replicas of the bifurcated channel arrays. As described in a previous section, 2D network designs are applied to photomasks for photolithographic pattern transfer onto a silicon master mold wafer. Molds are generated using high aspect ratio lithographic negative molds, plasma etching techniques, or electroplating, as previously discussed. In physiological vascular networks, the blood vessels are cylindrical, while microfabrication technology typically produces rectangular vessel geometries. In addition, the ability to produce vessel networks with varying channel depths is limited by the increasing complexity of lithographic processes when they are used to superimpose structures with changing heights on a planar surface. Micromachined or electroplated molds can obviate this difficulty, thereby enabling

microvascular scaffolds with vessel geometries more closely approximating physiological systems.

These approaches have been used to generate single layers of planar microvessel networks; by stacking and laminating multiple layers with vertical interconnects, highly integrated 3D microvascular networks are developed. Applications aimed at *in vitro* studies or for *ex vivo* organ support may utilize biocompatible nondegradable scaffolding materials, but for implantable natural organ replacements, fully biodegradable scaffolds are preferred. In the former set of applications, PDMS is a preferred embodiment, due to its simplicity and ease of fabrication and assembly, and PMMA has also been investigated. For microvessel networks in a biodegradable embodiment, PLGA, PGS, silk fibroin, and calcium alginate hydrogels have all been successfully demonstrated. Endothelialization of the microvascular network has been shown to result in confluent coatings of endothelial cells on channels of varying dimension, providing a pathway toward functional capillary beds for oxygen and nutrient transport within tissue-engineered organ replacements.

Application of the microfabrication technology described in the previous section can be applied to specific organ structures as well as to blood vessel networks. Organs such as the heart, the liver, kidneys, and lungs are essentially microfluidic processing chambers, with vast arrays of microscale chambers for transport, filtration, generation, and excretion of fluids. Examples of these microfluidic structures include the liver sinusoids, renal proximal tubules of the kidney, and alveolar beds within the lung. In each of these instances, microvascular networks are essentially interpenetrated within networks or compartments of organ-specific structures. The microvascular networks supply oxygen and nutrients to the parenchymal cells of the organoid.

In the simplest embodiment, networks of microfabricated organ-specific structures such as liver sinusoids are cultured with hepatocytes and the cells are oxygenated by the surrounding environment. In the case of PDMS scaffolds, the permeability of oxygen in the bulk is high enough so that oxygen delivery through a separate microvessel network is not required (LeClerc *et al.* 2003). Oxygen and small molecule transport through alginate hydrogel is sufficient to sustain long-term cell viability. However, most structural biodegradable scaffolds made of PLGA or PGS, for instance, do not have oxygen permeability sufficient to support such a structure, and therefore, microvessel networks are placed proximal to the organ microstructures. In these cases, oxygen must be transported directly from the microvessel network into the parenchymal cells through open pores. Porous structures generated by porogen leaching, for instance, are possible, but a precision microfabricated device with micromolded compartments for the blood and parenchyma typically uses porous membranes to govern transport between compartments. Such membranes may be composed of materials such as polycarbonate, polysulfone (PS), or PES, depending on the requirements for protein fouling and hemocompatibility. Pores may be tortuous networks or line of sight as is formed by ion track etching, depending on the thickness and specific filtration requirements of the system.

An example of an acellular system constructed using the technology described above is the MEMS-based renal dialysis system (Kaazempur-Mofrad *et al.* 2004). In it, a microchannel network is juxtaposed beneath a dialysate layer, with an intervening PES membrane designed with a specific molecular weight cutoff (MWCO) for filtering molecules such as urea and creatinine in the blood. By increasing the MWCO of the filter, middle molecules may also be filtered more effectively than in conventional dialysis. The principal advantage of the microfabricated system is the ability to reproduce the dimensions of the glomerular capillaries more closely than in a conventional hollow fiber dialyzer. Cell-based systems such as liver bioreactors and renal replacement systems have also been demonstrated in a microfabricated construct, as will be described in Section 2.15.6.

2.15.5.2.5 Three-dimensional hydrogels

The application of photolithography-based processing to cell-seeded systems has resulted in promising new avenues for merging microfabrication techniques with cell culture in an integrated fashion. In one such technique, cell encapsulation techniques based on PEG have been merged with photolithography to construct and 3D scaffold with microscale features in a hydrogel network (Liu and Bhatia 2002). Living cells suspended in an uncrosslinked polymer solution are localized using photolithography to generate large-scale spatial organization of the cellular constructs. Numerous layers can be fabricated by a stepwise process in which each layer is patterned, the uncrosslinked cell-containing materials are washed away, and additional layers are photopatterned. Features as small as 50 μm in minimum dimension have been fabricated using this method, which has

the potential of being integrated with adhesion molecules, growth factors, and other substances.

An alternative approach that does not require the use of photocrosslinkable materials involves the use of micromolding technology to form structures such as microchannels in calcium alginate hydrogels (Cabodi *et al.* 2005a). Here, layers of the calcium alginate are micromolded against a PDMS transfer mold, which itself is produced from an SU-8 lithographically patterned silicon master mold. The calcium alginate films form a network of microchannels to support the flow of blood or media within the microfluidic construct. Multiple layers are bonded together by partially melting them with sodium citrate, a calcium ion chelator. Finished 3D hydrogel devices were used to study the transport of fluorescently labeled molecules, which flow through the microchannel networks and within the bulk of the hydrogel. Assisted extraction, in which pressure-driven flow occurs through the device, is compared with unassisted extraction, where there is no flow. This system represents a model for the development of a 3D scaffold in which microchannel networks and bulk transport can be engineered to support cell growth and tissue generation.

2.15.6 Applications of Microsystems Technology for Tissue Engineering

Applications of microfabrication technology and engineered microsystems in the field of tissue engineering span a range from skin, bone, and cartilage to complex vital organs such as the liver, kidneys, and lungs. The potential of microtechnology in this field is only beginning to emerge, and therefore, these investigations represent the first explorations in what promises to be a rapidly growing body of work in which the power of microtechnology in generating microenvironments suitable for cell culture is applied toward tissue regeneration and replacement. In the following section, several early applications of microscale tissue engineering will be discussed, with an emphasis on microfluidic devices with engineered biomaterials.

2.15.6.1 Skin, Bone, Cartilage, and Wound Healing

Skin and cartilage represent early successes in the field of tissue engineering, in that numerous commercial products have been developed and used successfully in a wide range of clinical applications. In the case of skin, patients suffering from severe burns and diabetic ulcers have benefited significantly from the ability to replace skin tissue using an engineered substitute, thereby avoiding donor site morbidity associated with a graft from another application. Cartilage tissue engineering leads solutions for huge numbers of patients requiring repair of meniscal tears and other injuries, and therefore, these applications represent the earliest realization of the promise of tissue engineering in the clinic. Together, they represent systems that have moved toward patient care most rapidly, in part because they do not require the use of an engineered microenvironment associated with a vascular supply or other complex microstructures. However, even in these cases, the tools of microscale tissue engineering can be useful in generating 3D structures for multicellular processes (Khademhosseini *et al.* 2006). In one recent example, Albrecht *et al.* (2006) have integrated multiple microscale tools to achieve chondrocyte biosynthesis, potentially leading to tissue repair after osteoarthritic injury. In this instance, photopolymerizable hydrogels were microfabricated in concert with dielectrophoretic forces in order to create 3D aggregates of approximately 20 000 bovine articular chondrocytes. This technology platform represents a potential pathway for investigation of fundamental mechanisms involved in embryonic development, cancer, and tissue regeneration.

In the field of bone tissue engineering, microfabrication technology has been utilized by Mata *et al.* (2002) to create 3D constructs. Based on the pioneering work of Caplan (Nakahara *et al.* 1991), in which the osteogenic potential of periosteal cells has been investigated, Mata and coworkers have attempted to induce bone regeneration using microengineered scaffolds. Multilayer PDMS constructs were fabricated using replica molding, with porous films produced using dual-sided molding techniques. These layers were integrated using microfluidic connectors in order to produce a 3D bioreactor. This approach has been applied toward bone formation as well as to the generation of artificial cornea using corneal epithelial cells. In addition, stem cell bioreactors have been produced using these 3D PDMS scaffolding techniques, in which human bone marrow-derived connective tissue progenitor cells were cultured to promote osteoblast differentiation.

The field of wound healing is driven by clinical needs such as the increasing number of type II diabetics with ulcers of the foot and other locations, the

incidence of full-thickness burns in accident victims, and the need for treating soldiers suffering from burns and other injuries in the battlefield. The worldwide market for wound care products exceeds $10 billion annually, and most therapies focus on management of pain and infection but do not address the fundamental issues associated with nonhealing wounds. Advances in the understanding of wound healing from the biological perspective have been significant, and numerous tissue engineering products have been developed and commercialized in response to the clinical need. A recent trend toward the use of biologically active scaffolds, developed for skin tissue engineering applications, has emerged in the field of wound care as well. Such bioactive therapies entail the delivery of antibiotics, growth factors, and other compounds to the wound site.

In addition to bioactive materials, recent developments are aimed at the application of mechanically active scaffolds and devices toward wound care. One such product utilizing mechanical forces to promote wound healing is the vacuum-assisted closure (VAC) device (Kinetic Concepts Inc.), in which a dressing is applied to the wound site and a vacuum is applied, thereby applying mechanical forces across the wound bed and promoting healing. Pioneering work by Orgill, Ingber, and coworkers (Saxena *et al.* 2004) has led to the development of a model for the mechanism associated with wound healing via the VAC device. Micromechanical forces are believed to promote cell division and amplify the effects of growth factors, and cytoskeletal rearrangements in response to these forces are believed to play an important role as well. Fluid transport of nutrients and waste products through the wound site, driven by the mechanical forces or the vacuum pressure, is also believed to play a major role in the enhancement of wound healing.

New microfluidic concepts for promoting wound healing are emerging, based on the ability to photopattern biomaterials and to construct 3D microfluidic devices therefrom. The work of Stroock and coworkers (Cabodi *et al.* 2005b) has resulted in the demonstration of a novel biopolymer, hydroxyethyl methacrylate, in a microfluidic device that optimizes mass transport in order to promote nutrient and waste exchange and to enable the delivery of drugs *in situ*. Such microfluidic devices combine a porous matrix with a well-defined network of microchannels engineered to provide transport of biomolecules to promote the healing process. In addition to the enhancement of mass transport, the fluid exchange also applies shear stresses to the wound site, a second avenue for accelerating the healing process.

2.15.6.2 Neural Tissue Engineering

Brain development is the subject of intense efforts, and microsystems technologies are providing new avenues for exploration of the mechanisms involved. Chemical gradients, as described earlier, are a critical element in neuronal behavior and in axon guidance. In one pioneering study (Dertinger *et al.* 2002), neurons from the hippocampus were cultured in systems involving a microfluidic mixer controlling the formation of a chemical gradient. As described therein, neuronal development involves the formation of processes at each end followed by the rapid elongation of one end toward axon formation. In the absence of chemical gradients, this process cannot be reproduced and hippocampal neurons do not become polarized. Substrate interactions such as contact with laminin trigger axon formation and polarization, and therefore, the microfluidic gradient generator was used to create a gradient of laminin versus bovine serum albumin to attempt to reproduce the axonal development process *in vitro*. A nearly linear gradient of laminin across a 180-μm channel was generated, first within the bulk of the microfluidic channel and then on the surface of the channel. These gradients were shown to promote axonal specification but not axonal guidance, as had been seen in earlier studies.

Other approaches to neural tissue engineering have utilized nanoscale techniques to control the microenvironment within engineered scaffolds. These matrices, based on the nanostructured self-assembled oligopeptides described earlier (Holmes *et al.* 2000), comprise alternating sequences of amino acids known as sapeptides. Neuronal attachment and differentiation, leading to synapse generation, have been demonstrated in scaffolds constructed with these sapeptide matrices. One of the specific cell types used, PC12 rat cells, are a clonal cell line that differentiates in the presence of neural growth factor. Primary cells from the rat cerebellum and hippocampus were also used in culture. When exposed to nerve growth factor (NGF), the PC12 rat cells seeded in specific sapeptides were shown to exhibit significant neurite outgrowth. Similar results were seen in primary cerebellar and hippocampal cell culture studies. Synapse formation was also dramatically enhanced in these cell culture studies, demonstrating the importance of these nanostructured peptides in promoting neural engineered tissues.

2.15.6.3 Vital Organ Tissue Engineering

Progress in the field of tissue engineering has been greatest in applications such as skin, bone, cartilage, and tubular structures such as blood vessels and urologic tissues (Kojima *et al.* 2003). While these advances have provided lifesaving therapies for newborns with congenital defects, burn victims, and many other patients, the greatest potential of tissue engineering technology is in the replacement of vital organ function. Vital organs such as the liver, kidneys, and lungs contain extraordinarily complex structure and function that arises from precise architecture spanning length scales from submicron features up to several centimeters. Prior to the introduction of microsystems technologies to the development of engineered tissues, such structures could only be produced by the introduction of growth factors, ECM materials, and other biochemical components into cell-seeded scaffolds with no significant microscale organization. For instance, microvascular networks have been produced by the stepwise introduction of VEGF and PDGF into a porous scaffolding to coax the formation of self-assembled capillaries (Richardson *et al.* 2001). Other studies have introduced sphingosine to generate vascular structures (Lee *et al.* 1999). Organ-specific structures such as sinusoids and biliary structures in engineered liver tissue, as well as nerve-like elements, have been demonstrated in response to the use of novel 3D nanostructured ECM materials (Semino *et al.* 2003). However, in each of these cases, the extraordinary complexity of the organ cannot be approached through the use of biochemical signals alone, and therefore, the field of microsystems technology has been instrumental in establishing the mechanical, spatial, and topographic microenvironment that comprises vital organs. In Sections 2.15.6.3.1–2.15.6.3.4, recent advances in engineered liver, kidney, cardiac, and pulmonary tissues will be described in brief.

2.15.6.3.1 Liver

The liver represents one of the most important and difficult challenges in the field of tissue engineering, because of the urgent need for solutions for liver failure beyond cadaveric organ transplantation, and because of the numerous requirements related to the mechanical and biochemical microenvironment for the hepatocytes and other cells. In addition, engineered liver replacement therapies require a means for the delivery of oxygen that reaches down to the 100-μm scale, due to the high OCR of liver tissue and the poor diffusivity of oxygen within liver tissue. Liver tissue engineering applications range from the development of artificial liver devices for drug screening to acellular and cell-based liver assist devices. Within the field of microscale cell and tissue engineering, more emphasis has been placed on the liver than on any other tissue, and a larger number of microscale tools have been developed in investigations involving hepatocytes than any other cell type.

Hepatotoxicity represents one of the principal reasons why candidate drug therapies are withdrawn from the market or never reach the market due to negative results obtained in toxicity and toxicology studies by pharmaceutical companies. The economic costs of failed drug compounds approach 1 billion dollars per event, and an untold number of promising candidates never reach patients who desperately need them due to inaccurate toxicity assessments in early-stage safety studies. Therefore, one of the most significant contributions of the field of tissue engineering, in addition to direct production of replacement tissues and organs, will be the development of laboratory screening tools based on human cells cultured in engineered scaffolds. In these applications, the power of microfabrication technology in replicating the microenvironment for suitable hepatocyte function will be critical.

Numerous drugs have encountered safety issues following FDA approval, including several (Rezulin/Parke–Davis, Duract/Wyeth–Ayerst, and Serzone/Bristol–Meyers) for hepatotoxicity. Other drugs have encountered safety problems with cardiotoxicity, such as the appearance of Q-T prolongation and other types of cardiac or vascular injury. In each case, the toxicity might have been predicted by the existence of early-stage screening tools based on the culture of human cells in microarrays.

Numerous studies have demonstrated the successful culture of primary hepatocytes in microfabricated arrays. Powers *et al.* (2002) have reported the development of microstructured silicon devices in which hepatocytes are cultured; in these studies the 3D nature of the scaffold results in a more physiological behavior of the cells. Kane *et al.* (2006) have developed a microfluidic array for the coculture of primary rat hepatocytes along with 3T3 fibroblasts. Coculture has resulted in improved albumin synthesis, drug detoxification through the cytochrome *P*450 pathway, and urea production. Random coculture of hepatocytes and fibroblasts, as described by

Bhatia *et al.*, does not produce results equivalent to those seen in micropatterned coculture studies. In the work of Kane *et al.*, glass substrates micropatterned with surface adhesion molecules were used as the foundation for a microfluidic array with PDMS wells and microchannels. Albumin and urea synthesis were monitored over periods of up to 5 weeks, and differences between the perfused arrays and control devices were observed and potentially attributed to shear stresses or oxygen distribution. **Figure 16** illustrates the time course of the albumin synthesis in this microengineered bioreactor (Kane *et al.* 2006).

Numerous approaches toward the replacement of liver function have been developed, with early systems based on acellular devices similar to those used for hemodialysis in renal failure patients. However, the complex array of biological functions such as protein synthesis and drug metabolism cannot be replicated without hepatocytes, and therefore, a spectrum of cell-based therapeutic devices have been pursued. In a manner similar to that of renal dialysis, hollow fiber membrane devices loaded with hepatocytes have also been utilized for liver assist, with mixed results. As discussed in Section 2.15.4.1, the inability of mesoscale technologies such as hollow fiber bundles to simulate the microenvironment of the liver sinusoids ultimately limits the effectiveness of such devices in therapeutic applications. A number of modifications to the conventional hollow fiber design have been reported, including the use of microcarriers and multiple coaxial geometries. The membrane itself can be modified to more closely approximate the adhesive and immunological properties of the liver, but long-term hepatocyte function has been limited even in the presence of these advanced configurations. Flat-plate geometries with radial flow have been implemented in an effort to more closely approximate the liver physiology, thereby providing the appropriate spatial microenvironment. Chemical signals and substrate–matrix interactions are equally critical, as are the specific coculture interactions that have been described earlier.

Hepatocyte immobilization on scaffolds and subsequent transplantation of these scaffolds for liver assist or replacement represents a towering technical challenge that will have enormous benefits for patients suffering from end-stage liver failure, hepatitis, and other diseases. Scaffolding for tissue-engineered liver devices can be synthetic or natural; synthetic materials include bioresorbable materials such as PLGA as well as polysaccharides. Biological scaffolds such as collagen sandwiches or hyaluronic acid have also been tested. Hepatocyte encapsulation

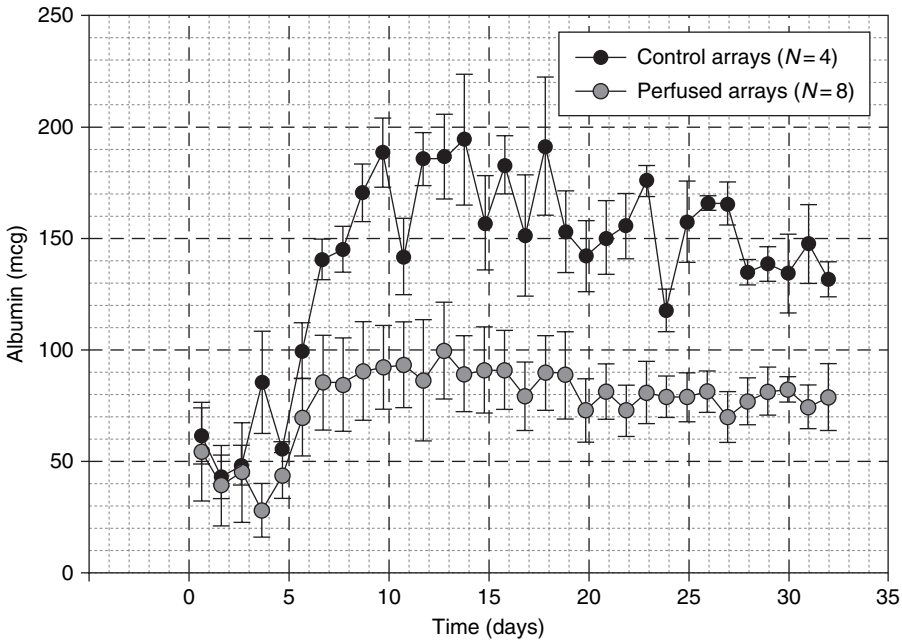


Figure 16 Albumin synthesis in a perfused microfabricated array compared with an open control device, using a coculture of primary rat hepatocytes and 3T3-J2 fibroblasts. (Source: Kane B J, Zinner M J, Yarmush M L, Toner M 2006 Liver-specific functional studies in a microfluidic array of primary mammalian hepatocytes. *Anal. Chem.* **78**, 4291–8.)

in alginate spheres has also been reported, and aggregates of hepatocytes and nonparenchymal cells have been studied as well. The high OCR of hepatocytes requires maintaining small distances between the hepatocytes and the oxygen source. Immune response and long-term viability and functionality of hepatocytes in these devices represent significant challenges, leading to developments such as the microvascular network technologies described earlier. Formation of a viable vasculature to provide oxygen to the highly metabolic liver tissue and development of a biliary network for excretion represent major challenges that need to be addressed in order to develop tissue-engineered liver devices.

2.15.6.3.2 Kidney

The kidney represents the first vital organ for which an artificial device was successfully used to replace function. Renal dialysis still represents the only long-term extracorporeal treatment for organ failure, despite huge efforts aimed at liver and lung augmentation. Kidney dialysis for patients with end-stage renal disease (ESRD) is a life-sustaining therapy but has many drawbacks including poor 5-year survival rates, high cost, severe impact on patient's quality of life, and poor temporal control over blood toxicity levels. The trend in the past decade has been toward slower, gentler dialysis, so that the levels of toxins such as urea and creatinine in the bloodstream do not reach the extreme high and low concentrations of typical hemodialysis. Home dialysis given each night reduces the excursions in toxin levels by replacing three 4-h dialysis treatments with seven 8-h treatments per week; wearable continuous dialysis would even more closely approximate the operation of healthy kidneys for ESRD patients. The goal of microsystems research in microfabricated dialysis devices is to approach the size, capacity, and reliability of a healthy kidney to enable wearable continuous dialysis for the 300 000 ESRD patients in the United States. One group has demonstrated improved clearance of urea and creatinine using a microfabricated device in which multiple layers of PDMS microchannels are stacked with intervening PES nanoporous membranes. In **Figure 17**, a 35-layer device containing a bonded stack of bilayers, each with a semipermeable PES membrane separating a vascular and dialysate chamber, is shown.

ARF represents another opportunity for microsystems technology to improve therapy for patients suffering from kidney disease. ARF can follow serious infections, poisonings and drug overdose, and numerous other causes, and currently the most prevalent therapy for ARF is supportive hospitalization to prevent complications such as sepsis while the kidney function regenerates. New technologies present the possibility for improved therapy for ARF, based on hollow fibers lined with renal proximal tubular cells. Improved glomerular filtration is achieved using silicon nanomachined membranes with slit pores to improve the permselectivity of solutes while maintaining physiological levels of hydraulic permeability.

Advanced membrane technology represents a tremendous opportunity for microfabrication and nanofabrication in the field of renal replacement. Traditional renal dialysis membranes utilized regenerated cellulose as the substrate material, because of its low cost and easy availability. These cellulose membranes were incorporated into various dialyzer configurations, including rotating drums, parallel plates, spirals, and hollow fiber-based systems. Biocompatibility concerns led to the modification of cellulose by vitamin E bonding and heparin bonding, and eventually to the replacement of cellulose by synthetic membranes such as PS and PES. Biocompatibility issues such as thrombogenicity and biofouling are only one aspect in hemodialysis; the efficiency (clearance) of ultrafiltration is another critical aspect of the technology. Diffusive transport governed the filtration process in early dialysis machines, separating small solutes from high-molecular-weight constituents and the solid fraction of the blood and removing them in an efficient manner. Convective transport has emerged as a means to remove middle molecules more efficiently and has accompanied the shift from cellulose to synthetic membranes such as PS and PES. These newer membranes have larger pore sizes in order to

Figure 17 Image of renal filtration device constructed by stacking 35 individual bilayers, each containing a vascular network and dialysate chamber separated by a nanoporous membrane. (Source: Borenstein, Weinberg, Kaazempur-Mofrad, Krebs, and Vacanti, unpublished.)

accommodate the clearance of materials such as $\beta 2$-microglobulin, a middle molecule with a molecular weight of roughly 10 kDa that is responsible for diseases such as amyloidosis in renal failure patients.

One of the most potentially significant advances in the field of organ assist and organ replacement technology based on MEMS fabrication techniques is the silicon nanopore membrane reported by Humes *et al.* (2006). Detailed investigations of the physiology of the glomerulus have shown that a slit-shaped pore with engineered spatial and electrical properties approaches physiological performance more closely than do conventional cylindrical or sponge-like pores. Analysis of the dependence of the hydraulic permeability on the pore geometry indicates that slit-shaped pores may provide improved control over the filtration process. In early studies, Humes' group has collaborated with the Cleveland Clinic to produce slit-shaped silicon nanopores in polysilicon membranes, using a sacrificial silicon dioxide layer that can be controlled to within ± 1 nm. Initial flow characterization of these silicon nanopores matches Hele–Shaw flow through slit-shaped tubes very closely.

2.15.6.3.3 Cardiovascular tissues

The development of cardiovascular tissues such as heart valves (Sodian and Hoerstrup 2000), pulmonary artery grafts (Shinoka *et al.* 1996), and large and small blood vessels represents the largest set of applications for tissue engineering because of the prevalence of heart disease in the United States. As in the case of the liver and other tissues, multiple cell types are required in order to construct replacement cardiovascular tissues, and the need for coculture as well as the requirement for a microvascular supply makes this set of applications an ideal target for microscale tissue engineering. Tissue engineering approaches involving cell-seeded rolled tubes (Shinoka and Ma 1996) have shown extraordinary success for arterial grafts in the clinic in early studies, but the goal of a vascular graft with suitable long-term mechanical and structural properties similar to natural vessels remains a challenge for the field. Topographic cues such as the micropatterning and nanopatterning techniques described earlier have been shown to produce appropriate morphology for the vascular SMCs that comprise the tunica media of blood vessels (Sarkar *et al.* 2005). These topographic cues can be integrated into a biodegradable scaffold (Sarkar *et al.* 2006) to generate the classic herringbone organization of cells present in the smooth muscle layer of native vessels.

Numerous studies have demonstrated the ability to imprint microscale features in biodegradable constructs such as PLGA and PGS (Bettinger *et al.* 2006b). In the blood vessel construct described earlier, PCL was selected because of its biocompatibility, mechanical properties, ease of processing, and the ability to modify the degradation times within a wide range. One drawback of PCL is its poor bulk oxygen and nutrient transport properties. In this instance, PLGA nanoparticles were used to generate an interconnected network of pores in the bulk PCL structure, and this particulate leaching was combined with the formation of microtopography as a cue for orienting the vascular SMCs. This application represents an excellent example of the merger of microfabrication technology and biomaterials modification in order to produce a physiological engineered construct.

2.15.6.3.4 Lung assist devices

The lung is yet another vital organ that is composed of a vast interconnected microfluidic network of channels proximate to an equally complex network of capillaries, in this case placing the blood vessels adjacent to the alveolar structures. Current therapies for pulmonary failure include ECMO, a cardiopulmonary assist device that oxygenates the blood automatically in the absence of normal lung function. Several microsystems-based technologies are currently under development for applications in lung assist and replacement devices. These devices include a microchannel network of blood vessels placed adjacent to a network of oxygenated vessels, with a semipermeable membrane placed in between.

Hollow fiber bioreactors similar to those used in renal and liver tissue engineering applications have been reported for a biohybrid lung by Polk *et al.* These fibers are chemically modified by the addition of adhesion proteins and by plasma processing and then endothelialized to confluence. Gas transfer studies are under way for the development of a mixing oxygenator for long-term patient applications.

The Federspiel group at Pittsburgh (Henchir *et al.* 2003) has reported the development of artificial alveolar capillary (AAC) modules for biohybrid artificial lungs, using a MEMS-based approach with numerous advantages over hollow fiber technology. These devices utilize PDMS microchannel networks generated by micromolding around thin metal wires, to increase the surface-to-volume ratio and enhance oxygen diffusivity. In order to reduce the incidence of thrombosis, these workers are endothelializing the constructs onto fibronectin-coated channels.

Researchers at Northwestern University are also pursuing an artificial lung technology based on microfabrication technology, with the current focus on analytical models for high-efficiency oxygenation. Another group is utilizing photocatalytic generation of oxygen to reduce the need for access to bulky oxygen sources for long-term operation of the artificial lung.

2.15.7 Future Directions

In response to the urgent need for therapeutic solutions for an increasing population of patients suffering from tissue loss and end-stage organ failure, the field of tissue engineering is undergoing explosive growth due to the convergence of advances in cell biology, novel biomaterials, and microscale and nanoscale fabrication techniques. These technologies are achieving success in applications as divergent as skin, bone, and nerve repair and liver, kidney, and lung assist and replacement devices. In each case, the power of the microfabrication technology lies in its ability to replicate the cellular microenvironment and the mechanical, spatial, topographic, and chemical cues that control cell behavior in these complex living systems. Before the introduction of microfabrication technologies to the field, the technological emphasis was principally on the development of biochemical environments in bioreactors that were capable of eliciting biologically based development of structure and function in these systems. While these biologically based approaches are still predominant in many applications, it is becoming clear that complex multiphenotype systems such as the liver and kidney will benefit substantially from the ability to provide microfabricated patterns and structures to drive cellular organization.

Scaffold preparation techniques, ranging from SFF to micromolding large-scale fluidic networks, represent a critical element of microfabricated constructs for tissue engineering. Each of the techniques described herein possesses advantages and disadvantages specific to each tissue engineering application, related to issues such as cost, biocompatibility, processing constraints, and manufacturability. The emergence of micromolding techniques with nanoscale resolution has resulted in rapid advances in vital organ tissue engineering. The use of lasers, ink-jet printers, and other tools developed for other applications has highlighted the vast opportunities for technologies developed for other industries to provide for rapid and precise construction of cell-based systems. Micropatterning of chemical cues on the surfaces of tissue engineering scaffolds has enabled the replication of complex coculture systems in the *in vitro* setting and has provided opportunities to merge these techniques so that cellular and biochemical deposition can be integrated with microfabrication of scaffolds to build complex systems to replace organ function.

The potential for advances in tissue engineering based on microscale and nanoscale systems and technologies can be further realized through the integration of MEMS and nanoelectromechanical systems (NEMS) active elements in the tissue engineering process. These advances are yet to be realized, and therefore, they have not been discussed here. However, control over the flows and concentrations of culture media, oxygen, nutrients, and other biochemical factors in bioreactors for tissue engineering can be enhanced through the use of MEMS sensors and actuators such as microvalves and micropumps, flow sensors, biosensors, and other microscale elements. Integration of these elements into extracorporeal devices seems to be a natural development that will improve control, lower the cost, and increase the reliability of organ assist devices, for instance. Integration of these types of devices into fully implantable systems, and ultimately into fully biodegradable systems, represents a future challenge.

Advances in tissue engineering depend on numerous factors beyond the scaffolding and microenvironment associated with microscale technologies. These include advances in biomaterials for scaffolding substrates as well as matrix components, integration of delivery vehicles into scaffolding, advances in basic cell biology, and cell sourcing. The opportunity for microsystems technology lies in the integration of these various elements, and therefore, an engineered tissue construct represents a classic example of a microsystem in the same manner as miniaturized sensing and actuation systems developed using MEMS technology. Much of the added complexity in these systems lies in the extraordinarily intricate functioning of biological elements such as cells that are far too complex to be modeled using the types of physical laws and analyses applied to most MEMS systems. Therefore, most of the advances here are still at a very early stage, and for the foreseeable future, most systems will be represented by an integration of microfabricated devices and cells, because cell functions will be too difficult to replicate in a fully artificial system. Future developments in tissue engineering are

expected to continue to leverage advances in microfabrication and nanofabrication technology in order to more closely replicate the physiological systems these devices are being developed to replace.

Acknowledgments

The author would like to acknowledge the extraordinary contributions of colleagues and collaborators too numerous to mention. Special thanks go to James Hsiao, Ernest Kim, Kim Bonner, Eleanor Pritchard, Mark Keegan, Chris Bettinger, Eli Weinberg, Ahmad Khalil, and John Aceti of Draper Laboratory for their contributions to the tissue engineering program there. The contributions of collaborators Joseph Vacanti, Cathryn Sundback, Craig Neville, Katherine Kulig, and Nicholas Krebs of Massachusetts General Hospital are also gratefully acknowledged, as are those of Mohammad Kaazempur-Mofrad of the University of California at Berkeley, Robert Langer, Ali Khademhosseini, Kevin King, and Roger Kamm of MIT, and David Kaplan of Tufts University. The support of Draper Laboratory and the Center for Integration of Medicine and Innovative Technology (CIMIT) US Army DAMD17-02-2-0006 in the preparation of this chapter is gratefully acknowledged.

References

Abrams G A, Schaus S S, Goodman S L, Nealey P F, Murphy C J 1997 Nanoscale topography of the basement membrane underlying the corneal epithelium of the non-human primate. *IOVS* **38**, 350

Agrawal C M, Athanasiou K A, Heckman J D 1997 Biodegradable PLA–PGA polymers for tissue engineering in orthopaedics. *Mater. Sci. Forum* **250**, 113–28

Albrecht D R, Underhill G H, Wassermann T B, Sah R L, Bhatia S N 2006 Probing the role of multicellular organization in three-dimensional microenvironments. *Nature Methods* **3**, 369–75

Allen J W, Bhatia S N 2002 Engineering liver therapies for the future. *Tissue Eng.* **8**, 725–37

Allen J W, Bhatia S N 2003 Formation of steady-state oxygen gradients *in-vitro*. *Biotechnol. Bioeng.* **82**, 253–62

Armani D K, Liu C 2000 Microfabrication technology for polycaprolactone, a biodegradable polymer. *J. Microelectromech. Syst.* **10(1)**, 80–4

Athanasiou K A, Schmitz J P, Agrawal C M 1998 The effects of porosity on *in vitro* degradation of polylactic acid–polyglycolic acid implants used in repair of articular cartilage. *Tissue Eng.* **4**, 53–63

Auger F A, Valle C A L, Guignard R, Tremblay N, Noel B, Goulet F, Germain L 1995 Skin equivalent produced with human collagen. *In Vitro Cell Dev. Biol. Anim.* **31**, 432–9

Basu S, Cunningham L P, Pins G D, Bush K A, Howell A R, Campagnola P J 2005 Multi-photon excited fabrication of collagen matrices crosslinked by a modified benzophenone dimer: Bioactivity and enzymatic degradation. *Biomacromolecules* **6**, 1465–74

Bell E, Ehrlich P, Buttle D J, Nakatsuji T 1981 Living tissue formed *in vitro* and accepted as skin-equivalent of full thickness. *Science* **221**, 1052–54

Berthiaume F, Yarmush M L 1995 Tissue engineering. In: Bronzino J D (ed.) *The Biomedical Engineering Handbook.* CRC Press, Boca Raton, FL, pp. 1556–66

Bettinger C J, Orrick B, Langer R, Borenstein J 2006a Tunable contact guidance using feature geometry in biodegradable substrates. *Biomaterials* **27**, 2558–65

Bettinger C J, Weinberg E J, Kulig K M, Vacanti J P, Wang Y D, Borenstein J T, Langer R 2006b Three-dimensional microfluidic tissue engineering scaffolds using a flexible biodegradable polymer. *Adv. Mater.* **18**, 165–9

Bhatia S N, Yarmush M L, Toner M 1997 Controlling cell interactions by micropatterning in co-cultures: Hepatocytes and 3T3 fibroblasts. *J. Biomed. Mater. Res.* **34**, 189–99

Bhatia S N, Yarmush M L, Toner M 1998a Micropatterning cells in tissue engineering. *Methods Mol. Med.* **18**, 349–63

Bhatia S N, Balis U J, Yarmush M L, Toner M 1998b Microfabrication of hepatocyte/fibroblast co-cultures: Role of homotypic cell interactions. *Biotechnol. Prog.* **14**, 378–87

Bhatia S N, Balis U J, Yarmush M L, Toner M 1999 Effect of cell–cell interactions in preservation of cellular phenotype. *FASEB J.* **13**, 1883–900

Borenstein J T, Terai H, King E J, Weinberg M R, Kaazempur-Mofrad J P, Vacanti J P 2002 Microfabrication technology for vascularized tissue engineering. *Biomed. Microdevices* **4**, 167–75

Brunette D M, Kenner G S, Gould T R 1983 Grooved titanium surfaces orient growth and migration of cells from human gingival explants. *J. Dent. Res.* **62**, 1045–8

Cabodi M, Choi N W, Gleghorn J P, Lee C S D, Bonassar L J, Stroock A D 2005a A microfluidic biomaterial. *J. Am. Chem. Soc.* **127**, 13788–9

Cabodi M, Havenstrite K L, Curtis V, Schwartz S, Stroock A D 2005b A microfluidic wound dressing and wound analysis tool. *ASME Summer Bioeng. Conf.,* Vail, CO

Chen C S, Mrksich M, Huang S, Whitesides G M, Ingber D E 1997 Geometric control of cell life and death. *Science* **276**, 1425–8

Chin V, Collins B E, Sailor M J, Bhatia S N 2001 Compatibility of primary hepatocytes with oxidized nanoporous silicon. *Adv. Mater.* **13**, 1877–80

Chu T-M G, Orton D G, Hollister S J, Feinberg S E, Halloran J W 2002 Mechanical and *in vivo* performance of hydroxyapatite implants containing controlled internal architecture. *Biomaterials* **23**, 1283–93

Clark P, Connolly P, Curtis A S G, Dow J A T, Wilkinson C D W 1991 Cell guidance by ultrafine topography *in-vitro*. *J. Cell. Sci.* **99**, 73–7

Cooke M N, Fisher J P, Dean D, Rimnac C, Mikos A G 2003 Use of stereolithography to manufacture critical-sized 3D biodegradable scaffolds for bone ingrowth. *J. Biomed. Mater. Res. B Appl. Biomater.* **64B**, 65–9

Curtis A S, Casey B, Gallagher J O, Pasqui D, Wood M A, Wilkinson C D W 2001 Substratum nanotopography and the adhesion of biological cells. Are symmetry or regularity of nanotopography important? *Biophys. Chem.* **94**, 275–83

De Silva M N, Paulsen J, Renn M J, Odde D J 2005 Two-step cell patterning on planar and complex curved surfaces by precision spraying of polymers. *Biotechnol. Bioeng.* **93**, 919–27

Den Braber E T, de Ruijter J E, Ginsel L A, von Recum A F, Jansen J A 1998 Orientation of ECM protein deposition, fibroblast cytoskeleton, and attachment complex

components on silicon microgrooved surfaces. *J. Biomed. Mater. Res.* **40**, 291–300
Dertinger S K W, Jiang X, Li Z, Murthy V N, Whitesides G M 2002 Gradients of substrate-bound laminin orient axonal specification of neurons. *Proc. Natl. Acad. Sci. U S A* **99**, 12542–7
Desai T A 2000 Micro- and nanoscale structures for tissue engineering constructs. *Med. Eng. Phys.* **22**, 595–606
Dowling R D, Etoch S W, Stevens K A, Johnson A C, Gray Jr L A 2000 Results of preclinical investigations of the Abiomed totally implantable replacement heart. *Circulation* **102(N 18 Supplement)**, II763
Downing B R, Cornwell K G, Toner M, Pins G D 2005 The influence of microtextured basal lamina analog topography on keratinocyte function and epidermal regeneration. *J. Biomed. Mater. Res* **72A**, 47–56
Escaned J, Segovia J, Flores A, Aragoncillo P, Salas C, Alfonso F, Lopez M, Garcia-Touchard A, Fernandez-Ortiz A, Hernandez R, Banuelos C, Sabate M, Alonso-Pulpon L, Macaya C 2001 Assessment of coronary microcirculation in cardiac allografts. A comparison of intracoronary physiology, intravascular ultrasound and histological morphometry. *J. Heart Lung Transplant.* **20**, 204–5
Essig M, Friedlander G 2003 Tubular shear stress and phenotype of renal proximal tubular cells. *J. Am. Soc. Nephrol.* **14**, S33–5
Flemming R G, Murphy C J, Abrams G A, Goodman S L, Nealey P F 1999 Effects of synthetic micro- and nano-structured surfaces on cell behavior. *Biomaterials* **20**, 573–88
Foo C W P, Patwardhan S V, Belton D J, Kitchel B, Anastasiades D, Huang J, Naik R R, Perry C C, Kaplan D L 2006 Novel nanocomposites from spider silk-silica fusion (chimeric) proteins. *Proc. Natl. Acad. Sci. U S A* **103**, 9428–33
Frey W, Woods C K, Chilkoti A 2000 Ultraflat nanosphere lithography: A new method to fabricate flat nanostructures. *Adv. Mater.* **12**, 1515–19
Garcia-Cardena G, Comander J, Anderson K R, Blackman B R, Gimbrone M A Jr. 2001 Biomechanical activation of vascular endothelium as a determinant of its functional phenotype. *Proc. Natl. Acad. Sci. U S A* **98**, 4478–85
Gimbrone M A Jr., Cotran R S, Folkman J 1973 Endothelial regeneration: Studies with human endothelial cells in culture. *Ser. Haematol.* **6**, 453–5
Goodberlet J G, Kavak H 2002 Patterning sub-50 nm features with near-field embedded-amplitude masks. *Appl. Phys. Lett.* **81**, 1315–17
Goodman S L, Sims P A, Albrecht R M 1996 Three-dimensional extracellular matrix textured biomaterials. *Biomaterials* **17**, 2087–95
Griffith L G, Swartz M A 2006 Capturing complex 3D tissue physiology *in vitro*. *Nat. Rev. Mol. Cell Biol.* **7**, 211–23
Gu W, Zhu X, Futai N, Cho B S, Takayama S 2004 Computerized microfluidic cell culture using elastomeric channels and Braille displays. *Proc. Natl. Acad. Sci. U S A* **101**, 15861–6
Guyton A C, Hall J E 2000 *Textbook of Medical Physiology* WB Saunders, Philadelphia, PA
Hartgerink J D, Beniash E, Stupp S I 2001 Self-assembly and mineralization of peptide-amphiphile nanofibers. *Science* **294**, 1684–8
Henchir K A, Lin Q, Wagner W R, Federspiel W J 2003 Microfabrication of biohybrid artificial alveolar capillary modules. *Int. J. Cardiovasc. Med. Sci.* **6**, 123
Holmes T C, de Lacalle S, Su X, Liu G, Rich A, Zhang S 2000 Extensive neurite outgrowth and active synapse formation on self-assembling peptide scaffolds. *Proc. Natl. Acad. Sci. USA* **97**, 6728–33
Humes H D 1996 Tissue engineering of a bioartificial kidney: A universal donor organ. *Transplant. Proc.* **28**, 2032–5
Humes H D, Fissell W H, Weitzel W F 2006 The future of hemodialysis membranes. *Kidney Int.* **69**, 1115–19
Hutmacher D W 2001 Scaffold design and fabrication technologies for engineering tissues – State of the art and future perspectives. *J. Biomater. Sci. Polym. Ed.* **12**, 107–24
Hutmacher D W, Sittinger M, Risbud M V 2004 Scaffold-based tissue engineering: Rationale for computer-aided design and solid free-form fabrication systems. *Trends Biotechnol.* **22**, 355–62
Ingber D E 1993 Cellular tensegrity: Defining new rules of biological design that govern the cytoskeleton. *J. Cell Sci.* **104**, 613–27
Ingber D E 2003 Mechanobiology and diseases of mechanotransduction. *Ann. Med.* **35**, 564–77
Jaffe E A, Nachman R L, Becker C G, Minick C R 1973 Culture of human endothelial cells derived from umbilical veins. Identification by morphologic and immunologic criteria. *J. Clin. Invest.* **52**, 2754–6
Jo B-H, Beebe D J 1999 Fabrication of three-dimensional microfluidic systems by stacking molded PDMS layers. *SPIE* **3877**, 222
Kaazempur-Mofrad M R, Vacanti J P, Kamm R D 2001 Computational modeling of blood flow and rheology in fractal microvascular networks. *Comput. Fluid Solid Mech.* **2**, 864–7
Kaazempur-Mofrad M R, Krebs N J, Vacanti J P, Borenstein J T 2004 A MEMS-based renal replacement system. *Proc. 2004 Sensors Actuators Conf.,* Hilton Head, SC, USA, pp. 67–70
Kaihara S, Borenstein J T, Koka R, Lalan S, Ochoa E R, Ravens M, Pien H, Cunningham B, Vacanti J P 2000 Silicon micromachining to tissue engineer branched vascular channels for liver fabrication. *Tissue Eng.* **6(2)**, 105–17
Kane R S, Takayama S, Ostuni E, Ingber D E, Whitesides G M 1999 Patterning proteins and cells using soft lithography. *Biomaterials* **20**, 2363–76
Kane B J, Zinner M J, Yarmush M L, Toner M 2006 Liver-specific functional studies in a microfluidic array of primary mammalian hepatocytes. *Anal. Chem.* **78**, 4291–8
Kaplan D L, Moon R T, Vunjak-Novakovic G 2005 It takes a village to grow a tissue. *Nat. Biotechnol.* **23**, 1237–9
Kassab G S, Fung Y C 1994 Topology and dimensions of pig coronary capillary network. *Am. J. Physiol. Heart Circ. Physiol.* **267**, H319–25
Kassab G S, Fung Y C 1995 The pattern of coronary arteriolar bifurcations and the uniform shear hypothesis. *Ann. Biomed. Eng.* **23**, 13–20
Kassab G S, Rider C A, Tang N J, Fung Y C 1993 Morphometry of pig coronary arterial trees. *Am. J. Physiol.* **265**, H350–65
Keegan M E, Whittum-Hudson J A, Saltzman M W 2003 Biomimetic design in microparticulate vaccines. *Biomaterials* **24**, 4435–43
Keegan M E, Falcone J L, Leung T C, Saltzman W M 2004 Biodegradable microspheres with enhanced capacity for covalently bound surface ligands. *Macromolecules* **37**, 9779–84
Khademhosseini A, Yeh J, Eng G, Karp J M, Kaji H, Borenstein J, Langer R 2005 Cell docking inside microwells within reversibly sealed microfluidic channels for multiphenotype cellular arrays. *Lab Chip* **5**, 1380–6
Khademhosseini A, Borenstein J T, Langer R, Borenstein J, Vacanti J 2006 Microscale approaches for tissue engineering and biology. *Proc. Natl. Acad. Sci. U S A* **103**, 2480–7
Kim S S, Utsonomiya H, Koski J A, Wu B M, Cima M J, Sohn J, Mukai K, Griffith L G, Vacanti J P 1998 Survival and function of hepatocytes on a novel three-dimensional synthetic biodegradable polymer scaffold with an intrinsic network of channels. *Ann. Surg.* **228**, 8–13
King K R, Wang C C, Shin M, Vacanti J P, Borenstein J T 2002 Biodegradable polymer microfluidics for tissue engineering microvasculature. *MRS Press* **729**, U1.3

King K R, Wang C J, Kaazempur-Mofrad M R, Vacanti J P, Borenstein J T 2004 Biodegradable microfluidics. *Adv. Mater.* **16**, 2007–12

Kisiday J, Jin M, Kurz B, Hung H, Semino C, Zhang S, Grodzinsky A J 2002 Self-assembling peptide hydrogel fosters chondrocyte extracellular matrix production and cell division: Implications for cartilage tissue repair. *Proc. Natl. Acad. Sci. U S A* **99**, 9996–10001

Klien T J, Schumacher B L 2003 Tissue engineering of stratified articular cartilage from chondrocyte subpopulations. *Osteoarthritis Cartilage* **11(8)**, 595–602

Kojima K, Bonassar L J, Roy A K, Mizuno H, Cortiella J, Vacanti C A 2003 A composite tissue-engineered trachea using sheep nasal chondrocyte and epithelial cells. *FASEB J.* **17**, 823–8

Kolff W, Berk H et al 1944 The artificial kidney: A dialyzer with a great area. *Acta Med. Scand.* **117**, 121–34

Krenz G S, Dawson C A 2002 Vessel distensibility and flow distribution in vascular trees. *J. Math. Biol.* **44**, 360–74

Langer R, Vacanti J 1993 Tissue engineering. *Science* **260**, 920–6

LaVan D A, George P M, Langer R 2003 Simple, three-dimensional microfabrication of electrodeposited structures. *Angew. Chem. Int. Ed. Engl.* **42**, 1262–5

Leclerc E, Sakai Y, Fujii T 2003 A multi-layer PDMS microfluidic device for tissue engineering applications. *Proc. 2003 Transducers Conf.,* Kyoto, Japan, pp. 415–18

Lee M J, Thangada S, Claffey K P, Ancellin N, Liu C H, Kluk M, Volpi M, Sha'afi R I, Hla T 1999 Vascular endothelial cell adherens junction assembly and morphogenesis induced by sphingosine 1-phosphate. *Cell* **99**, 301–12

Leong K F, Cheah C M, Chua C K 2003 Solid freeform fabrication of three-dimensional scaffolds for engineering replacement tissues and organs. *Biomaterials* **24**, 2363–78

Levenberg S, Rouwkema J, Macdonald M, Garfein E S, Kohane D S, Darland D C, Marini R, van Blitterswijk C A, Mulligan R C, D'Amore P A, Langer R 2005 Engineering vascularized skeletal muscle tissue. *Nat. Biotechnol.* **23**, 879–84

Li M, Glawe J D, Green H, Mills D K, McShane M J, Gale B K 2000 Effect of high-aspect-ratio microstructures on cell growth and attachment. *Proc.1st Annu. Int. IEEE-EMBS Spec. Top. Conf. Microtechnologies in Medicine and Biology*, Lyon, France, Vol. 66, pp. 531–6, Poster

Lim D, Kamotani Y, Cho B, Mazumder J, Takayama S 2003 Fabrication of microfluidic mixers and artificial vasculatures using a high-brightness diode-pumped Nd:YAG laser direct write method. *Lab Chip* **3**, 318–23

Liu V A, Bhatia S N 2002 Three-dimensional photopatterning of hydrogels containing living cells. *Biomed. Microdevices* **4**, 257–66

Liu W, Chen C S 2005 Engineering biomaterials to control cell function. *Mater. Today* **8**, 28–35

Lund L W, Hattler B G, Federspiel W J 2002 A comparative *in vitro* hemolysis study of a pulsating intravenous artificial lung. *ASAIO J.* **48**, 631–5

Lysaght M J, Reyes J 2001 The growth of tissue engineering. *Tissue Eng.* **7**, 485–93

Malek A M, Izumo S 1996 Mechanism of endothelial cell shape change and cytoskeletal remodeling in response to fluid shear stress. *J. Cell Sci.* **109(Part 4)**, 713–26

Mata A, Boehm C, Fleischman A, Muschler G, Roy S 2002 Growth of connective tissue progenitor cells on microtextured polydimethylsiloxane surfaces. *J. Biomed. Mater. Res.* **62**, 499–506

McDevitt T C, Woodhouse K A, Hauschka S D, Murry C E, Stayton P S 2003 Spatially organized layers of cardiomyocytes on biodegradable polyurethane films for myocardial repair. *J. Biomed. Mater. Res* **66A**, 586–95

McKnight N L, Frangos J A 2003 Strain rate mechanotransduction in aligned human vascular smooth muscle cells. *Ann. Biomed. Eng.* **31**, 239–49

Mescher M J, Dube C E, Varghese M, Fiering J O 2003. Surface mount microfluidic flow regulator on a polymer substrate. MicroTAS 2003, Squaw Valley, CA

Mironov V, Boland T, Trusk T, Forgacs G, Markwald R R 2003 Organ printing: Computer-aided jet-based 3D tissue engineering. *Trends Biotechnol.* **21**, 157–61

Mooney D, Hansen L 1992 Switching from differentiation to growth in hepatocytes – Control by extracellular matrix. *J. Cell. Physiol.* **151(3)**, 497–505

Mooney D J, Organ G, Vacanti J P, Langer R 1994 Design and fabrication of biodegradable polymer devices to engineer tubular tissues. *Cell Transplant.* **3**, 203–10

Nakahara H, Dennis J E, Buder S P, Haynesworth S E, Lennon D P, Caplan A I 1991 *In vitro* differentiation of bone and hypertrophic cartilage from periosteal-derived cells. *Exp. Cell. Res.* **195**, 492–503

Nelson C M, Chen C S 2003 Engineering cell adhesion for biotechnology applications. *J. Cell. Sci.* **116**, 3571–81

Niklason L E, Gao J, Abbott W M, Hirschi K, Houser S, Marini R, Langer R 1999 Functional arteries grown *in vitro*. *Science* **284**, 489–93

Nuttelman C R, Henry S M, Anseth K S 2002 Synthesis and characterization of photocrosslinkable, degradable poly(vinyl alcohol)-based tissue engineering scaffolds. *Biomaterials* **23**, 3617–26

Palsson B O, Bhatia S N 2004 *Tissue Engineering*. Pearson Prentice Hall, Upper Saddle River, NJ, Chapter 1

Parenteau N, Sabolinski N, Prosky S, Nolte C, Oleson M, Kriwet K, Bilbo P 1996 Biological and physical factors influencing the successful engraftment of a cultured human skin substitute. *Biotechnol. Bioeng.* **52**, 3–14

Patzer J F II 2004 Oxygen consumption in a hollow fiber bioartificial liver – Revisited. *Artif. Organs* **28**, 83–98

Peters M C, Polverini P J, Mooney D J 2002 Engineering vascular networks in porous polymer matrices. *J. Biomed. Mater. Res* **60**, 668–78

Petka W A, Harden J L, McGrath K P, Wirtz D, Tirrell D A 1998 Reversible hydrogels from self-assembling artificial proteins. *Science* **281**, 389–92

Powers M J, Janigian D M, Wack K E, Baker C S, Stolz D B, Griffith L G 2002 Functional behavior of primary rat liver cells in a three-dimensional perfused microarray bioreactor. *Tissue Eng.* **8**, 499–513

Poznansky M C, Evans R H, Foxall R B, Olszak I T, Piascik A H, Hartman K E, Brander C, Meyer T H, Pykett M J, Chabner K T, Kalams S A, Rosenzweig M, Scadden D T 2000 Efficient generation of human T cells from a tissue-engineered thymic organoid. *Nat. Biotechnol.* **18**, 729–34

Purdue G F, Hunt J L, Still J M Jr., Law E J, Hernson D N, Goldfarb I W, Schiller W R, Hansbrough J F, Hickerson W K, Himel H N, Kealey G P, Twomey J, Missavage A E, Solem L D, Davis M, Totoritis M, Gentzkow G D 1997 A multicenter clinical trial of a biosynthetic skin replacement, dermagraft-TC, compared with cryopreserved human cadaver skin for temporary coverage of excised burn wounds. *J. Burn Care Rehabil.* **18**, 52–7

Quake S R, Scherer A 2000 From micro- to nanofabrication with soft materials. *Science* **290**, 1536–40

Raghavan S, Chen C S 2004 Micropatterned environments in cell biology. *Adv. Mater.* **16**, 1303–13

Richardson T P, Peters M C, Ennett A B, Mooney D J, 2001 Polymeric system for dual growth factor delivery. *Nature Biotech.* **19**, 1029–34

Rosenberg M D 1962 Long-range interactions between cell and substratum. *Proc. Natl. Acad. Sci. U S A* **48**, 1342–9

Roy P, Baskaran H, Roy P, Baskaran H, Tilles A W, Yarmush M L, Toner M 2001 Analysis of oxygen transport to hepatocytes in a flat-plate microchannel bioreactor. *Ann. Biomed. Eng.* **29**, 947–55

Ryu W, Min S, Fasching R J, Prinz F B 2006 3D multi-layered micro-fabricated tissue scaffolds of biodegradable polymers. *Soc. Biomater. Conf.,* Pittsburgh, PA

Sarkar S, Dadhania M, Rourke P, Desai T A, Wong J Y 2005 Vascular tissue engineering: Microtextured scaffold templates to control organization of vascular smooth muscle cells and extracellular matrix. *Acta Biomater.* **1**, 93–100

Sarkar S, Lee G Y, Wong J Y, Desai T A 2006 Development and characterization of a porous micro-patterned scaffold for vascular tissue engineering applications. *Biomaterials* **27**, 4775–82

Saxena V, Hwang C W, Huang S, Eichbaum Q, Ingber D, Orgill D P 2004 Vacuum assisted closure: Micro-deformations of wounds and cell proliferation. *Plast. Reconstruct. Surg.* **114**, 1086–96

Seebach J D, Schneider M K, Comrack C A, LeGuern A, Kolb S A, Knolle P A, Germana S, DerSimonian H, LeGuern C, Sachs D H 2001 Immortalized bone-marrow derived pig endothelial cells. *Xenotransplantation* **8**, 48–51

Semino C E, Merok J R, Crane G G, Panagiotakos G, Zhang S 2003 Functional differentiation of hepatocyte-like spheroid structures from putative liver progenitor cells in three-dimensional peptide scaffolds. *Differentiation* **71**, 262–70

Shen J, Luscinskas F W, Connolly A, Dewey C F Jr., Gimbrone M A Jr. 1992 Fluid shear stress modulates cytosolic free calcium in vascular endothelial cells. *Am. J. Physiol.* **262**, C384–90

Shinoka T, Ma P X 1996 Tissue engineered heart valves. Autologous valve leaflet replacement study in a lamb model. *Circulation* **94(Suppl II)**, II164–8

Sodian R, Hoerstrup S P 2000 Tissue engineering of heart valves: *In vitro* experiences. *Ann. Thorac. Surg.* **70**, 140–4

Sodian R, Sperling J S, Martin D P, Stock U, Mayer J E Jr., Vacanti J P 2002 Application of stereolithography for scaffold fabrication for tissue engineered heart valves. *ASAIO J.* **48**, 12–16

Tan J L, Tien J, Pirone D, Gray D S, Chen C S 2003 Cells lying on a bed of microneedles: An approach to isolate mechanical force. *Proc. Natl. Acad. Sci. U S A* **100**, 1484–9

Tsang V L, Bhatia S N 2004 Three-dimensional tissue fabrication. *Adv. Drug Del. Rev.* **56**, 1635–47

Vacanti J P, Langer R 1999 Tissue engineering: The design and fabrication of living replacement devices for surgical reconstruction and transplantation. *Lancet* **354(Suppl I)**, 32SI–4SI

Van Kooten T G, Whitesides J F, von Recum A F 1998 Influence of silicone (PDMS) surface texture on human skin fibroblast proliferation as determined by cell cycle analysis. *J. Biomed. Mater. Res.* **43**, 1–14

Voldman J, Gray M L, Toner M, Schmidt M A 2002 A microfabrication-based dynamic array cytometer. *Anal. Chem.* **74**, 3984–90

Vozzi G, Flaim C, Ahluwalia A, Bhatia S 2003 Fabrication of PLGA scaffolds using soft lithography and microsyringe deposition. *Biomaterials* **24**, 2533–40

Wang Y, Guillermo A, Sheppard B J, Langer R 2002 A tough biodegradable elastomer. *Nat. Biotechnol.* **20**, 602–6

Wang G J, Chen C L, Hsu S H, Chiang Y L 2005 Bio-MEMS fabricated artificial capillaries for tissue engineering. *Microsyst. Technol.* **12**, 120–7

Warren S M, Fong K D, Chen C M, Loboa E G, Cowan C M, Lorenz H P, Longaker M T 2003 Tools and techniques for craniofacial tissue engineering. *Tissue Eng.* **9**, 187–200

Weaver V M, Petersen O W, Wang F, Larabell C A, Briand P, Damsky C, Bissell M J 1997 Reversion of the malignant phenotype of human breast cells in three-dimensional culture and *in vivo* by integrin blocking antibodies. *J. Cell. Biol.* **137**, 231–245

Weinberg E J, Borenstein J T, Kaazempur-Mofrad M R, Orrick B, Vacanti J P 2004 Design and fabrication of a constant shear microfluidic network for tissue engineering. *MRS Symp. Proc.* MRS Press, Warrendale, PA, Vol. 820, pp. 121–6

Weiss P 1945 Experiments on cell and axon orientation *in vitro*: The role of colloidal exudates in tissue organization. *J. Exp. Zool.* **100**, 353–86

Whitesides G M, Stroock A D 2001 Flexible methods for microfluidics. *Phys. Today* **54**, 42–8

Wilson C J, Clegg R E, Leavesley D I, Pearcy M J 2005 Mediation of biomaterial–cell interactions by adsorbed proteins: A review. *Tissue Eng.* **11**, 1–18

Xu T, Jin J, Gregory C, Hickman J J, Boland T 2005 Inkjet printing of viable mammalian cells. *Biomaterials* **26**, 93–9

Yang S, Leong K F, Du Z, Chua C K 2001 The design of scaffolds for use in tissue engineering. Part I. Traditional factors. *Tissue Eng.* **7**, 679–89

Yang S, Leong K-F, Du Z H, Chua C K 2002 The design of scaffolds for use in tissue engineering. Part II. Rapid prototyping techniques. *Tissue Eng.* **8**, 1–11

Young C S, Terada S, Vacanti J P, Honda M, Bartlett J D, Yelick P C 2002 Tissue engineering of complex tooth structures on biodegradable polymer scaffolds. *J. Dent. Res.* **81**, 695–700

Zein I, Hutmacher D W, Tan K C, Teoh S H 2002 Fused deposition modeling of novel scaffold architectures for tissue engineering applications. *Biomaterials* **23**, 1169–85

Zhang S 2002 Emerging biological materials through molecular self-assembly. *Biotechnol. Adv.* 321–39

Zhang S, Holmes T, Lockshin C, Rich A 1993 Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proc. Natl. Acad. Sci. U S A* **90**, 3334–8

Ziegler T, Nerem R M 1994 Tissue engineering a blood vessel: Regulation of vascular biology by mechanical stresses. *J. Cell. Biochem.* **56**, 204–9

Biography

Jeffrey Borenstein is currently Director of the Biomedical Engineering Center and is a Distinguished Member of the Technical Staff at the Charles Stark Draper Laboratory in Cambridge, Massachusetts. There he serves as the Technical Director for Draper's new biomedical engineering program, with projects in drug delivery, clinical diagnostics, regenerative medicine and biodefense. He is also a Program Leader and the Draper Laboratory Liason for the Center for the Integration of Medicine and Innovative Technology (CIMIT). Dr. Borenstein is Principal Investigator for several commercial and government programs in tissue engineering as well as an NIH Bioengineering Research Partnership in drug delivery. Dr. Borenstein has a Ph.D. in Physics from the University at Albany and 20 years of experience in microsystems technology and biomedical devices. Dr. Borenstein's current work is focused on the application of microsystems technologies towards organ assist and organ replacement for transplant medicine and drug discovery, diagnostics and sensing applications, and drug delivery. He has eight issued patents, as well as nine published patent applications, and drug delivery. He has eight issued patents, as well as nine published patent applications and over 80 peer-reviewed journal articles and conference proceedings.

3.01 Micro-Mirrors

K. Hane, Department of Nanomechanics, Tohoku University, Sendai, Japan

M. Sasaki, Department of Advanced Science and Technology, Toyota Technological Institute, Nagoya, Japan

Glossary

AVC Angular Vertical Comb drive
CTP Computer-To-Plate
ECL External Cavity Diode Laser
OCT Optical Coherence Tomography
OXC Optical Cross Connect
PSD Position-Sensitive Detector
SOI Silicon-On-Insulator
SVC Staggered Vertical Comb drive
TIR Total Internal Reflection
VOA Variable Optical Attenuator
WDM Wavelength-Division-Multiplexing

3.01.1 Introduction

The definition of the micromirror may be the small-sized device including some new mechanisms as compared to bulk mirror devices, which usually use bulk motors for actuation. The micromirror requires a smaller amount of power for driving, and some integration in numbers or functions can be combined. At present (2006), Digital Micromirror Devices (DMD), a product from Texas Instruments, is a typical micromirror because of mass production and commercial success. Of course, there are many other variations and potential uses of micromirrors. The continuing growth in optical system applications (e.g., display technologies or optical-fiber telecommunication infrastructure) has stimulated the progress of miniaturized, reliable, inexpensive photonic devices for light beam manipulation. In integrated optics, mechanical moving elements are avoided, as they cause unreliability. Electro- or magneto-optic effects are preferred. Such devices, however, generally suffer from low contrast and usually operate at low efficiencies. Movable mechanical elements such as mirrors can realize optical beam manipulation more effectively. The devices can be manufactured inexpensively by using the microelectromechanical systems (MEMS) fabrication technique. Microfabrication offers high-volume capability and therefore low-cost products. In general applications, not only cost reduction but also power reduction or portability and shrinkage in size will promote the introduction of micromirrors when compared to present designs. The foreseen goal will be to integrate the micromirrors and electrical circuits for their controls.

3.01.2 History at the Beginning

Historically, significant improvements have been carried out introducing new ideas and technologies. The first published report of a micromirror is the mirror matrix tube for mechanical light modulation developed at Westinghouse Research Labs in 1975 (Guldberg *et al.* 1975, Thomas *et al.* 1975). The mirror array is electrostatically deflected to the substrate when charged by an electron beam. With the saturated irradiation of the electron beam with the surrounding bias voltage, the mirror gets a bias voltage. The micromirror consists of an SiO_2 membrane grown on epitaxial Si on sapphire on which Al electrodes are patterned. SiO_2 layers are released by undercutting of Si forming the deflectable mirror surfaces. **Figure 1** shows the image of the mirror having the four-leaf clover shape and its cross section.

When the bias voltage is applied between the mirror and the underlying electrodes, the segmented leafs of mirror deflect. The reflected light beam generates four spots corresponding to each mirror leaf area. The rectangular mirror generates the cross-type reflection pattern (described in Section 3.01.3.1.1). **Figure 2(a)** shows the diffraction pattern illuminated with a laser beam. When the cross-type light shield is inserted (illustrated by white lines in **Figure 2(a)**), the reflection beam can be on/off modulated. **Figure 2(b)** shows the reflected white light beams that have passed through the cross-type shield. **Figure 3** shows the demonstration of the devices as the projection display for airport information.

In 1977, Petersen (1977) at the IBM research laboratory reported SiO_2 cantilevers for light modulators fabricated by anisotropic Si etching. An individual cantilever can deflect attracted by the electrostatic force. A 16-element light modulator array was used for parallel light processing. A beam of He–Ne laser light was reflected by the cantilever, apertured, focused, and scanned by a galvanometer. In 1980, Petersen presented the scanning mirror with the torsion bar using single-crystalline silicon (c-Si). **Figure 4(a)** shows the mirror image. This shape of the micromirror is fundamentally the same as that of typical micromirrors in use currently. For a 2 mm $\times$ 2 mm size mirror, a scan angle of $\pm 1^\circ$ was

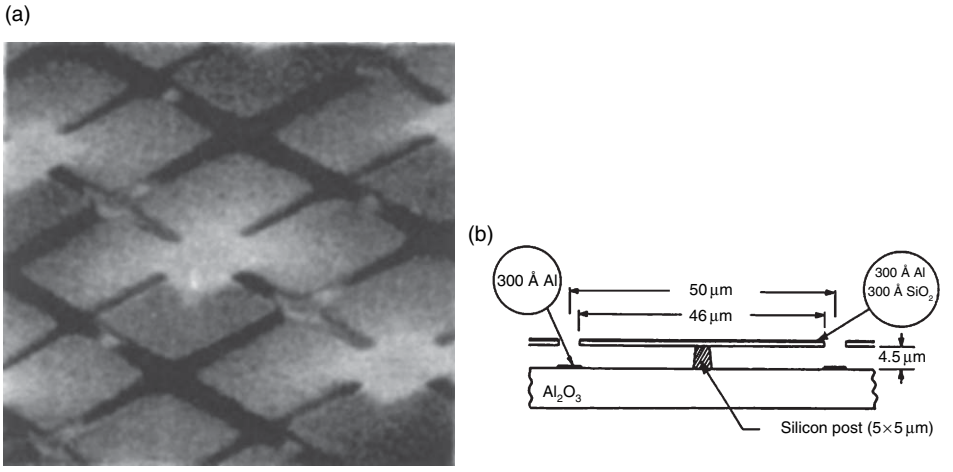


Figure 1 (a) Photographs of light valve elements and (b) cross-sectional view of a single element. (Source: Thomas R N, Guldberg J, Nathanson H C, Malmberg P R 1975 The mirror-matrix tube: A nonel light valve for projection displays. *IEEE Trans. On Electron Devices* **ED-22(9)**, 765–75.)

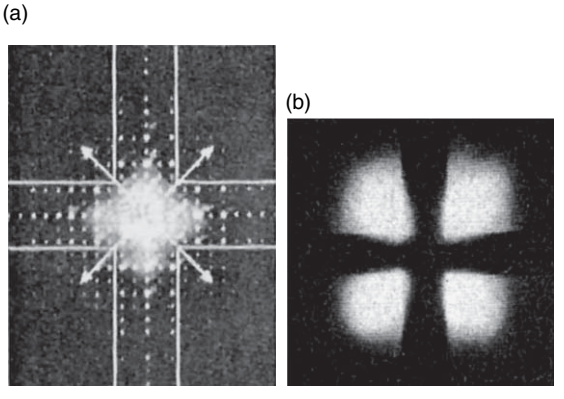


Figure 2 The diffraction pattern from the individual element illuminated by (a) laser and (b) white light. (Source: Thomas R N, Guldberg J, Nathanson H C, Malmberg P R 1975 The mirror-matrix tube: A nonel light valve for projection displays. *IEEE Trans. On Electron Devices* **ED-22(9)**, 765–75.)

Figure 3 Photographs of projection display produced with the mirror matrix light valve. (Source: Thomas R N, Guldberg J, Nathanson H C, Malmberg P R 1975 The mirror-matrix tube: A nonel light valve for projection displays. *IEEE Trans. On Electron Devices* **ED-22(9)**, 765–75.)

demonstrated at the resonance frequency of 15 kHz. Figure 4(b) shows the cross section of the device. A glass substrate was etched to form a well into which the torsion plate could deflect. A support ridge is used to prevent it from being pulled down in parallel to the base of the well. A pair of driving electrodes was fabricated in the well. This idea is used in the recent micromirror device design (Yamamoto *et al.* 2003). Continuous operation of the devices for months ($\sim 10^{11}$ cycles) followed by dislocation-sensitive Si etching showed no damage exhibiting high fatigue strengths and high reliability (Petersen 1982). This highlights one of the important advantages of using single-crystal materials and of scaling down the device dimensions: fatigue mechanisms are not the same as those found in macroscopic materials due to the absence of grain boundaries.

3.01.3 Diffraction Characteristics of Micromirrors

3.01.3.1 Diffraction Image Obtained through Lens

The micromirror is usually used for making a spot far from the device. The spot can be treated as the (Fraunhofer) diffraction pattern of the micromirror

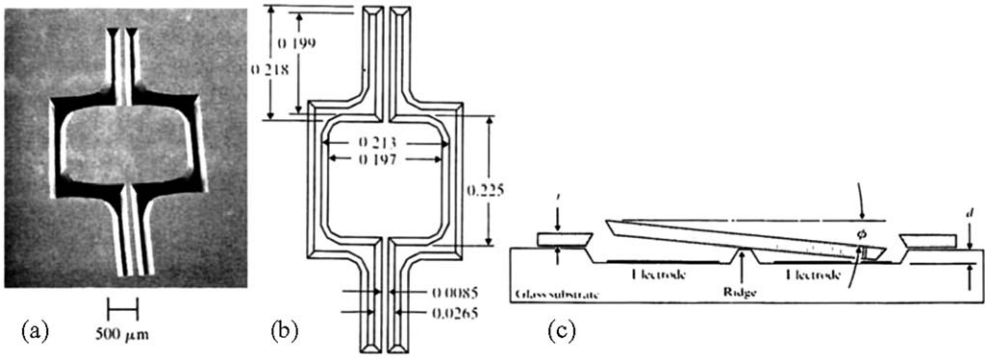


Figure 4 (a) Scanning electron microscope (SEM) image of typical torsion mirror. (b) Dimensions of 15-kHz resonating mirror (cm). (c) Cross section of mirror element. The depth of well is 12.5 μm. (Source: Petersen K E 1980 Silicon torsional scanning mirror. *IBM J. Res. Develop.* **24(5)**, 631–7.)

(Hecht 2002). **Figure 5** shows the optical setup. Q_0 is the point light source on the optical axis and Q is its image. The wavelength of the light emitted from Q_0 is λ. The folded orbit of the light beam reflecting at the mirror is straightened in this model. The reflection mirror is replaced with the screen aperture. The light, which has passed outside of the mirror, does not contribute to generating the spot image. The reflection mirror has the same function in the aperture. The following examples show the images of the micromirrors using the projection lens. When the imaging through the lens is ideal without aberration, the wave, which has passed through the aperture, will become a part of the spherical wave converging to its image. The emission pupil of the lens is considered as an aperture. The spherical surface Σ is defined to have the rim at the pupil. The center of Σ is point Q, and its radius f is the propagation distance from the aperture to the image. Huygen's principle states that the wave amplitude $u(x', y')$ on the image plane including point Q can be expressed as follows:

$$u(x', y') = C \iint \sum \exp\left\{ i \frac{2\pi}{\lambda} \frac{(\xi' x' + \eta' y')}{f} \right\} d\xi'\, d\varsigma' \quad [1]$$

where (x', y') is the coordinate near the image point Q. (ξ', ζ') is the coordinate for the plane of the emission pupil. Here, it is assumed that $\sqrt{\xi'^2 + \eta'^2} << f$. This assumption can be expressed as $(a/f)^2 << 1$, where a is the size of the emission pupil or the mirror. In the following case, the integral over the spherical surface Σ is replaced with the integral over the planar surface. This replacement is justified if observation is at a distant position.

3.01.3.1.1 Rectangular mirror

Consider a rectangular mirror having the side lengths of $2a$ and $2b$ along ξ' and ζ' directions, respectively. **Figure 6** shows the optical setup. From eqn [1], the following result is obtained:

$$I(x', y') = |u(x', y')|^2 = |C|^2 (2a \cdot 2b)^2 \left[\frac{(\sin(2\pi a/\lambda f))x'}{(2\pi a/\lambda f)x'} \right]^2 \left[\frac{(\sin(2\pi b/\lambda f))y'}{(2\pi b/\lambda f)y'} \right] \quad [2a]$$

At the center position $(x', y') = (0, 0)$, the intensity $I(x', y')$ becomes the maximum value $|C|^2 (2a \cdot 2b)^2$.

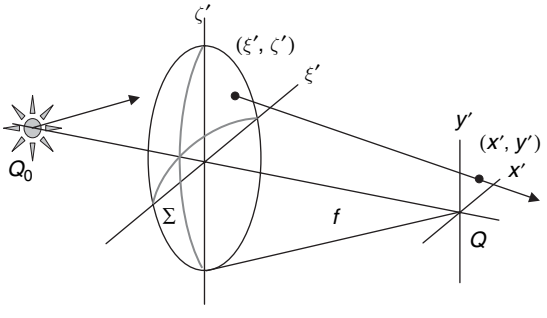


Figure 5 Optical setup for considering Fraunhofer diffraction through lens.

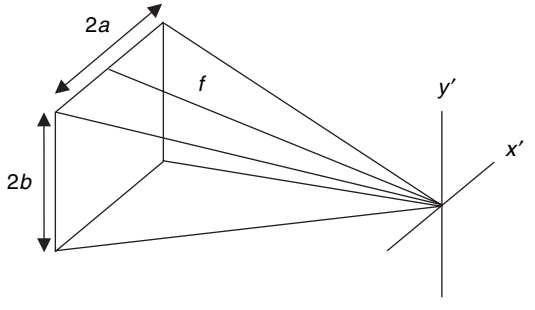


Figure 6 Diffraction from a rectangular mirror.

There is a proportional relation between the intensity and the square of the area. There are positions where the intensity $I(x', y')$ becomes minimum reaching a value 0 where dark nodes are generated. The position of dark nodes is decided by the numerator sine function in eqn [2a]. x' and y' will satisfy the following equation:

$$\frac{2\pi a}{\lambda f}x' = m\pi, \quad \frac{2\pi a}{\lambda f}y' = n\pi \quad (m, n = \pm 1, \pm 2, \pm 3, \ldots) \qquad [2b]$$

Figure 7(a) shows the intensity distribution obtained from eqn [2b]. **Figure 7(b)** shows the magnified image around the center. Beside the main center peak there are subpeaks. The magnitudes of the peak intensity are 4.7% (second peak), 1.7% (third peak), and 0.8% (fourth peak) of that of the center peak. **Figure 8** shows the calculated density plot of the diffraction pattern. The dark lines of the node are observed following eqn [2b]. The diffraction spots generate a ± shape pattern like the constellation. This is the reason why a cross-type shield is used in the mirror matrix tube (see **Figure 2**). When the four segmented leafs deflect, the center of the deflection pattern splits and shifts to the four corners in **Figure 2(b)**. The light passes through the shield realizing the light switching.

Consider the circular mirror shape having a radius a. **Figure 9** shows the optical setup. The coordinates (ξ', ζ') and (x', y') can be replaced by the polar coordinates. Calculating eqn [1], the following equation showing the symmetry against the origin can be obtained:

$$I(r') = |C|^2 (\pi a^2)^2 \left[\frac{2\mathcal{J}_1\left(\frac{2\pi a}{\lambda f}r'\right)}{\frac{2\pi a}{\lambda f}r'}\right]^2 \qquad [3]$$

where r' is the radius from the center of the spot image. $\mathcal{J}_1$ is the Bessel function of order 1. **Figure 10** shows the curve. $\mathcal{J}_1$ is a slowly decreasing oscillatory function. Its shape is similar to that of a sine function. The curve passes the origin showing the odd function.

Therefore, the main feature of eqn [3] is similar to that of eqn [2a] and [2b] and relates to the telescope image of the star, and was first derived by G B Airy (1801–1892), Astronomier Royal of England. When the twinkling due to the air flow can be ignored, the center brightness of the star is proportional to a to the power of 4 as can be seen from the factor $(\pi a^2)^2$. In this case, $2a$ is the aperture diameter of the lens of the telescope. This is the reason why the telescope uses the larger lens. In the case of a micromirror with a reduced size a, the larger diffraction effect appears, even though the almost light power is reflected using the mirror having the diameter larger than the size of the laser beam diameter.

Figure 11 shows the intensity distribution expressed by eqn [3]. **Figure 12** shows the calculated density plot of the diffraction pattern. This circular spot pattern is called an Airy disk. Around the center peak, there are circular rings having a radius r'_{m} for the black one and a radius r'_{M} for the bright one. The condition of these rings can be expressed by the following equations:

$$\mathcal{J}_1\left(\frac{2\pi a}{\lambda f}r'_{\mathrm{m}}\right) = 0 \qquad [4]$$

$$\frac{\mathrm{d}}{\mathrm{d}r}\left[\frac{2\mathcal{J}_1\left(\frac{2\pi a}{\lambda f}r'_{\mathrm{M}}\right)}{\frac{2\pi a}{\lambda f}r'_{\mathrm{M}}}\right] = \mathcal{J}_2\left(\frac{2\pi a}{\lambda f}r'_{\mathrm{M}}\right) = 0 \qquad [5]$$

Table 1 lists the radius of bright (maximum intensity) and black (minimum intensity) rings and the relative intensity compared to the center value. The position of the first minimum ring is given by

$$\frac{2\pi a}{\lambda f}r'_{\mathrm{m}} = 1.22\pi, \quad r'_{\mathrm{m}} = 1.22\frac{\lambda f}{2a} \qquad [6]$$

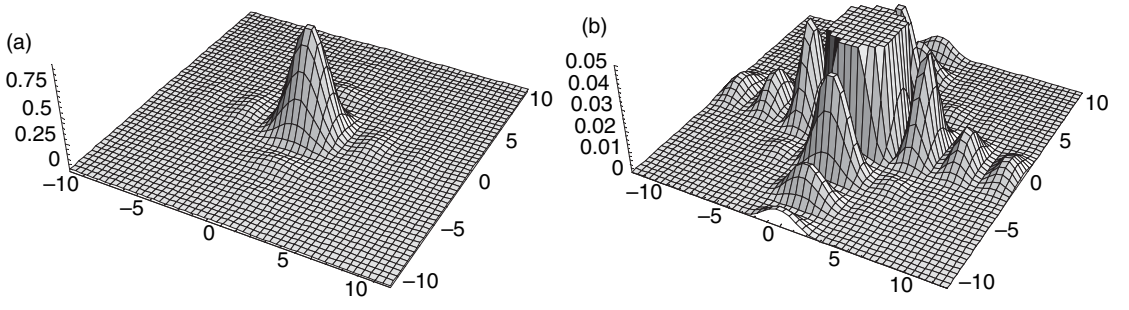


Figure 7 (a) Intensity pattern calculated from eqn [2]. (b) The magnified view is around the center peak. The coordinates in the figures are normalized by the equation $x = (2\pi a/\lambda f)x'$, $y = (2\pi b/\lambda f)y'$.

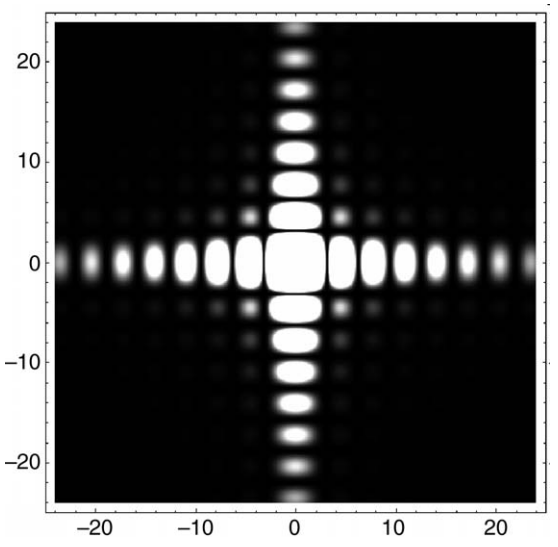


Figure 8 Calculated density plot of the diffraction pattern from a square aperture (or mirror).

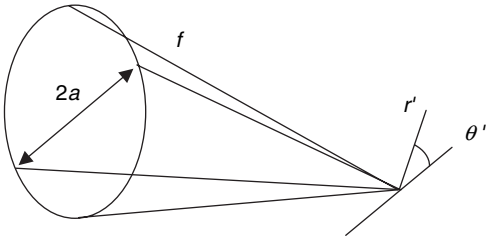


Figure 9 Diffraction from circular mirror.

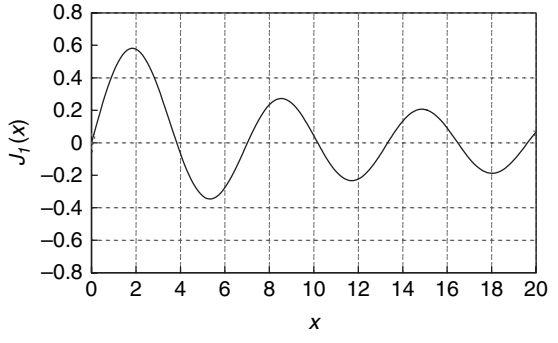


Figure 10 Shape of Bessel function of order 1 of $J_1(x)$.

Comparing to the condition in eqn [2a] and [2b], the difference is the factor 1.22. This larger value can be attributed to the smaller area of the circular aperture πa^2 (the area of the rectangular aperture is $4a^2$). The light intensity of the spot inside the first black ring is ~84% of the total light bundle that has passed through the circular emission pupil.

Figure 13 shows the normalized intensity distributions from rectangular and circular mirrors. The lateral axis is normalized multiplying the factor $2\pi a/\lambda f$ to r or x. The main difference is outside peaks. The inset of Figure 13 shows the magnified profile. The subpeaks of the circular mirror are smaller than those of the rectangular one. Compared to the circle area, the rectangle area includes a larger percentage of the side area far from the center. This will generate the larger contribution of the phase change in eqn [1].

3.01.3.2 Number of Resolvable Spots – Resolution

Setting integer m to be 1 in eqn [2b], the width of the center bright spot obtained from the rectangular mirror can be expressed as follows:

$$x' = \frac{\lambda f}{2a} \qquad [7]$$

This width means the limit of the optically resolvable size of the resolution. For making this peak width smaller, the factor f/a has to be small. The wavelength λ is usually fixed depending on the application. f is the parameter of the outside optics. The parameter of interest for the micromirror is a, and the mirror size $2a$ needs to be large. Figure 14 shows two diffraction profiles. The additional profile shown by the gray curve is set at the black node of the center profile. Here, the relative position shift is $\Delta x' = \lambda f/2a$. When the light sources are incoherent, the image profiles become simple summation of two profiles. The summed intensity profile expressed by the dotted curve is added showing the existence of two peaks. If two peaks further approach each other, these two spots will mix resulting in a single spot. One definition of the resolution that is usually used is the nearest black node $\Delta x' = \lambda f/2a$, which was proposed by Rayleigh. When this value is transformed to the angle, the resolvable diffraction angle $\theta_{\text{diffraction}}$ can be expressed by

$$\theta_{\text{diffraction}} \approx \tan^{-1}\frac{x'}{f} \approx \frac{\lambda}{2a} \qquad [8]$$

For making this result more general, the diverging angle can be expressed as follows:

$$\theta_{\text{diffraction}} = C\frac{\lambda}{2a} \qquad [9]$$

where C is the shape factor of the mirror ($C = 1$ for the rectangle, $C = 1.22$ for the circle). The resolvable

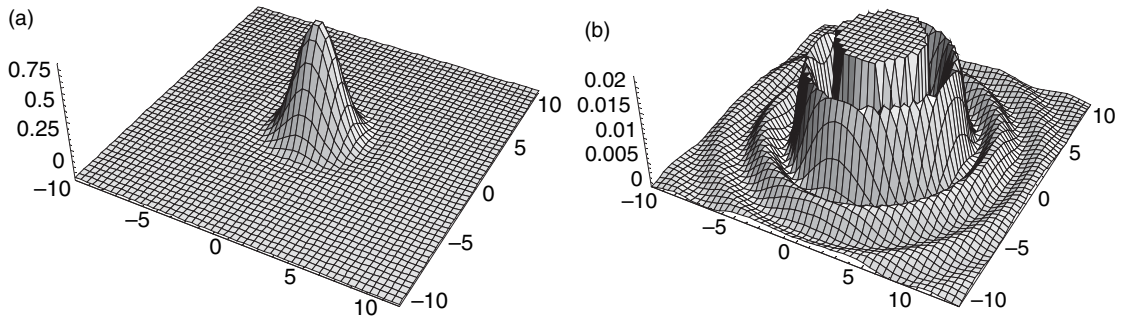


Figure 11 (a) Intensity pattern calculated from eqn [3]. (b) The magnified view around the center peak.

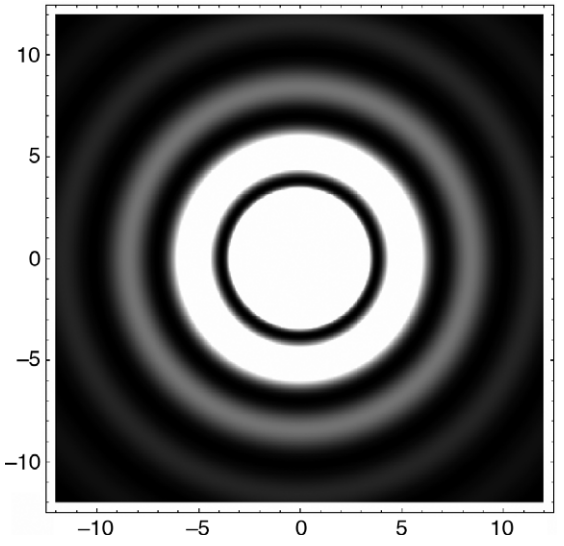


Figure 12 Calculated density plot of Airy disk.

Table 1 The maximum and minimum values of diffracted light from the circular aperture

X (*rad*)	$(2J_1(x)/x)^2$	
0	1	max
$1.220\pi = 3.833$	0	min
$1.635\pi = 5.136$	0.0175	local max
$2.233\pi = 7.016$	0	min
$2.679\pi = 8.417$	0.0042	local max
$3.328\pi = 10.174$	0	min
$3.699\pi = 11.620$	0.0016	local max

pixels of the image obtained by the laser display are discussed using this factor.

Now the resolution of the image obtained from the parallel electrostatic torsion mirror, as shown in Figure 15, can be discussed. The light modulator developed by Petersen (Figure 4) has this structure. The mirror is suspended in the air at a height h from the gap. The driving method is arbitary. The

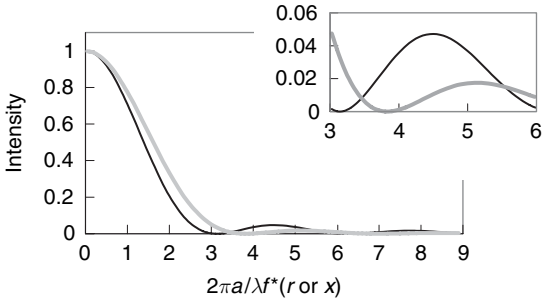


Figure 13 Intensity profiles from circular and rectangular mirrors.

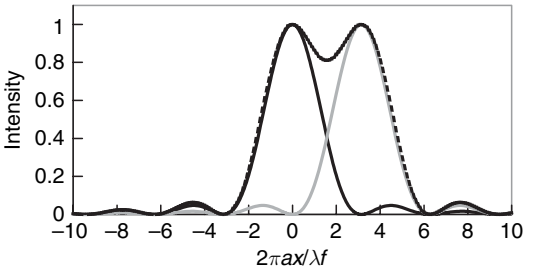


Figure 14 Overlapping point images for illustrating Rayleigh criteria.

maximum rotation angle of the mirror is decided by the geometry as shown by the dashed line.

$$\theta_{\max} = \sin^{-1}\left(\frac{b}{a}\right) \approx \frac{b}{a} \qquad [10]$$

For the center pivot mirrors, $\theta_{\max}a$ corresponds to the motion of the edge of the mirror. The optical scan angle is $2\theta_{\max}$ due to the light reflection.

The number of resolvable spots is obtained as follows:

$$N = \frac{2\theta_{\max}}{\theta_{\text{diffraction}}} \approx \frac{2b}{a}\frac{2a}{C\lambda} = \frac{4b}{C\lambda} \qquad [11]$$

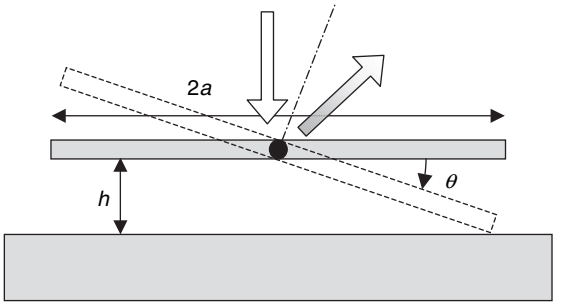


Figure 15 Simplified drawing of the micromirror for discussing the resolution.

To increase the number of resolvable spots, the air gap b has to be large. Supposing the values of $C=1$, $\lambda=0.5\,\mu m$, and $N=1000$ for the high-quality laser display, the height b has to be $125\,\mu m$. This vertical distance is large for micromirrors, requiring special techniques in microfabrication.

When the parallel plate electrostatic actuator is used for driving the mirror, the stable obtainable rotation angle decreases to $\sim 1/3$ of the maximum rotation due to the pull-in instability (Senturia 2001). With the larger height b, the decrease of the electrostatic force becomes the restrictive factor for the actuator design.

3.01.3.3 Static Flatness of Mirror

Let us consider the criterion for the static flatness of the mirror (Nee 2001). **Figure 16** is the illustration of the mirror. The mirror is considered to deform having a diameter D as a section of a sphere with radius r. The total amount of deviation from the flat line at the center of the mirror is assumed to be b. A collimated laser beam incident on a curved mirror will reflect at an angle ϕ and will conform to the following equations:

$$\sin\phi = \frac{D/2}{r} \quad [12]$$

$$\cos\phi = \frac{r-b}{r} \quad [13]$$

From the relation of $\sin^2\phi + \cos^2\phi = 1$, r can be expressed as follows:

$$r = \frac{D^2 + 4b^2}{8b} \quad [14]$$

Inserting this result into eqn [12] and assuming small angle yields:

$$\phi = \frac{4bD}{D^2 + 4b^2} \quad [15]$$

The mirror deformation that results in beam divergence equal to the diffraction Rayleigh limit gives the following relation:

$$b = \frac{\lambda}{4} \quad [16]$$

This is the criterion for static flatness of the mirror. A round trip of the reflecting light beam propagation corresponds to a delay of $\lambda/2$. For the mirror that can generate a diffraction-limited spot, the condition is as follows:

$$b << \frac{\lambda}{4} \quad [17]$$

The bulk mirror element usually satisfies the condition of $<\lambda/10$.

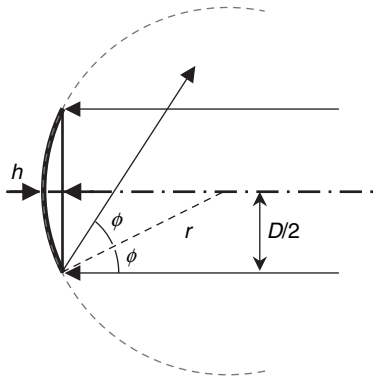


Figure 16 Sketch of relationship between beam divergence and nonplanar mirror surface approximated as part of a sphere.

3.01.3.4 Dynamic Flatness of Mirror

The dynamic deformation of the mirror can be the limiting factor for realizing the laser display, which requires low distortion of the image. The dynamic torque produces a surface distortion at the flyback position where the angular acceleration becomes maximum. This issue was discussed at an early stage (Petersen 1980). **Figure 17** illustrates the schematic drawing where the mirror no longer conforms to the original shape. The high acceleration forces during mechanical deflection of the mirror result in bending of the mirror. Dynamic deformation calculation does not take damping into account. The maximum acceleration is attained at the extremities of the scan, where the velocity is zero.

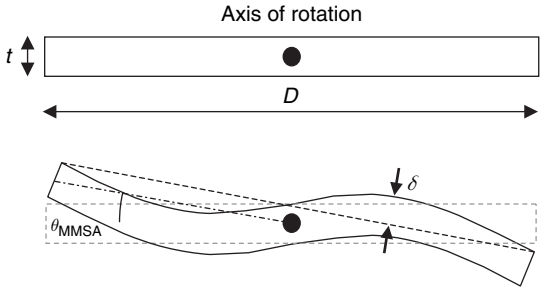


Figure 17 Schematic drawing of a dynamic mirror deformation.

Damping force is related to velocity; thus, the damping should have no effect in the calculation.

For rectangular mirrors, the following formulas can be used for calculating deformation, which is defined as the deviation from linear line shown in **Figure 17**. **Figure 18** shows the shape of the mirror surface as a function of u $(-1 < u < 1)$, which is the normalized mirror coordinate perpendicular to a rotation axis. The positions $u = \pm 1$ correspond to edges of the mirror, where the deformation δ is set to be 0, subtracting the effect of the slant angle.

$$\delta(u) = \frac{\delta_{\max}(u^5 - 10u^3 + 20|u| \cdot u - 11u)}{1.83} \quad [18]$$

$$\delta_{\max} = 0.217 \frac{\rho f^2 D^5 \theta_{\max}}{Et^2} \quad [19]$$

where ρ is the material density, E is the modulus of elasticity (Young's modulus), and f is the scanner frequency, and $\theta_{\max}$ is in radian. The large mirrors scanning at high speeds will have considerable dynamic deformation.

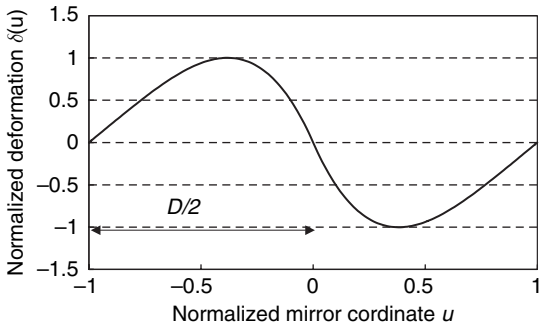


Figure 18 Surface profile as a result of deformation. Maximum deviation occurs at ~40% way from the center to the edge of the scanner.

3.01.4 Recent Progress after *c.* 1999

Around 1999, the technique of the surface micromachining using structural polysilicon (poly-Si) layers was mainly adopted for devices fabrication. This is because the multilayer structures give more design flexibility and are advantageous for realizing high fill factors with actuators and mechanical springs hidden underneath the mirrors (Bustillo *et al.* 1998). The micromirrors prepared by surface micromachining have narrow gaps between the moving element and the substrate at the initial state. As justified in Section 3.01.3, there were many efforts to increase this gap for obtaining larger rotation angles. **Figure 19(a)** shows an example lifting the frame for a 4×4 micromirror array (Su *et al.* 1999). The lifted height is 72 μm taking the balance with the disadvantage of the increase of the driving voltage.

Figure 19(b) shows another example developed in the MicroStar project at Lucent Technologies for realizing a large-scale optical switch (Aksyuk *et al.* 2000). The lifting height from the substrate is 50 μm. The parallel plate-type electrostatic actuator with the underlayered electrodes is adopted. The angular range needed for switching applications is about ±9°. A self-assembly mechanism powered by tailored residual stress in a separate metallized layer is developed.

The limitation of the parallel plate layout can be overcome taking another layout reported from the group of UC Berkeley (Kiang *et al.* 1998a, b, Muller and Lau 1998). **Figure 20** shows an example. Suspended by a frame that is connected to a hinged slider at its back, the micromirror is inclined at 70° to the substrate. The backspace of the mirror is open, and the mirror motion does not suffer from the geometrical limit. Amplification of the mirror motion at resonance is realizable. The linear comb-drive actuator on the substrate drives the mirror. The electrostatic comb drive consists of 100 interdigitated fingers on the shuttle comb and 101 fingers on the stationary comb. The comb fingers are 2 μm wide and 40 μm long, and the entire comb region extends 1 mm from side to side. The shuttle of the comb drive is connected to the bottom of the mirror. The maximum rotation angle $\theta_{\max}$ is decided by the stroke of the linear motion and the position of the rotational axis. In this example, the maximum displacement of the shuttle comb is ±15 μm. Using eqn [11], the number of resolvable spots can increase with the

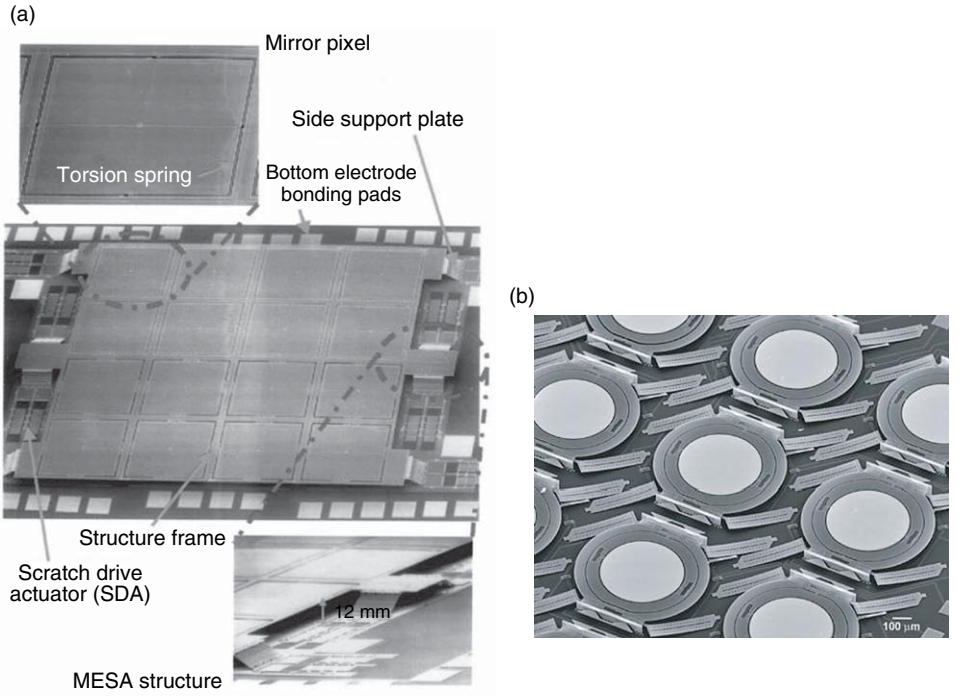


Figure 19 (a) 4×4 mirror array lifted at the structure frame. (Source: Su G-D J, Fan L, Wu M C 1999 Surface-micromachined adaptive micromirror arrays with large strokes. *Proc. 10th Int. Conf. Solid-State Sensors and Actuators*, pp. 578–81.). (b) Arrayed two-axis tilting mirror lifted individually. (Photo courtesy: Alcatel-Lucent.)

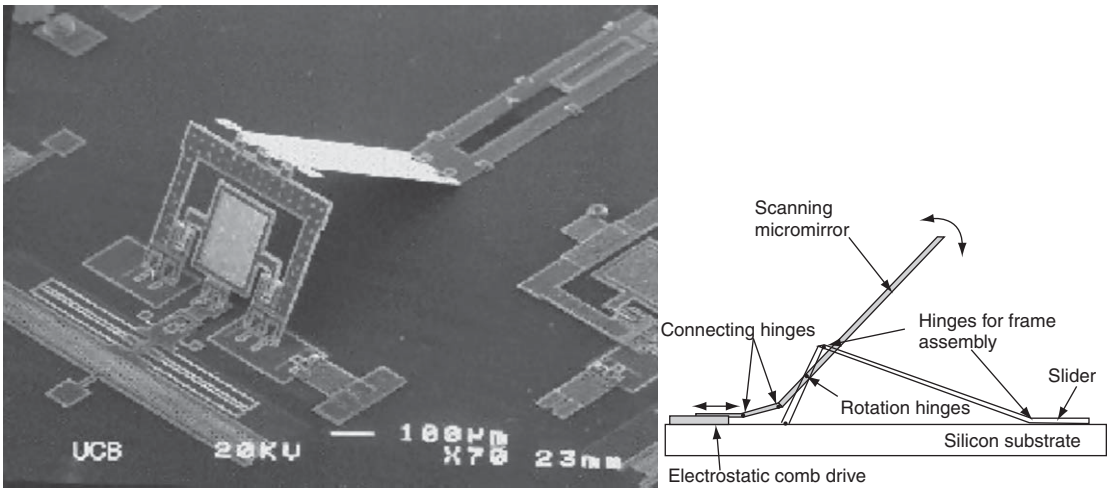


Figure 20 A pin-and-staple hinge scanner of poly-Si films flipped up from the substrate. (Source: Muller R S, Lau K Y 1998 Surface-micromachined microoptical elements and systems. *Proc. IEEE* **86(8)**, 1705–20; Kiang M-H, Solgaard O, Lau K Y, Muller R S 1998a Polysilicon optical microscanners for laser scanning displays. *Sens. Actuators A Phys*. **70(1–2)**, 195–9; Kiang M-H, Solgaard O, Lau K Y, Muller R S 1998b Electrostatic combdrive-actuated micromirrors for laser-beam scanning and positioning. *J. Microelectromech. Syst.* **7(1)**, 27–37.)

maximum rotation angle and the mirror size and is expressed as follows:

$$N = 2\theta_{\max}\frac{2a}{C\lambda} \qquad [20]$$

This is the relation under the static condition. At resonance, $\theta_{\max}$ will be doubled taking the full stroke of the scanning motion. **Figure 21** shows the scanning electron microscope (SEM) image of another

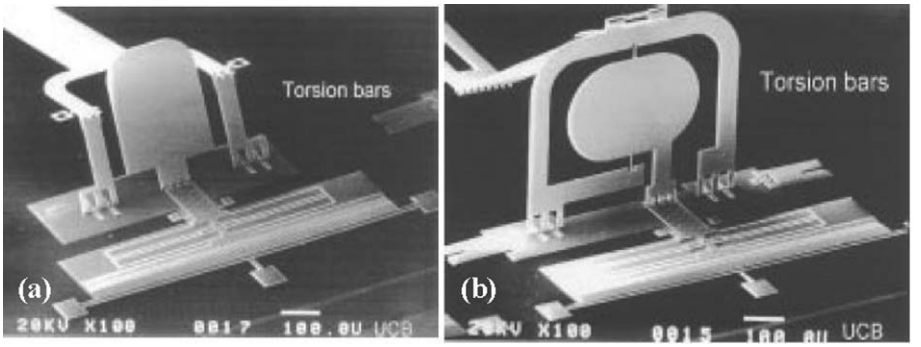


Figure 21 Scanning electron microscope (SEM) photographs of two prototype resonant scanners. (Source: Muller R S, Lau K Y 1998 Surface-micromachined microoptical elements and systems. *Proc. IEEE* **86(8)**, 1705–20.) (a) A 300 × 400 μm scanning mirror that deflects the laser beam in the out-of-plane direction. (b) A 440 × 300 μm scanning mirror that scans the beam in a direction parallel to the substrate.

microscanner constructed with a 300 μm × 400 μm and a 440 μm × 300 μm mirror. The mirror stands upright on the wafer and its rotational axis is normal to the substrate. The system has a resolution of 7 × 11 pixels (Kiang *et al.* 1998a, b).

Figure 22 shows a raster-scanning display using two scanners (Hagelin and Solgaard 1999). The incoming optical beam first reflects off the fast mirror and then strikes the slow mirror. The laser beam is modulated with an acousto-optic modulator that is synchronized to the mirror driving voltages for drawing some images. The fundamental structure of the microscanner is similar to that shown in **Figures 20** and **21**. The diameter of the scanning mirrors is increased to ∼500 μm having nearly circular shape. The maximum optical scanning angle is 15°. The resolution achieved is 102 × 119 pixels. The pixel size is observed to vary according to mirror angle. The smallest pixel size is at the center of the display with no voltage applied to the actuators.

The transition of the micromirror made of thin poly-Si layers to that made of a bulk c-Si layer using

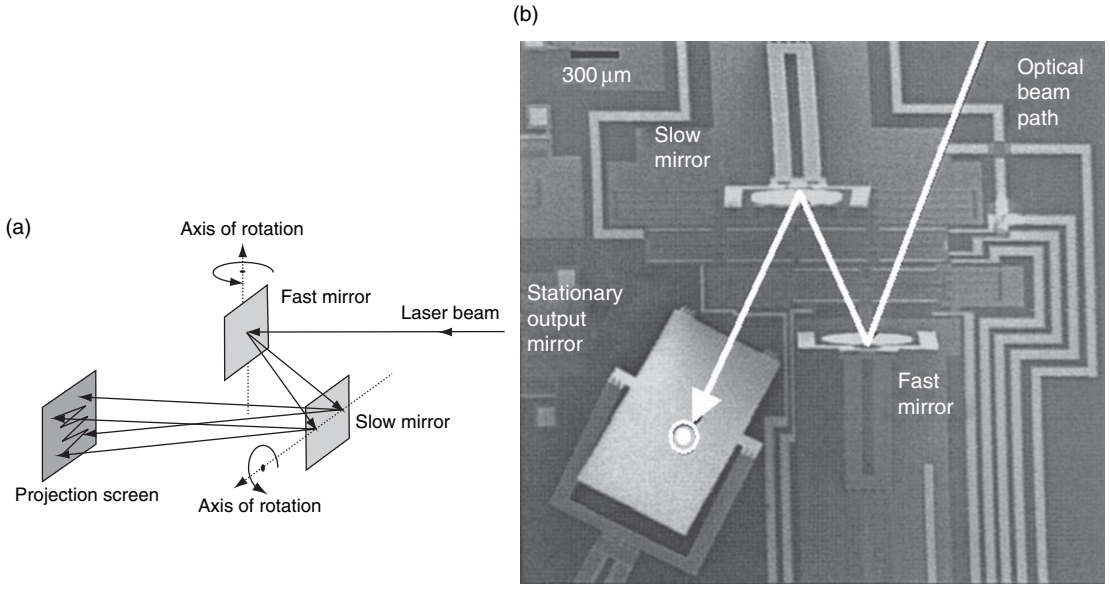


Figure 22 (a) Schematic drawing of a laser display incorporating two micromirrors rotating on orthogonal axes. (b) Top view of a single-chip raster scanner. (Source: Hagelin P M, Solgaard O 1999 *IEEE J. Sel. Top. Quant. Electron.* **5(1)**, 67–74.) The surface deformation of the micromirror was characterized using the stroboscopic interferometer reported in Hart *et al.* (1999). (Source: Hart M R, Conant R A, Lau K Y, Muller R S 1999 Time-resolved measurement of optical MEMS using stroboscopic interferometry. *Proc. 10th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '99)*, pp. 470–3.)

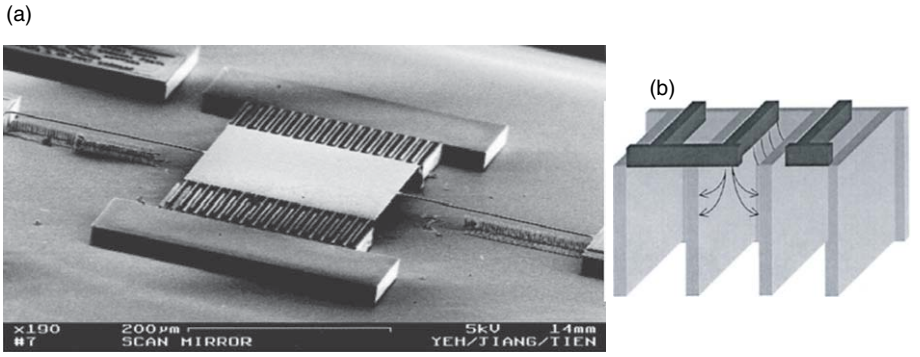


Figure 23 (a) Asymmetric comb-drive actuated membrane fabricated using surface micromachining and deep reactive ion etching (DRIE). (b) Schematic of the electric field lines in an asymmetric comb drive. (Source: Yeh J-L A, Jiang H, Tien N C 1999 Integrated polysilicon and DRIE bulk silicon micromachining for an electrostatic torsional actuator. *J. Microelectromech. Syst.* **48(4)**, 456–65.)

deep reactive ion etching (DRIE) was tried by the group at Cornell University (Yeh *et al.* 1999). The process is a combination of surface micromachining and DRIE of bulk Si. **Figure 23(a)** shows the image of a micromirror. A torsional actuator consists of poly-Si fingers and bulk Si fingers. The suspended mirror is actuated by comb structures moving in the out-of-plane direction. The electrostatic force increases the overlapping capacitance of the electrode when the voltage is applied between the movable and fixed combs as shown in **Figure 23(b)**.

One of the hurdles of the surface micromachined mirror is the mirror curvature. The stress gradient appears in the film and its reduced thickness cannot prevent the mirror from bending. Although this static deformation is overcome using the stress control, the dynamic deformation remains. The dynamic mirror deformation is the deflection caused by the scanning motion of the mirror. The curved mirror surface affects the wavefront of the reflected light beam. The pixel size change is observed (Hagelin and Solgaard 1999). Satisfying the requirements of high resolution and high scan speed has been difficult. The dynamic mirror deformations limit the resolution to less than 20% of the diffraction-limited resolution.

In 2000, Conant *et al.*, reported a new type of micromirror. **Figure 24** shows the micromirror with the so-called staggered vertical comb drive

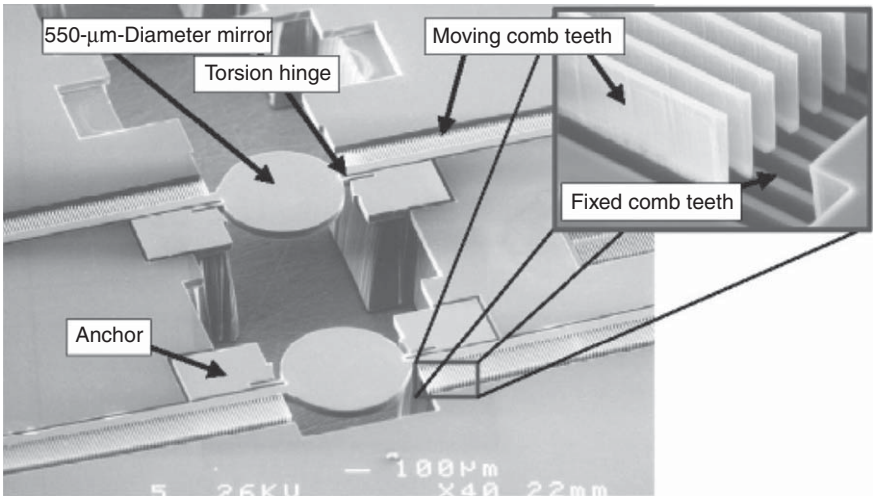


Figure 24 Scanning electron micrograph (SEM) of micromirrors with a staggered vertical comb drive. The inset shows a close-up view of the fixed and moving comb teeth. (Source: Conant R A, Nee J T, Lau K Y, Muller R S 2000 A flat high-frequency scanning micromirror. *Proc. Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, June 4–8, pp. 6–9.)

(SVC) for electrostatic driving, which is made from two layers of bulk c-Si separated by a 1.7-μm-thick SiO_2 layer. The mirror diameter is 550 μm and the resonant frequency is 42 kHz. The mirror, torsion hinge, and moving comb teeth are in the top Si layer. The fixed comb teeth are in the bottom Si layer. Applying the voltage between the top and the bottom layers, the moving comb teeth are attracted to the fixed comb teeth, exerting torque on the mirror generating the rotation. There are no hinges to couple the moving slider and the mirror as in the previous design, as shown in **Figures 20** and **21**. This vertical comb drive offers an advantage over gap-closing parallel plate-type actuators because the energy density in the comb drive is higher; moreover, parallel plate pull-in instability in the actuation is avoided.

The 50-μm-thick c-Si plate improves the rigidity of the mirror. This is valid for not only static but also dynamic deformation. Compared to the surface micromachined mirror, the mirror can be more than 10 times thicker. **Figure 25** shows the deformation of the micromirror where the dynamic deformation reaches its maximum. The total deformation is less than 30 nm.

The spot size and separation at eight different regions across the scan give the device performance. The optical resolution of 350 pixels is measured at the resonant frequency of 34 kHz with the scan angle of ±6.25°. The resolution of this 550-μm-diameter mirror with a 24.9° optical scan and 655-nm laser light is near the diffraction-limited resolution of 356 pixels based on eqn [20] by setting the value C to be 1.03.

Let us consider the dynamic deformation using eqn [19]. For a 550-μm-long rectangular c-Si mirror of 50 μm thickness, half-angle mechanical scan of 6.25°, and resonant frequency of 34 kHz, the estimated dynamic deformation is ~8 nm. This value is much lower than the Rayleigh limit of $\lambda/4$ for 655-nm light. For comparison, a 550-μm-long surface micromachined mirror of thickness 1.5 μm having the same resonant frequency will have the dynamic deformation ~9 μm. Because the dynamic deformation δ_{max} scales as the fifth power of the mirror length D, the small size of the mirror works advantageously. If the system performance is deformation limited, significant gains can be attained by slightly reducing the mirror size.

With greatly improved flatness, the micromirrors fabricated from the silicon-on-insulator (SOI) wafer are developed using a simplified process. The next main issue is increasing the rotation angle. Milanovic *et al.* showed a mirror design using the leverage. The inset of **Figure 26** shows a schematic drawing. By applying the lateral force F on the arm by either pushing or pulling on the arm, the lateral

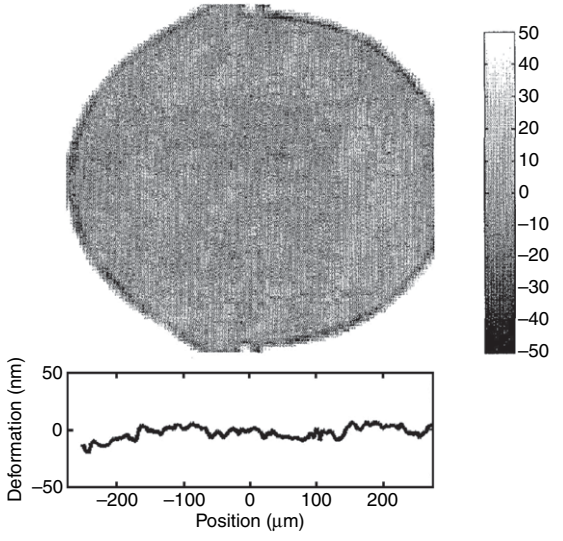


Figure 25 The dynamic deformation of the micromirror at the end of the scan where the dynamic deformation reaches its maximum. (Source: Conant R A, Nee J T, Lau K Y, Muller R S 2000 A flat high-frequency scanning micromirror. *Proc. Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, June 4–8, pp. 6–9.)

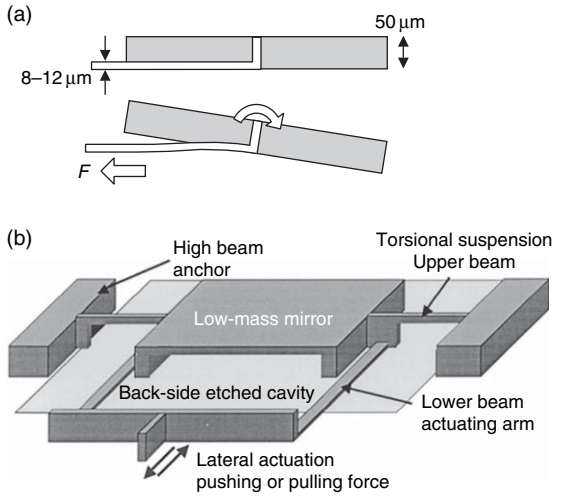


Figure 26 Micromirror with the leverage mechanism for converting the lateral actuation to rotation by applying actuation at points displaced from the shear centers of torsional suspensions. (a) Schematic of the mechanism. (b) Micromirror structure. (Source: Milanovic V, Last M, Pister K S J 2003 Laterally actuated torsional micromirrors for large static deflection. *IEEE Photonics Technol. Lett.* **15(2)**, 245–7.)

force is transferred along linkages and applied below the axes of rotation of the mirror torsion bars, resulting in mirror rotation (Milanovic *et al.* 2003). The fabrication of the time-controlled multi-level DRIE of Si realizes types of beams providing the needed distance between axes and loads. The thickness of the device Si layer is 50 μm. **Figure 26** shows an example of a micromirror with three level structures, which experienced the largest mirror rotation angle in their designs. Lower beams are thinned to a height of 8-12 μm. **Figure 27** shows the demonstrated optical beam deflection. Applying the bias from 0 to 130 V, a static mirror rotation reached up to 22° optical (mirror rotation is half that value, i.e., 11°) in the pull mode, and 21° optical in push mode. The first resonant mode is >2 kHz.

After the invention of SVC, another category of actuator, the angular vertical comb drive (AVC) was developed. **Figure 28(a)** and **28(b)** shows the schematic drawing for explaining the difference between AVC and SVC, respectively. In an SVC, the moving fingers have the offset vertically from the fixed fingers. The AVC has the slant fingers, which require some new fabrication technique. Hah *et al.* (2004a) have demonstrated an AVC using reflow of photoresist. **Figure 28** shows the fabricated micromirror. After the release of the structures, the wafer was placed in deionized water at about 100°C to reflow the photoresist hinge. The inset is an enlarged view of the AVC. The moving fingers are patterned in the same layer as the fixed fingers and then tilted upward using the surface tension of the photoresist. The fingers are self-aligned. Compared to the SVC, AVC can generate a larger overlap capacitance between the moving and the fixed comb fingers. The maximum scan angle of the AVC increases up to 60% compared to that of the SVC for the same comb-finger dimensions with the enlarged stable region.

3.01.5 Classification of Micromirrors

3.01.5.1 Actuation Methods

Micromirrors using the electrostatic force have been described previously. A wide variety of actuation methods have been developed. Although the magnitude of the electrostatic force is generally small, electrostatic actuation seems to be adequate as light does not introduce any load. On the contrary, the other types of actuators can produce larger forces, and are suitable for performing external tasks (e.g., manipulation of objects). A comparison of actuation methods is important for satisfying the application requirements.

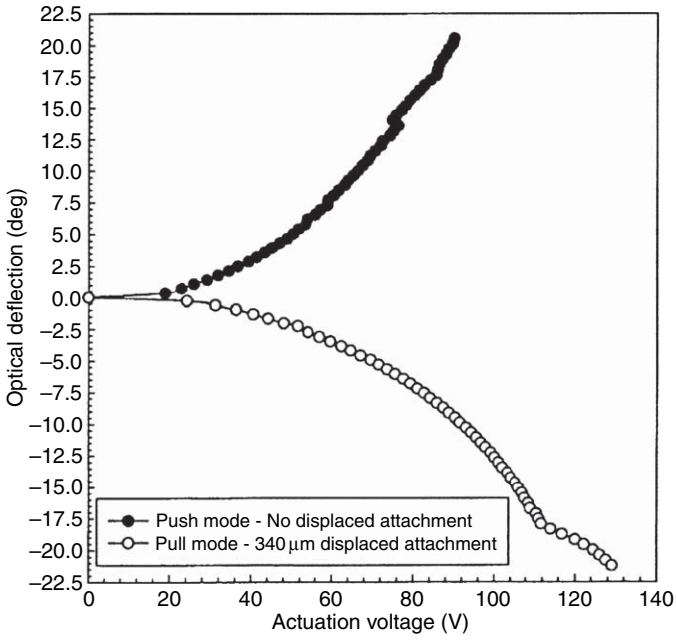


Figure 27 Demonstrated static optical deflection angle. (Source: Milanovic V, Last M, Pister K S J 2003 Laterally actuated torsional micromirrors for large static deflection. *IEEE Photonics Technol. Lett.* **15(2)**, 245–7.)

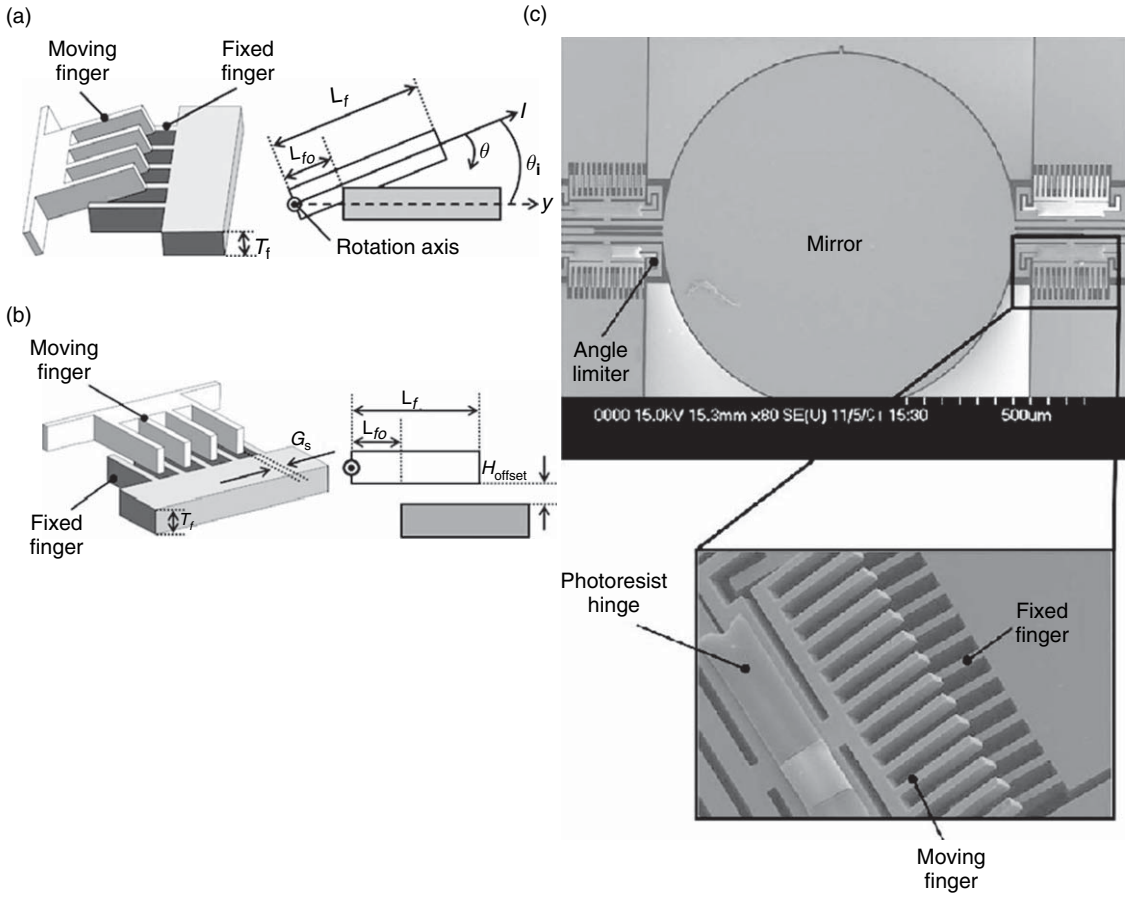


Figure 28 Schematic diagrams of (a) angular and (b) staggered vertical comb-drive actuators. (c) Scanning electron microscope (SEM) image of a micromirror with angular vertical comb-drive actuators assembled by reflow of photoresist hinge. (Source: Hah D, Patterson P R, Nguyen H D, Toshiyoshi H, Wu M C 2004a Theory and experiments of angular vertical comb-drive actuators for scanning micromirrors. *IEEE J. Sel. Top. Quant. Electron.* **10(3)**, 505–13.)

The actuators mainly consist of four categories: electrostatic, piezoelectric, thermal, and magnetic driving (Bell *et al.* 2005). In the category of magnetic actuators, a magnetostrictive actuator is included. Piezoelectric and magnetostrictive materials have intrinsic actuation capabilities. Magnetostrictive materials produce enough force to build microactuators (Bourouina *et al.* 2002). Thermal expansion and phase transformations such as the shape–memory effect cause shape or volume changes. It is appropriate to conduct the assessment and comparison for these types of devices. In the following sections, examples of individual actuators are described.

3.01.5.1.1 Electrostatic actuators

The electrostatic actuator is rather preferred as the device can be completed within a chip, making integration easy. The actuator consumes little power. Heat dissipation is zero, in principle, as there are no steady-state currents caused by capacitive loads. The degradation in device reliability due to heat generation can be prevented. Here, different variations are described.

Figure 29 shows one of the sophisticated surface micromachined mirror devices for 2D optical switching developed by Lin *et al.* (1999) for the AT&T project. The free-rotating mirrors are monolithically integrated on a substrate. The collimated light beam is switched to the desired output port by rotating a mirror. **Figure 30** shows a schematic drawing showing the mechanism in a single micromirror. The mechanical linkage converts the linear displacement of the microactuators to the rotation of the mirror. The hinges anchor the mirror on the Si substrate. The mirror is connected to the translation stage by push rods. The translation stage is actuated by arrays of scratch-drive actuators (Akiyama *et al.* 1993, 1997).

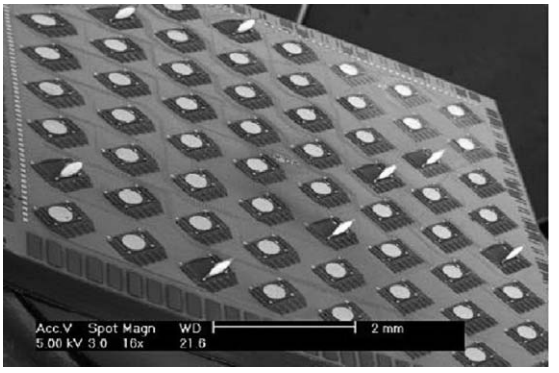


Figure 29 Two-dimensional optical cross-connect consists of mirrors, which can be flipped up from the substrate using the microactuators. (Photo courtesy: Lih Y. Lin.)

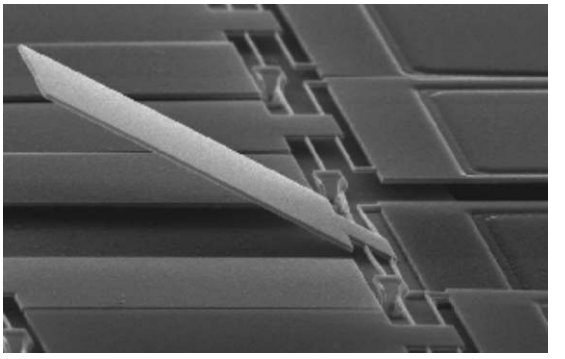


Figure 31 Scanning electron microscope (SEM) image of the double-hinged tilting mirror showing the snap down position of the actuator. (Photo courtesy: Alcatel-Lucent.)

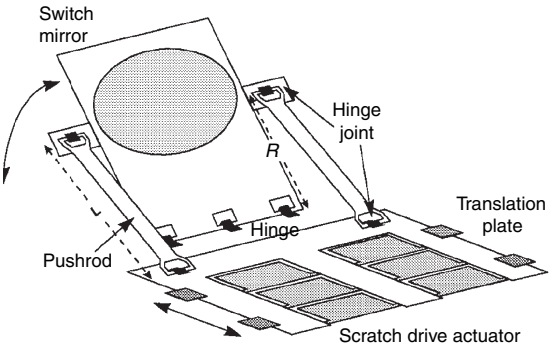


Figure 30 Schematic of the micromirror illustrating a free-rotating mechanism. (Source: Lin L Y, Goldstein E L, Tkach R W 1999 Free-space micromachined optical switches for optical networking. *IEEE J. Sel. Top. Quant. Electron.* **5(1)**, 4–9.)

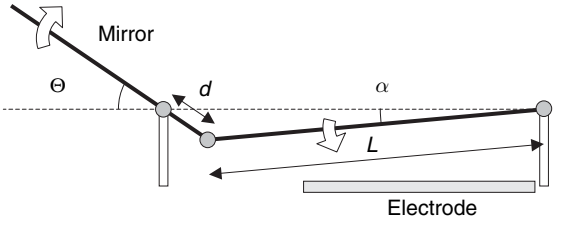


Figure 32 Schematic drawing of the double-hinged linkage for the angular amplification.

Some surface micromachined micromirrors are used to break the limitation of the rotation angle expressed by eqn [11]. **Figure 31** shows a 1D mirror array developed by Lucent Technologies (Lopez *et al.* 2002). The mirror size is $30\,\mu m \times 50\,\mu m$. The mirror is suspended at the side end, and the rotation toward the out-of-plane direction is generated with the neighboring actuator motion via the mechanical linkage. With the lever arm layout, the small downward displacement inside the actuator is transmitted as a large mirror rotation. The parallel plate-type electrostatic actuator shown at the right-hand side of **Figure 31** matches the structure. **Figure 32** shows a schematic drawing of the mechanism. The device consists of a double-hinged tilting mirror. One hinge anchors the mirror to the substrate and the other attaches the mirror to the moving plate electrostatically. When a driving voltage is applied, the moving plate is attracted downward, causing the mirror to pivot around the anchored hinge. This design has advantages. The micromirror is decoupled from the need for a large gap under the moving element. A double-hinged mechanism works as an angle amplification tool allowing large out-of-plane rotating mirror angles with small gap actuator designs. Neglecting sag and actuation plate deformation, the mirror rotates an angle θ given by:

$$\sin\theta = \frac{1}{d}\sin\alpha \qquad [21]$$

where α is the plate deflection angle, l is the length of the actuation arm, and d is the distance between the hinges. The larger rotation angle can be obtained with the amplification factor l/d. The maximum angle can be increased either by reducing the length d or by increasing the gap l. By separating the areas of actuators and mirrors, this mirror consists of a 1D array with a high fill factor. The larger area can be used for preparing micromirrors. **Figure 33** shows the voltage dependence of the angle θ. About 14° of the mirror angle is achieved at a voltage of <60 V with amplification factor of ~10. The inset shows a contour plot of an actuator element.

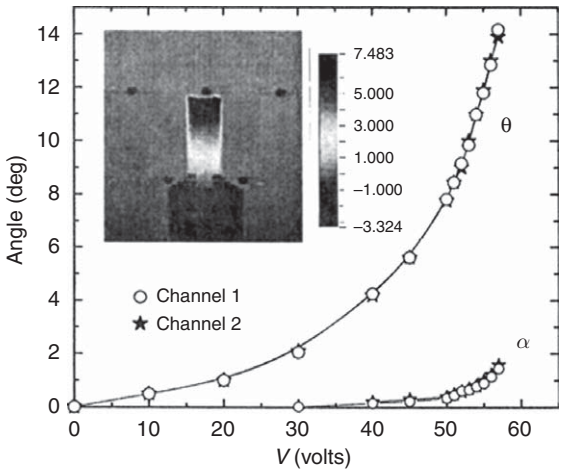


Figure 33 Rotation angle (θ) of two different mirrors and its respective actuator plate (α) as a function of the driving voltage. (Source: Lopez D, Simon M E, Pardo F, Aksyuk V, Klemens F, Cirelli R, Neilson D T, Shea H, Sorsch T, Ferry E, Nalamasu O, Gammel P L 2002 Monolithic MEMS optical switch with amplified out-of-plane angular motion. *Proc. IEEE/LEOS Int. Conf. Optical MEMS*, pp. 165–6.)

Figure 34 shows another design for realizing large mirror rotation (Greywall *et al.* 2003). The mirrors are made of c-Si and are covered with a thin layer of aluminum. The basic idea is to use thick poly-Si electrodes located on top of the SOI layer. The thick electrodes generate the offset geometry, generating a significant electrostatic interaction between the nearly horizontal mirror plate and the vertical sidewall via electrostatic fringing fields. **Figure 35** shows the mirror rotation as a function of the driving voltage. The maximum rotation of 9.2° is obtained at 175 V. The mirror is made from the c-Si device layer of the SOI substrate with a thickness of 1 mm. A 10-μm-thick *in situ* doped poly-Si layer is used to create the thick electrodes.

3.01.5.1.2 Piezoelectric actuators

Piezoelectric actuators are attractive because they can generate large stresses and high-energy density; they can be operated within a large bandwidth and potentially require lower operating voltages. The induced strain is proportional to the applied electric field. The same material can be used as the sensor based on the inverse effect measuring the piezoelectric voltage or the piezoelectric current charge when the separated electrodes are prepared. Disadvantages are the smaller displacement and the need for developing the delicate deposition process of the material. A thick film (1–100 μm) is necessary for the actuator. The deposition of piezoelectric material including the preparation of electrodes and the micromachining process are challenging. For skipping the deposition process, an external bulk piezoelectric actuator is sometimes assembled with passive micromachined structures (Yamada *et al.* 1998).

For amplifying the small displacement of the lead zirconate titanate (PZT) film and generating angular displacement of the mirror, the micromirrors are frequently combined with adapted mechanisms. **Figure 36** shows the schematic drawing for generating the larger rotation of a mirror using a pair of piezoelectric unimorph actuator beams. With the driving voltage, the right beam curves downward because of expansion of a PZT film, whereas the left beam bends upward as a result of film contraction with a reverse polarity. These bending motions are converted into a tilting motion of the torsion bar attached at the middle.

Figure 37 is an example of a 1D microscanner (Kobayashi *et al.* 2005b). The cantilever driven by the PZT film deforms the center hinge generating a steep angle at around the torsion bar. The scanner has Pt (0.2 μm)/Ti (0.05 μm)/PZT/Pt (0.2 μm)/Ti (0.05 μm)/SiO_2/Si as a multilayered structure. The PZT film is deposited by chemical solution deposition.

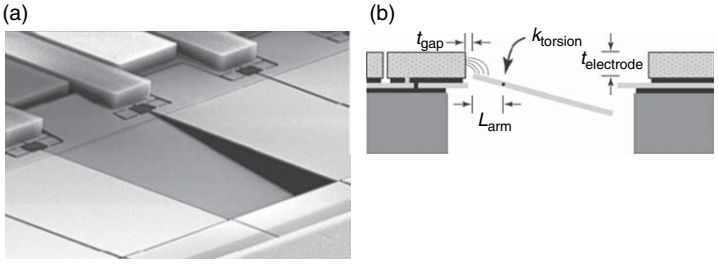


Figure 34 (a) Scanning electron microscope (SEM) image showing a linear array of 80-μm-wide micromirrors. (b) Schematic drawing shows the cross-sectional view of a tilting mirror with fringe-field activation. (Source: Greywall D S, Pai C-S, Oh S-H, Chang C-P, Marom D M, Busch P A, Cirelli R A, Taylor J A, Klemens F P, Sorsch T W, Bower J E, Lai W Y-C, Soh H T 2003 Monolithic fringe-field-activated crystalline silicon tilting-mirror devices. *J. Microelectromech. Syst.* **12(5)**, 702–7.)

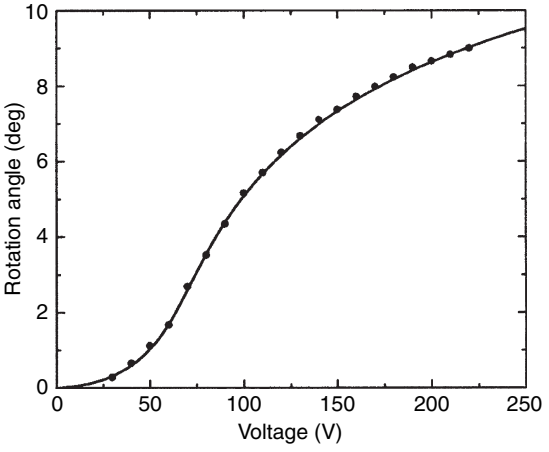


Figure 35 Measured rotation angle as a function of activation voltage. (Source: Greywall D S, Pai C-S, Oh S-H, Chang C-P, Marom D M, Busch P A, Cirelli R A, Taylor J A, Klemens F P, Sorsch T W, Bower J E, Lai W Y-C, Soh H T 2003 Monolithic fringe-field-activated crystalline silicon tilting-mirror devices. *J. Microelectromech. Syst.* **12(5)**, 702–7.)

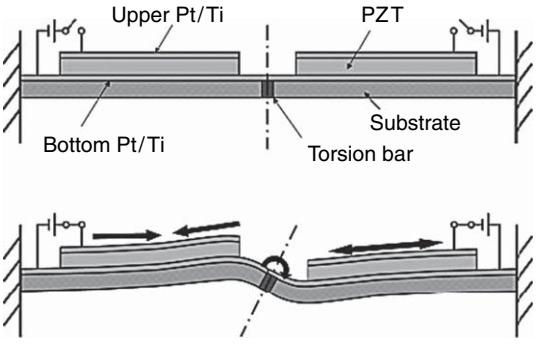


Figure 36 A typical mechanism of generating rotational motion using a pair of piezoelectric unimorph actuator beams. (Source: Yasuda Y, Akamatsu M, Tani M, Iijima T, Toshiyoshi H 2006 Piezoelectric 2D-optical micro scanners with PZT thick films. *Integrated Ferroelectrics* **80**, 341–53.)

A commercially available solution (PZT-20, Kojundo Chemical Co., Japan) with a Pb:Zr:Ti ratio of 120:52:48 was used as the precursor solution. Coating, drying (120°C), pyrolysis (250°C), and rapid thermal annealing (630°C) were repeated more than 10 times to form a 1.1-μm-thick PZT film (Kobayashi *et al.* 2005a). A scanning angle of 25° at the driving voltage of 20 V is demonstrated at the resonant frequency.

Yasuda *et al.* (2006) report 2D optical scanners. The arc discharged reaction ion plating is developed for depositing a thick PZT film. **Figure 38** shows the image of a scanner. A double gimbal structure is used to tilt the ellipse mirror (1 mm × 2 mm) at the

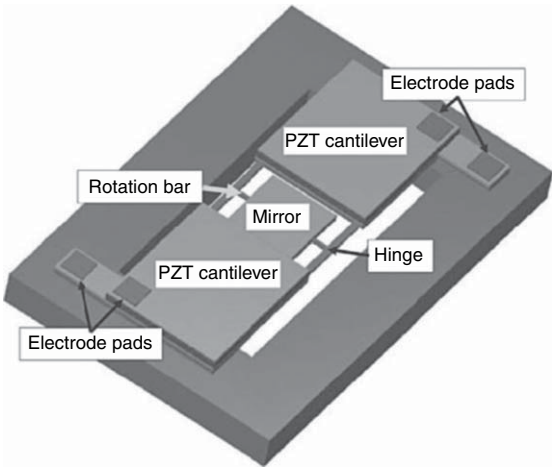


Figure 37 Schematic design of a 1D microscanner. The 50-μm-wide rotation bars, the hinges and the lead zirconate titanate (PZT) cantilevers (1.5-mm-long × 1.7-mm-wide) support the square mirror (1 mm × 1 mm). (Source: Kobayashi T, Tsaur J, Maeda R 2005b Fabrication of optical micro scanner driven by PZT actuators. *Jpn. J. Appl. Phys.* **44(9B)**, 7078–82.)

center two-dimensionally. The inner set of short torsion bars and small U-shaped unimorph actuators is used for the fast motion of the mirror around the *x*-axis. A slow scan around the *y*-axis is made using the outer set of long torsion bars and larger U-shaped actuators.

3.01.5.1.3 Thermal actuators

Although thermal actuators consume more power and have a lower bandwidth (determined by thermal time constants), they can generate a larger force. The fabrication is generally simple and their structure allows easy integration. **Figure 39** shows a well-designed micromirror with the cantilever beams (Schweizer *et al.* 1999). The mirror is attached to the end of two unimorph cantilever beams that are thermally actuated (Schweizer *et al.* 2000). **Figure 39** shows the state at the rest position. With the bending using the residual stress, the mirror is reversed as the initial condition. The unimorph cantilever consists of a metal film (Cr) and SiO_2 layers. The cantilever bends upward. The residual film stresses (the tensile stress in the metal layer and the compressive stress in the oxide) can control the rest position θ_{initial} setting the mirror at an out-of-plane position.

The freestanding cantilever can be resistively heated with an electric current through the conductive layer. The electrical current flows forth on one

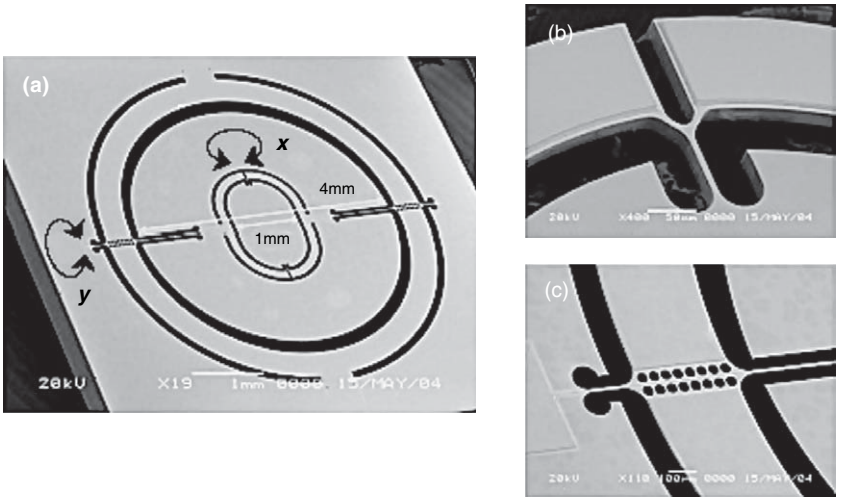


Figure 38 (a) Micrograph of fabricated piezoelectric 2D scanner. The magnified views of (a) and (b) are the tilting mechanisms for the *x*- and *y*-axis, respectively. (Photo courtesy: Stanley Electric Co. Ltd.)

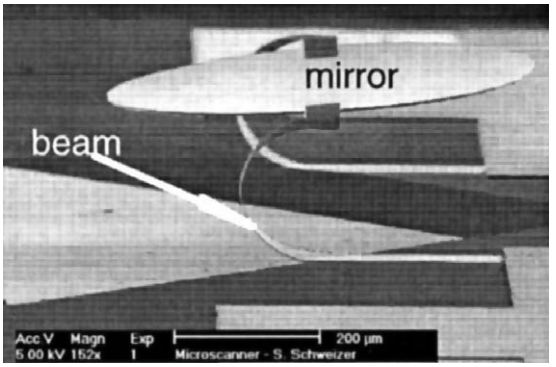


Figure 39 The image of the lateral fixed beams design: mirrors 600 μm × 600 μm, beam lengths 300 μm and thickness 5 μm, and resonance frequency 450 Hz. The mirror swings with the bending of the cantilever beams. (Source: Schweizer S, Calmes S, Laudon M, Renauld Ph 1999 Thermally actuated optical microscanner with large angle and low consumption. *Sens. Actuators* **76**, 470–7.)

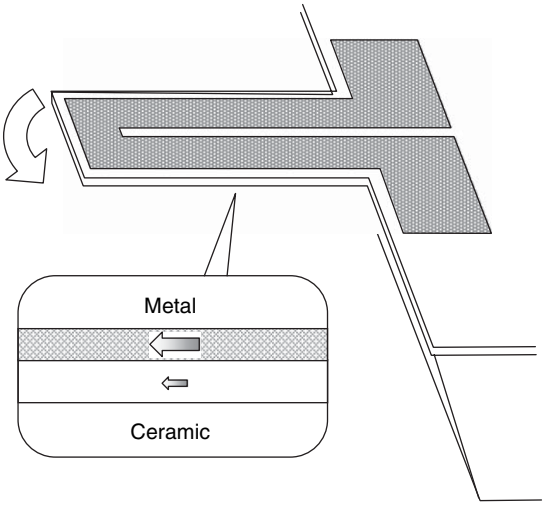


Figure 40 Schematic drawing for illustrating the unimorph cantilever with the thermal actuation. With the increase of the temperature, the cantilever bends downward due to the larger extension of the metal layer compare to the ceramic layer (e.g., oxide).

beam and back on the other. The temperature increase induces the thermal expansion.

$$\varepsilon = \alpha \cdot \Delta T \quad [22]$$

$$\sigma = E\varepsilon = E\alpha \cdot \Delta T \quad [23]$$

Because the stresses generated by thermal expansion are large, whereas the displacements are small, the mechanism amplifying the displacement is a good combination. Unimorph or bimorph structures are usually used. Figure 40 illustrates the situation inside the unimorph cantilever. The materials are selected for obtaining a large difference of α between two layers, as the difference of thermal expansion between the two layers generates the bending motion. The combination of a metal and a ceramic are popular (e.g., Si/Al, SiN/Al). Because the driving force is proportional to ΔT, the temperature control of the device is important for the accurate rotation angle.

Figure 41 shows the scan amplitude of the mirror as a function of the excitation frequency. The onset of thermal cutoff occurs at ~20 Hz and the resonance

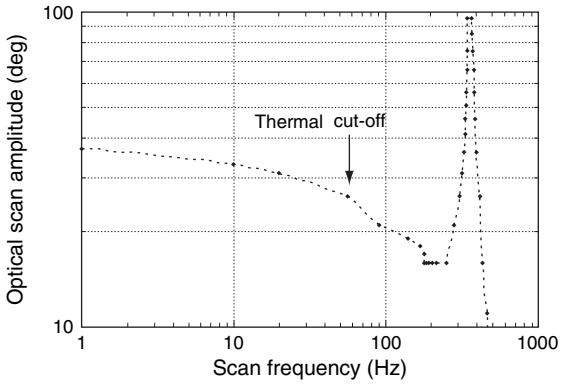


Figure 41 Frequency response of a 500 μm × 800 μm micromirror with a 300-μm-long beam. (Source: Schweizer S, Calmes S, Laudon M, Renauld Ph 1999 Thermally actuated optical microscanner with large angle and low consumption. *Sens. Actuators* **76**, 470–7.)

frequency is 30 Hz. Note that the device can be excited at its mechanical resonance frequency over the thermal bandwidth. The quality factor Q corresponding to the energy loss of the resonator was about 20–40 for various mirror designs.

Figure 42 shows another thermal actuator demonstrating the integration with the bulk 3D micro-optical bench (Sasaki and Yamaguchi 2003b). The V-shaped cantilever consists of doped poly-Si, SiN, and backside Al layers. The cantilever deflects upward with the temperature increase. A skew micromirror at the free end of the cantilever is scanned. The region on the left is the terrace and the laser diode (LD) chip mounted facing the scanning mirror. The region on the right is the thermal actuator. **Figure 42(a)** and **42(b)** shows the thermal actuator under resting and driving conditions, respectively. In **Figure 42(b)**, the bright spot of LD emission is observed on the scanning mirror.

3.01.5.1.4 Magnetic actuators

As described in Section 3.01.3, the mirror size has to be large for obtaining fine spots. The mirror size can be a few millimeters. The electromagnetic force becomes an attractive candidate for the actuation. The electromagnetic driving is adopted in some promising products as described later. Papers on the electromagnetic scanner can be found from 1991 onwards (Wagner *et al.* 1991). **Figure 43(a)** shows a schematic drawing for explaining torque generation. A one-axis mirror is shown for simplicity. The basic design consists of an integrated planar coil on a micromirror plate. **Figure 43(b)** is an example of a moving coil (Sasaki and Miura 2002). When placed in an external field of magnetic flux densities $\boldsymbol{B}$, applied in directions perpendicular to torsion bars, the current $\boldsymbol{I}$ in the driving coil generates paired Lorentz forces $\boldsymbol{F} = \boldsymbol{i} \times \boldsymbol{B}$. The paired forces are generated on the coil. The pure torque is obtained for driving the mirror. The rotation center is at the center axis of the coil. The electromagnetic force and the driving torque are proportional to the current $\boldsymbol{I}$. This linearity is advantageous for operation. After the mirror tilts, the restoring forces of the torsion bars are balanced with the corresponding rotational torques. Because the torsion bar restoring torque is nearly linear against the rotation angle, the system is stable. By controlling the magnitudes of the currents $\boldsymbol{I}$, the tilt angle control is stably obtained.

A series of micromirrors are commercially available from Nippon Signal Co. Ltd. **Figure 44** shows a photograph of a 2D micromirror. **Table 2** shows the performance of the micromirror. The tilt angle is in the ±10° range and the resonance frequency is a few kilohertz. The mirror size strongly relates to the performance of the scanner (the details are described later). The mirror is processed from single c-Si, as the energy loss can be suppressed.

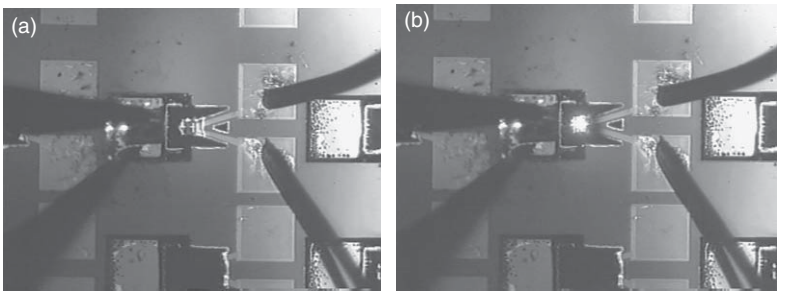


Figure 42 Mounted edge-emitting laser diode (LD) chip and the optical scanner at (a) resting and (b) driving conditions.

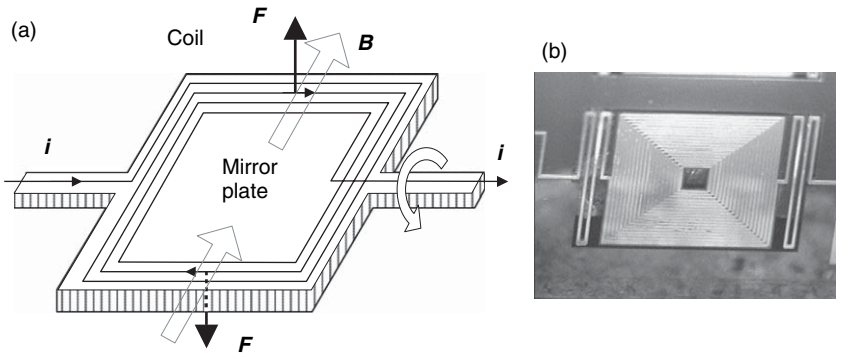


Figure 43 (a) Schematic drawing for showing the electromagnetic driving. (b) An example image of moving coil.

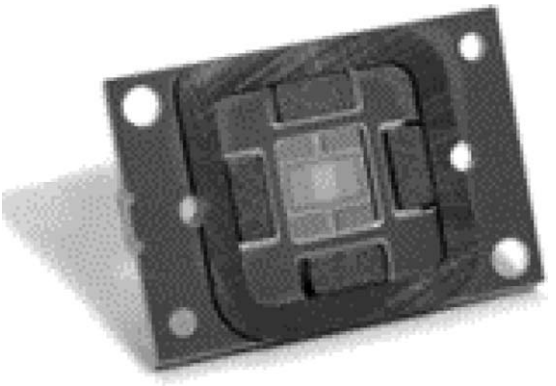


Figure 44 Optical micrograph of a 2D micromirror. (Photo courtesy: Nippon Signal Co. Ltd., http://www.signal.co.jp/vbc/mems/.)

Table 2 A series of micromirrors and their performances

Mirror size (mm^2)	*Resonance frequency (kHz)*	*Mirror vibration angle (°)*
5×6	0.5	± 17
5×10	1.2	± 5
4×4	1	± 9
4×4	2	± 6
4×4	4	± 3
4×4	8	± 1.5
3.5×2.2	16	± 1

Source: Nippon Signal Co. Ltd., http://www.signal.co.jp/vbc/mems/).

3.01.5.2 Mirror Motions

3.01.5.2.1 Rotation – one and two axes

Ever since the development of the micromirror by Petersen (1980), a mirror plate tilting around the torsional support is the standard design for a micromirror. The mirror with a two-axis rotation is the extension of the design shown in **Figures 19** and **38**. In addition to the simple mirror rotation, various movements can be obtained in the micromirror device. Here, the micromirrors are classified from the viewpoint of mirror motion. The mirror motion is closely related to the performance and the function of the optical system.

3.01.5.2.2 Linear motion

The linear mirror movement can realize the crossbar 2×2 optical switch (Marxer *et al.* 1997). **Figure 45** is an illustration for explaining the mirror motion. Two paired optical fibers with sharpened ends (20 μm in diameter) are placed face to face (separation between fiber ends is 50 μm). Because the propagating beam has a small diameter, the linear displacement of ~24 μm (estimated from the image) changes the optical propagation condition dramatically. **Figure 45(a)** is the bar state, and **Figure 45(b)** is the cross state. **Figure 46** shows a device image. The micromirror is at the end of the moving part of the lateral comb-drive actuator. Compared to the devices using waveguides, the switching function does not depend on the wavelength, which is the common feature of the mechanical principle. This device uses the etched sidewall as the reflecting mirror. The thickness of the device Si layer is 75 μm. The core height of the standard single-mode fiber is nearl the top region ($\sim 62.5 \pm 5$ μm) of the sidewall. The gold is deposited for a proper reflectivity of the mirror.

This device opened the new area called SOI-MEMS demonstrated by the group from the University of Neuchatel in *c.* 1997 (before the development of SVC). This 2×2 optical switch was later commercialized by Sercalo Microtechnology Ltd., a

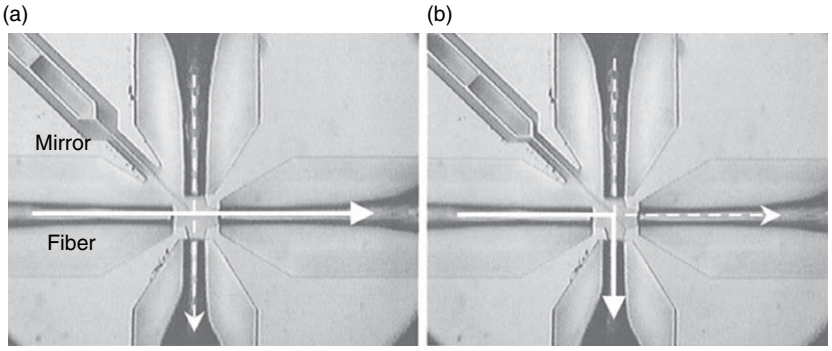


Figure 45 Linear motion of the shutter realizing 2 × 2 optical switch: (a) bar state and (b) cross state. (Source: Marxer C, Thio C, Gretillat M-A, de Rooij N F, Battig R, Anthamatten O, Valk B, Vogel P 1997 Vertical mirrors fabricated by deep reactive ion etching for fiber-optic switching applications. *J. Microelectromech. Syst.* **6(3)**, 277–85.)

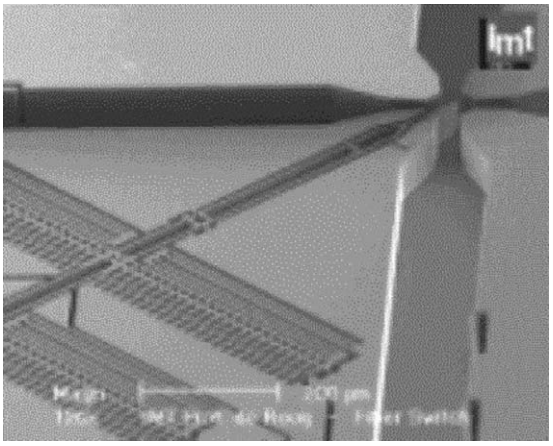


Figure 46 2 × 2 optical switch. The lateral shutter actuator is fabricated with the U-shaped fiber guides at the same time. (Photo courtesy: Nico F. de Rooij.)

venture company (Noell *et al.* 2002). The fabrication process is simplified to a single patterning and one etching step. The critical process step is the DRIE that determines the structure itself and surface roughness of the sidewalls. Four optical fibers are attached to the device. The mechanical guiding structure of U-grooves and the leaf springs to hold the fiber are integrated to guarantee alignment accuracy. To enhance optical performance and minimize loss, all the optical components are completely surrounded by a refractive index matching fluid. The index step between the fiber end faces and the ambient is decreased, and the divergence of the freely propagating light in the ambient and hence the beam diameter is decreased. The coupling efficiency between the fibers is increased. The performance of the devices shown in the homepage of Sercalo Microtechnology Ltd. includes repeatability of <0.01 dB, a response time of <1 ms, a cycling rate of >500 Hz, an insertion loss of <0.9 dB (0.3–0.5 dB in the paper), and the crosstalk of <50 dB (−70 dB in the paper) (Noell *et al.* 2002).

For using the etched sidewall as the mirror, the surface roughness determines the optical performance. Even a small degree of roughness produces measurable light scatter. For a gently sloped surface with a Gaussian distribution of the surface height, the amount of scattered light is estimated from the rms surface roughness σ using the relationship

$$\frac{P_{\text{scat}}}{P_{\text{tot}}} = 1 - \exp\left(-\left(\frac{4\pi\sigma\cos\theta_{\text{i}}}{\lambda}\right)^2\right) \qquad [24]$$

where P_{scat} is the flux of light scattered away from the specific direction, P_{tot} is the total reflected flux, and θ_{i} is the incidence angle.

Figure 47 shows the sidewall surface reported (Marxer *et al.* 1997). The surface quality is found to be best from the top surface to an etched depth of 30 μm. Between the 30- and 45-μm etch depth, vertical notches are visible; finally, between the 45- and 75-μm etch depth, the surface quality is degraded with a surface roughness of up to 1000 nm. The area for light reflection is selected to be the top region. The above features depend on the nature of the plasma etching and the performance of the etching system. The repeated pattern in the side profile is called scalloping. This is the trace of the cyclic etching and deposition steps (Bosch process).

There are researches for smoothening this scalloping. Decreasing the period of this etching/deposition is one method for decreasing the surface roughness of this scalloping (Johnson *et al.* 2005). The

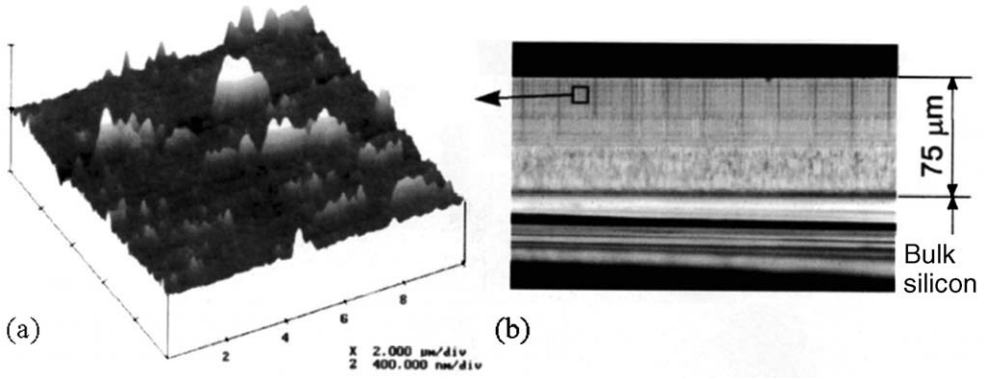


Figure 47 Surface profile of (a) magnified and (b) large area of etched sidewall used as the mirror. (Source: Marxer C, Thio C, Gretillat M-A, de Rooij N F, Battig R, Anthamatten O, Valk B, Vogel P 1997 Vertical mirrors fabricated by deep reactive ion etching for fiber-optic switching applications. *J. Microelectromech. Syst.* **6(3)**, 277–85.)

switching time is usually limited by the time constant of the mass flow controller. The bypass or exhaust gas lines in the etching system can overcome this limit maintaining the pressure of the gas. When the sidewall can fit to a specific crystal orientation (e.g., {110} or {111} surfaces), combining anisotropic wet etching is effective for smoothening the surface. Etchants such as ethylene diamine pyrocatechol (EDP) or KOH have been tried (Ezoe *et al.* 2006, Sasaki *et al.* 2003a). For the arbitrary surfaces, hydrogen annealing can be applied to smoothen or round the surface by facilitating the migration of Si atoms (Lee *et al.* 2005).

Another linear motion consists in lifting the micromirror. **Figure 48** is a device for generating the large piston motion (Jain *et al.* 2005). Its application is the optical coherence tomography (OCT). OCT is a nondestructive imaging that produces high-resolution cross-sectional images of biological tissue. The core of this OCT system is a Michelson interferometer in which a broadband light beam is split into a reference arm and a measuring arm. The light beams from the reference mirror and the sample interfere with each other. The interference, which occurs only when the optical path difference is within the coherence length of the light source, is picked up by the photodetector. The coherence length determines the axial resolution. The axial linear motion range of the reference mirror generates cross-sectional images. An imaging depth of ~2 mm in tissue samples is typically required, so the optical delay line in the reference arm must perform axial scanning by a few millimeters.

Figure 48 shows a mirror design. The mirror plate is attached to a rigid frame supported by Al/

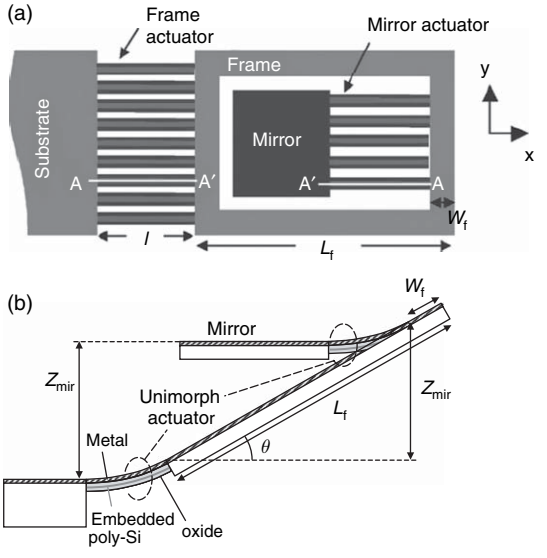


Figure 48 Schematic design of the micromirror generating a large vertical displacement: (a) top view and (b) cross-sectional schematic view of the device. (Source: Jain A, Qu H, Todd S, Xie H 2005 A thermal bimorph micromirror with large bi-directional and vertical actuation. *Sens. Actuators A Phys* **122**, 9–15.)

SiO_2 unimorph beams on the right side, where the thermal actuator is embedded. The rigid frame is connected to the substrate by another set of unimorph beams on the right side. This second set of unimorph beams actuates the frame. The mirror is actuated with two sets of unimorph actuators having an identical design. After the mirror is released, the bimorph beams curl up due to the tensile stress in the upper Al layer and the compressive stress in the lower SiO_2 layer, as illustrated in **Figure 48**. With the curls of the two sets of actuators, the tilt motions

compensate each other resulting in a zero tilt. At the initial state, the mirror plate elevates by Z_{mir}, which can be expressed by

$$Z_{mir} = (L_f - W_f)\sin\theta$$

where L_f and W_f are the length and beam width of the frame, respectively, and θ is the initial tilt angle of the frame. Because the frame dimensions are $L_f = 0.5$ mm, $W_f = 40\,\mu$m, and $\theta = 17°$, Z_{mir} is 135 μm. Because of thermal actuation, the mirror moves downward. The mirror moves vertically keeping its angular direction canceling the tilting motion of the two sets of actuator. The demonstrated piston motion is 200 μm.

3.01.5.2.3 Compound motions

Combining the rotation and the vertical movements of the mirror enables adjustment of the phase of the reflected light, which will be effective in an arrayed mirror like a grating (Krishnamoorthy *et al.* 2002). **Figure 49** shows another mirror device image (Lee *et al.* 2004). **Figure 50** is the schematic cross section illustrating the driving method and the mirror movement. Electrode 5 (the movable comb) is connected to the ground. Bidirectional rotation is achieved by applying a driving voltage on electrodes 1 and 4 or 2 and 3. Pure piston motion is achieved by applying the same voltage on electrodes 1 and 2. The static mirror demonstrates rotation of ±4.6° with 155 V driving voltage. The maximum static vertical movement is −7.8 μm with 110 V.

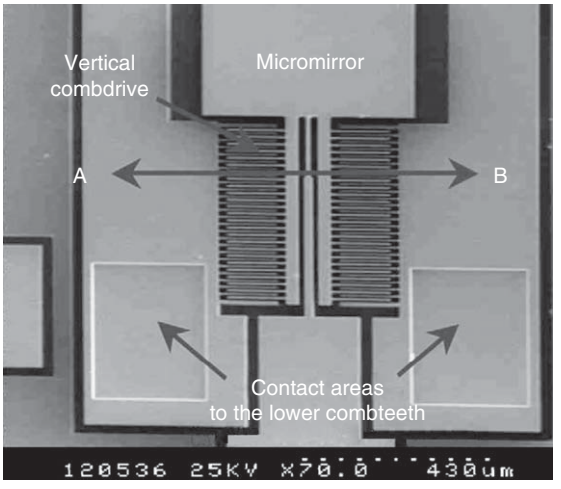


Figure 49 Image of the micromirror that can realize piston motion and rotation. (Source: Lee D, Krishnamoorthy U, Yu K, Solgaard O 2004 Single-crystalline silicon micrometers actuated by self-aligned vertical electrostatic combdrives with piston-motion and rotation capability. *Sensors and Actuators* **A114**, 423–8.)

When the accurate movement of the piston and the tilting motion of the micromirror can be realized as a large number of mirror arrays, the mirror array can work as a phase-shift mask. When the relative displacement between the neighboring mirrors is $\lambda/4$, the reflected lights interfere, canceling each other and thereby generating the dark field. This generates the higher resolution of the pattern. The maskless lithography with the high-end performance can be considered as an application (Martinsson *et al.* (2005) . The motivation for using optical maskless lithography is to decrease the mask cost, as the photomask set is prohibitively expensive for low-volume integrated circuit applications. The array of micromirrors is used as a reticle. The phase of the reflected light is controlled by shifting or tilting of the mirror. The function of the phase-shift mask is realized (Ljungblad *et al.* 2005). The reflected image from the micromirrors is projected and demagnified onto the wafer. The movement accuracy and the mirror-to-mirror uniformity are tightly required. Calibration and correction are carried out at present in the SIGMA7300 commercialized mask writer (Micronic Laser Systems AB).

The application of the tunable external cavity diode lasers requires a different mirror motion. This device will be useful in spectroscopy and especially in optical networks. Continuous wavelength tuning is important for wavelength-division-multiplexed (WDM) systems. Widely tunable lasers can be used as the sparing transponders in WDM systems. The ultimate goal is replacing fixed wavelength-distributed feedback (DFB) lasers with tunable ones (Bruce and Riezenman 2002). DFB lasers are stable but difficult to tune over a wide range. Valuable features like wavelength routing or dynamic provisioning to be incorporated into long-haul or metro-optical networks can be realized (Park *et al.* 2003). The continuous tuning of the wavelength is important, as mode hops mean the existence of an unusable wavelength. Among many tuning methods, the external cavity diode lasers (ECLs) offer advantages such as wide tuning range, high output power, and narrow line width. Although there have been tunable lasers using the bulk mechanism (Lazar *et al.* 2004), telecommunication needs small lasers with high yield. The micromirror is attractive because of its smaller size, larger shock resistance, lower power consumption, and potentially lower cost.

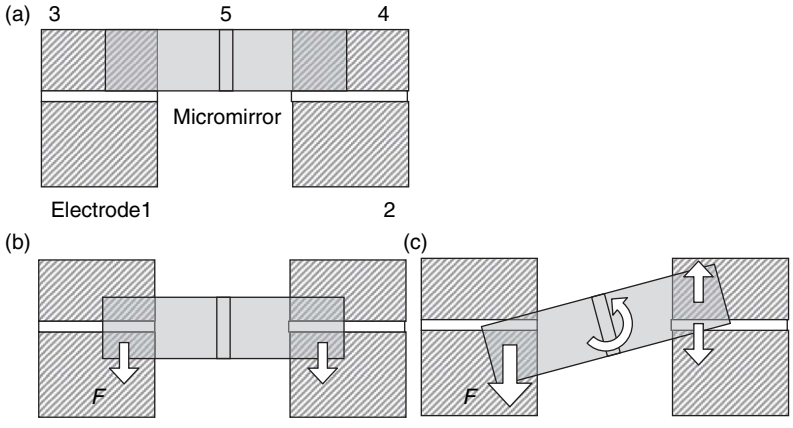


Figure 50 (a) Schematic cross section along AB (in **Figure 49**) without actuation; (b) pure piston motion is achieved by applying the same voltage to electrodes 1 and 2; and (c) electrodes 3 and 4 add actuation control in the adjustments of tilt and piston motion.

Figure 51(a) shows the schematic setup of ECL developed in Iolon Inc., which was acquired by Coherent Inc. in November 2005 (Anthon *et al.* 2002, Jerman *et al.* 2002). The Littman configuration has a folded resonator with a grating placed between the semiconductor amplifier and the resonator end mirrors. The grating is fixed and the mirror is moved. The first-order diffraction beam is incident on the resonator end mirror. The diffracted beam having the wavelength corresponding with the angle normal to the mirror is reflected back to the grating and to the amplifier. The grating provides a wavelength filter that selects a resonator mode whose exact frequency is determined by the cavity length. The zero-order reflection beam from the grating or the emission from the other end of the semiconductor amplifier become the output. The gain medium of the semiconductor amplifier is based on a high-power, 1.55-μm InGaAsP/InP multiple quantum well LD. The gain bandwidth is >80 nm, while the external cavity mode spacing is only 0.16 nm. A large number of external cavity modes are supported by the diode gain medium.

In addition to controlling the diffraction angle for the grating wavelength filter, it is necessary that the phase of the light in the cavity (the longitudinal modes of the cavity relating to the cavity length) be controlled for eliminating mode hops. For realizing wide continuous tuning, the pivot point is usually set at the specific point (near to the semiconductor amplifier). The mirror angle and cavity length are changed simultaneously so that an equal integer number of waves are included in the cavity.

Figure 51(b) shows the perspective view. The laser is tuned by applying voltage to the comb elements of the electrostatic actuator to produce an electrostatic force that rotates the mirror around a virtual pivot. The suspensions and comb fingers are not parallel to each other and have arch shapes. The angular range of the actuator is up to ±2.8° at ∼150 V, for tuning over a wavelength range of up to 42 nm. As shown in the inset photograph, the resonator end mirror is mounted on the microactuator. This mirror is prepared individually. The process for the mirror with high reflectivity can be optimized without the process restriction for combining the fabrication with the actuator. In addition to the part for mounting a mirror, there is another moving part. Two counter-rotating parts coupled by linkages through a flexural pivot give a zero-moment actuation.

Ishimori *et al.* (2007) reported another mirror design for controlling the virtual pivot position electrically. The mirror device has two parallel rotation axes for achieving mirror rotations and linear translation. **Figure 52** shows a schematic diagram of the external cavity diode laser using the mirror with two parallel rotation axes. The movable frame is suspended by torsion bars parallel to those of the mirror. To generate the translation, the micromirror is driven to the opposite rotational direction at a same angle around the two rotational axes. **Figure 53** shows the fabricated mirror. Two pairs of comb-drive actuators are included to enable an independent excitation of the actuators.

3.01.5.2.4 Mirror deformation

The deformable mirror is useful for controlling the wavefront. The dynamic focus control will realize the functions of field flattening or fine focus

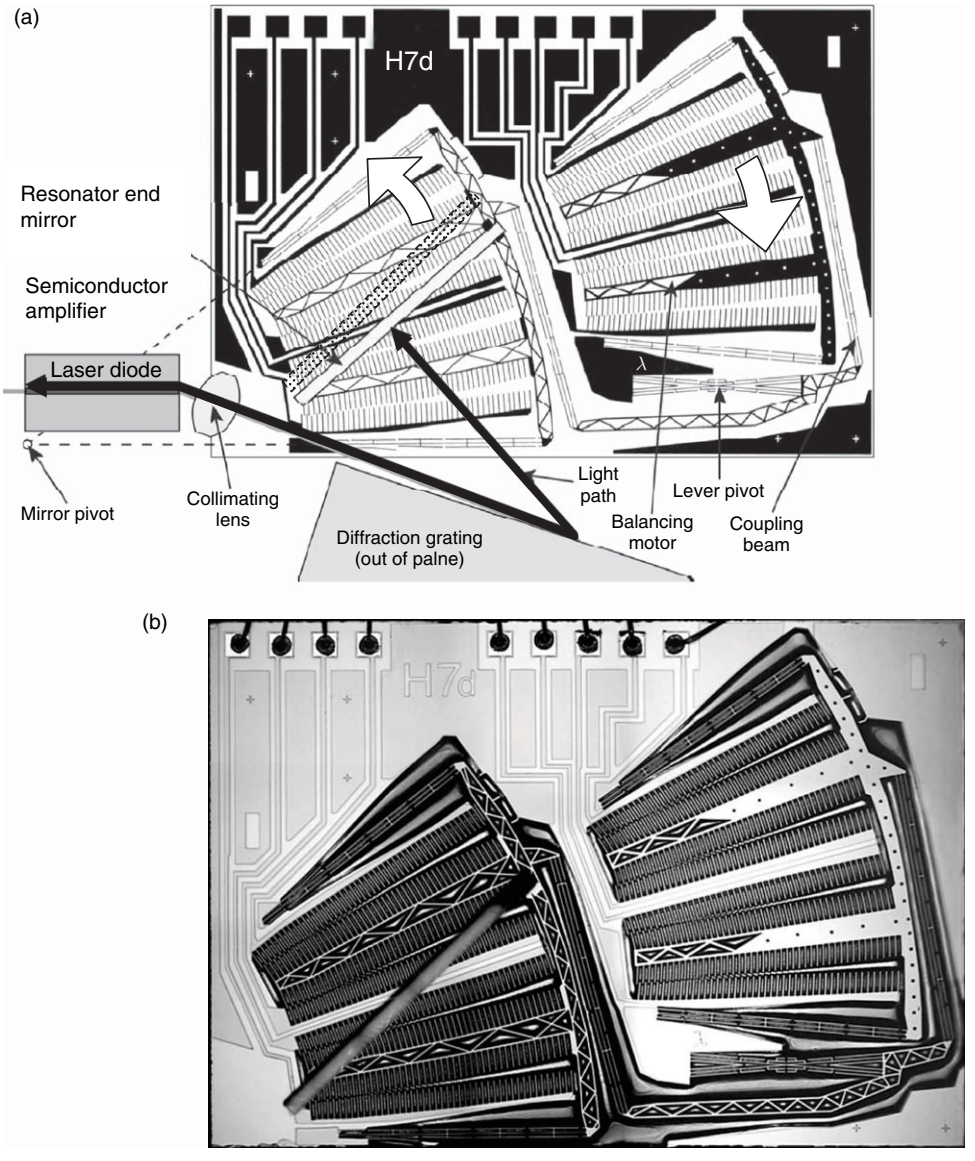


Figure 51 (a) Schematic drawing of external cavity laser diode with a balanced virtual pivot actuator. (b) Simplified perspective view with the inset photograph of actuator die after the attachment of the externally fabricated mirror. (Source: Jerman H, King D, Lee H, Tselikov A, Yasumura K 2002 External cavity diode lasers tuned with silicon MEMS. *Technical Digest, Optical Fiber Communication Conference*, Paper TUO7.)

control. **Figure 54** shows an example with variable focus membrane mirrors (Shao *et al.* 2004) integrated with a two-axis mirror. **Figure 54(a)** shows the cross-sectional view. A deformable SiN membrane is supported on a gimbal structure. **Figure 54(b)** shows the whole view of a mirror. The thickness of the center mirror is 0.9 μm. The low-stress low-pressure chemical vapor density (LPCVD) nitride membrane is used. A 100-nm chrome–gold layer on top of the membrane serves as the mirror surface.

Figure 55 shows the line profiles of the deformable membrane that realizes the flat surface. With no applied actuation voltage, the maximum surface fluctuation is <90 nm peak to peak across the membrane. The solid lines are the measured surface profile and the dotted lines are the parabolic curve fits. The depth of the air gap between the membrane and the Si substrate is ~12 μm. The center displacement is 3.5 μm at 240 V. The effective focal length f_{mir} of the mirror is given by

$$f_{mir} = D^2/16\delta \qquad [25]$$

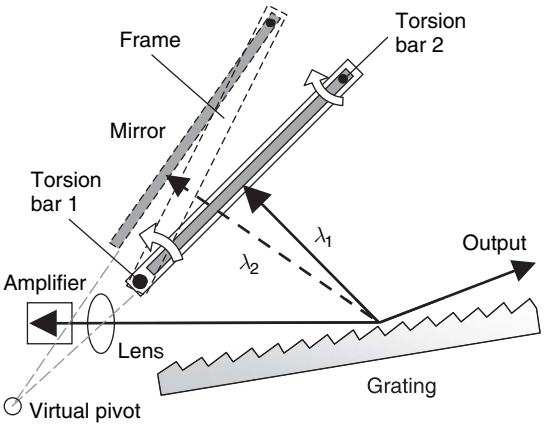


Figure 52 Illustration of the external cavity diode laser having Littman configuration and the micromirror movement.

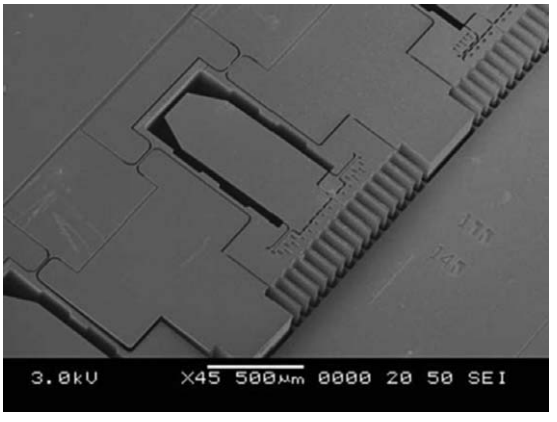


Figure 53 Fabricated micromirror with parallel two rotational axes.

where D (=700 µm) is the diameter of the mirror and δ is the center displacement. The demonstrated range of the focal lengths is from $f_{mir} = \infty$ with no displacement to $f_{mir} = 8.8$ mm with $\delta = 3.5$ µm.

The region of adaptive optics deals with the control of the further complicated wavefront. The final purpose is to increase the resolution of optical systems and to enhance fuzzy images by active compensation of phase aberrations. Aberrations are caused by imperfections of elements (e.g., mirrors deformed by mechanical or thermal stresses) and external conditions (e.g., turbulence of the ambient media). The active compensation of aberrations becomes important for large-aperture high-resolution systems, and for laser systems sensitive to energy losses. The principles of resolution enhancement have been applied successfully to large-aperture ground-based telescope systems, where the high cost of deformable mirror can be allowed. Passing through the atmosphere, the light wave is affected by fluctuations in density of the atmosphere due to turbulence, distorting the front of the wave. A blurred image is produced. To compensate for this distortion, the wavefront sensor and the deformable mirror are introduced. One inhibiting factor is the expense of conventionally manufactured deformable mirrors, which typically cost about $1000 per actuator (Xinetics Inc., according to Doble *et al.* 2002). Microfabrication can increase the number of actuators, decreasing the cost per actuator. Many applications (e.g., retinal imaging in ophthalmology, confocal microscopy, ultrashort laser pulse shaping (Bartels *et al.* 2000), and machine vision) can be realistic.

A group from Boston University (Bifano *et al.* 1997, Mali *et al.* 1997) reported a single compliant poly-Si membrane supported by multiple attachments to an

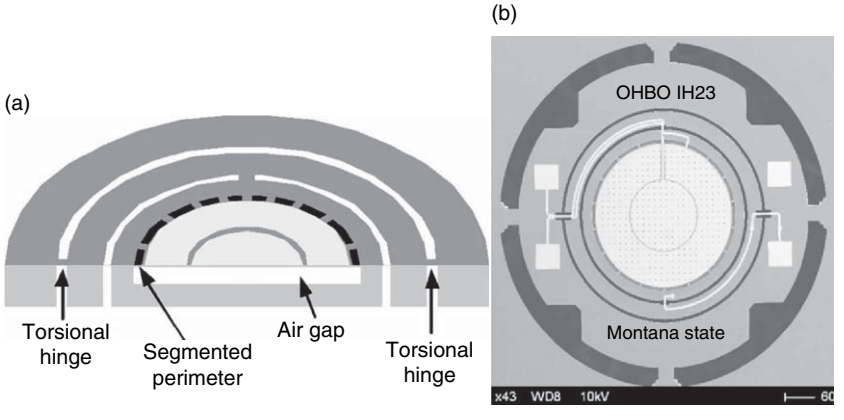


Figure 54 (a) Schematic cross-sectional view of the deformable mirror. (b) Scanning electron micrograph (SEM) of the deformable mirror. (Source: Shao Y, Dickensheets D L, Himmer P 2004 3-D MOEMS mirror for laser beam pointing and focus control. *IEEE J. Sel. Top. Quant. Electron.* **10**, 528–35.)

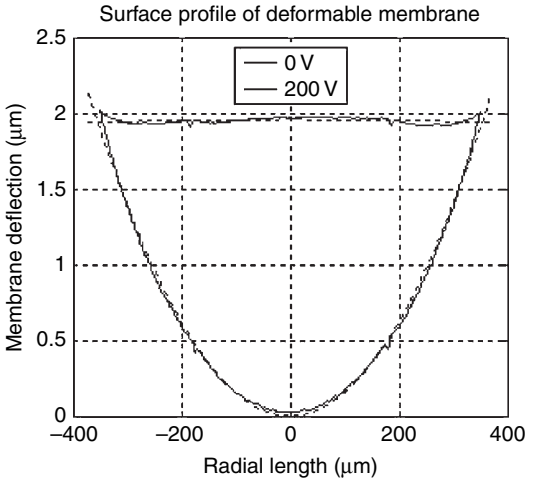


Figure 55 Surface line profiles (solid lines) of the deformable membrane with actuation voltages at 0 and 200 V. The dotted lines are parabolic fits to the data. (Source: Shao Y, Dickensheets D L, Himmer P 2004 3-D MOEMS mirror for laser beam pointing and focus control. *IEEE J. Sel. Top. Quant. Electron.* **10**, 528–35.)

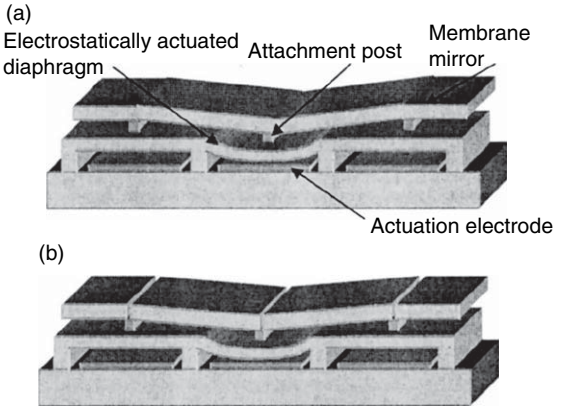


Figure 56 Cross-sectional schematics of deformable mirror array: (a) continuous and (b) segmented mirrors for stress relief having tip-tilt motion (not to scale). (Source: Perreault J A, Bifano T G, Levine B M, Horenstein M N 2002 Adaptive optic correction using microelectromechanical deformable mirrors. *Opt. Eng.* **41(3)**, 561–6.

underlying array of electrostatic actuators as shown in Figure 56. The micromirror based on the surface micromachining is now manufactured by Boston Micromachines Corporation. The mirror consists of 10×10 arrays of electrostatic actuators with the spacing of 300 μm supporting a thin-film continuous or semicontinuous mirror membrane via posts. Two types of deformable mirrors are reported. Figure 56(a) is a continuous membrane mirror with a fill factor of 99.7%. Figure 56(b) shows a continuous membrane with interpost cuts introduced into the mirror surface in order to relieve stress in the mirror membrane. The fill factor is 98.6%.

A group from Delft University reported a deformable mirror display consisting of a 10 mm × 10 mm SiN membrane fixed on another wafer with a 2D array of electrodes (Vdovin and Middelhoek 1995, Vdovin and Sarro 1994). OKO Technologies now produces membrane mirrors. The membrane is covered by Al and electrostatically deflected at the location of individual pixels. Figure 57 shows the schematic drawing. The membrane is thin (0.5–10 μm thick) and can have diameters of up to 50 mm. Voltages applied between the membrane and the individual electrode result in the deformation of the membrane. Combinations of voltages applied to different electrodes can form specific shapes on the membrane surface. Microfabrication realizes many electrodes with high density. The number of electrodes and its area are balanced with the amplitude of response of a single individual electrode. The feature of zero-hysteresis actuation enables the feed-forward control after preliminary calibration. Figure 58 shows a typical view of the deformable mirror.

An application of retinal imaging is introduced here. The micromirror designed by Boston Micromachines is applied for the human eye (Doble *et al.* 2002). The wavefront is corrected by modifying the optical phase profile. A wavefront sensor is coupled with the deformable mirror in a feedback loop. The wavefront is flattened with the conjugated phase profile controlling the optical path difference controlled by the mirror deformation. A 6.8-mm pupil is corrected for retinal imaging, and the central 6-mm pupil is used for retinal imaging. This eliminates edge effects that might disturb the wavefront. Figure 59 shows the retinal mosaic images before and after the correction, taken at 1° eccentricity. The field of view is 0.25°, corresponding to 75 μm on the retina. On examining the image power spectra, there was a fourfold increase in the relative power at the cone spatial frequency.

3.01.6 Applications and Specifications

Here, some products are overviewed, analyzing specifications and showing chosen technical solutions. This section aims in clarifying the advantages as well as the disadvantages of the micromirrors to

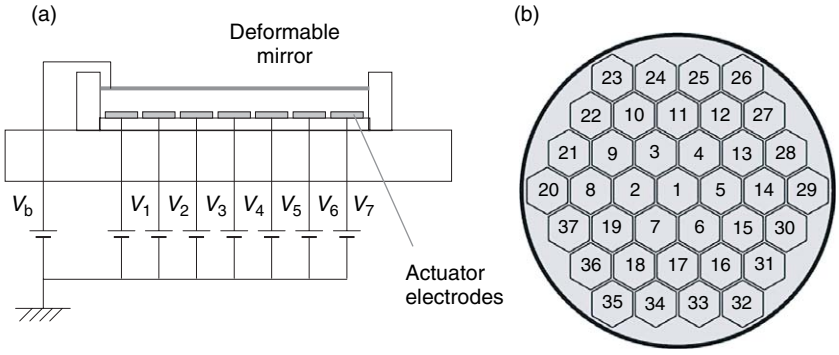


Figure 57 (a) Schematic drawing of deformable mirror and (b) the layout of driving electrodes. (Photo courtesy: Flexible Optical B.V. (aka OKO Technologies).)

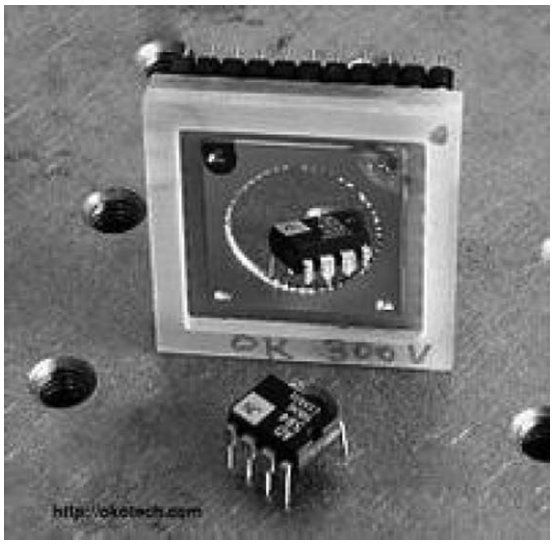


Figure 58 Typical view of micromachined deformable mirror. (Photo courtesy: Flexible Optical B.V. (aka OKO Technologies). http://www.okotech.com/mirrors/small.)

reach specifications. At first, the bulk mirror devices will be discussed as reference.

3.01.6.1 Bulk Mirror Devices

3.01.6.1.1 Polygon mirror

Figure 60 shows the typical polygon mirror. Applications include printer, facsimile, and copying. Table 3 lists the specifications of Sanyo LS7230 or LS3440. The actuator is the brushless motor driving at 24 V for example. The rotating motion of the motor stabilizes the scanning speed. Let us assume that for a hexagonal prism, the optical scan is six times of the motor rotation speed. Optical rotation speed can reach several thousand revolutions per second. The frequency corresponds to 2–4 kHz. Because one side of the surface of the hexagon generates a rotation of 60°, the optical scan angle of 120° is the maximum value. Compared to the polygon mirrors, the micromirror gives the reflecting surface

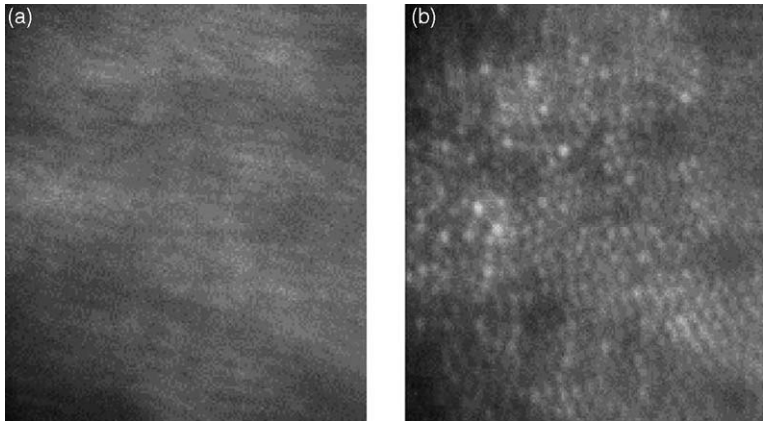


Figure 59 Retinal mosaic images before (a) and after (b) correction. (Source: Doble N, Yoon G, Chen L, Bierden P, Singer B, Olivier S, Williams D R 2002 Use of a microelectromechanical mirror for adaptive optics in the human eye. *Opt. Lett.* **27**, 1537–9.)

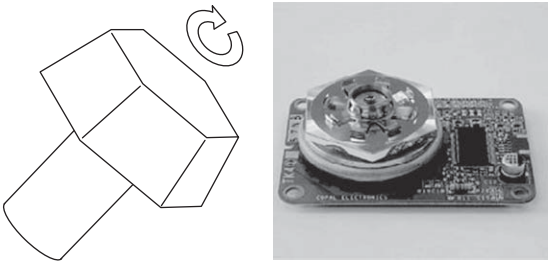


Figure 60 Examples of polygon mirror. (Photo courtesy: Nidec Copal Electronics Corp. http://www.copal-electronics.com/j/index.html.)

Table 3 Specifications of a typical polygon mirror

Voltage	24 V
Current	<700–850 mA
Rotation	370–580 rps
Actuator	Brushless motor

Source: Catalogue from SANYO Seimitsu Co., Ltd.)

placed at the same level with the rotation axis, allowing a steady reflecting point in geometry. The resonant frequency on the order of a few kilohertz is achievable. The scanning angle of the micromirror is relatively smaller.

3.01.6.1.2 Galvanometric mirror

Figure 61 shows galvanometric mirrors of another popular bulk mirror. These mirrors have the axis of rotation lying on the plane of the reflecting surface. The applications are, for example, optical processing (like rapid prototyping), machining of laser ablation or welding, laser marking or patterning, and optical metrology. For creating 2D scanning, two galvanometric mirrors are usually used. The biaxial micromirror can reduce the size simplifying the system. Table 4 lists typical specifications. The scanning frequency is relatively low. The settling time is ~1 ms. The rotation angle is about ±10. The mirror is exchangeable, and its size is on the order of 10 mm having a thickness of 3–7 mm. The tilt platforms are equipped with rotary encoder, capacitance, or strain gauge sensors. The rotation angle of the mirror is measured enabling the closed-loop control generating high-accuracy and repeatability scans. The resolution is, for example, 0.33 μrad, and the repeating accuracy is 10 μrad (0.0006°).

Figure 61 Example of a galvanometric scanner. (Photo courtesy: GSI Group Inc. http://www.gsilj.com/products/m-series.html.)

3.01.6.2 Scanning Performance of Micromirror

Consider the micromirror having a typical gimbal design as shown in Figure 62. Independent of the driving method, the resonant frequency is decided by the balance between the spring constant of the torsion bars and the moment of inertia of the micromirror. The rotational spring constant k_θ of two torsion bars having a rectangular cross section is expressed as follows:

$$k_\theta \approx 2\frac{Gwt^3}{3l}\left\{1-\frac{192}{\pi^5}\frac{t}{w}\tanh\left(\frac{\pi w}{2t}\right)\right\} \qquad [26]$$

where G is the shear modulus of the torsion bar, l is the length of the torsion bar, and w and t are the width and the thickness of torsion bars, respectively, supposing the relation $t < w$. The moment of inertia of the micromirror can be expressed as follows. Here, the inertia of the torsion bar is neglected as the torsion bar is usually small compared to the mirror.

$$I \approx 2\int_0^L \rho x^2 wtW\,\mathrm{d}x = \frac{2}{3}\rho wtWL^3$$

where ρ is the density of the mirror material and W and L are the width and half-length of the mirror. The mirror is supposed to be filled with uniform material. Using these relations, the resonant frequency f_r can be expressed as follows:

$$f_r \approx \frac{1}{2\pi}\sqrt{\frac{k_\theta}{I}} = \frac{1}{2\pi}\sqrt{\frac{Gwt^2}{\rho WlL^3}}\sqrt{1-\frac{192}{\pi^5}\frac{t}{w}\tanh\left(\frac{\pi w}{2t}\right)} \qquad [27]$$

Table 4 Specifications of a typical galvanometric mirror

	GSI General Scanning G series	*Canon FS-1M28*
Mirror diameter (mm)	9–50	~30
Power source (V)	±15 to ±24	5, 24
Moving range (deg)	±30	±10
Motor	Moving magnet	Brushless
Scanning speed	>1 kHz (bandwidth)	100 Hz
Repeatability	0.0006° capacitance detection-type sensor	0.0003° pp encoder and servo-control

Source: Data sheet from GSI Group Inc., Catalogue from Canon Marketing Japan Inc.

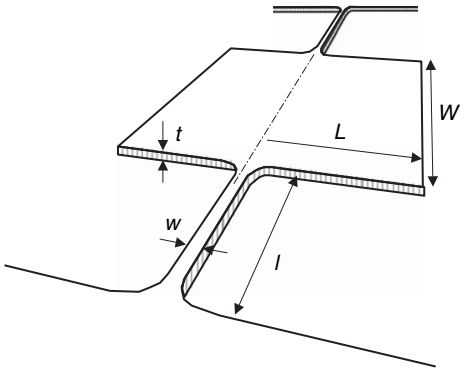


Figure 62 Schematic drawing of the typical micromirror for discussion.

Because the material used is usually Si, the resonant frequency is decided by the sizes of the torsion bar and the mirror. For increasing the resonant frequency, the larger spring constant k_θ and the smaller inertia I are necessary.

According to eqn [7], $x' = (\lambda f/2a)$, the spot diameter is $2x'$. The mirror diameter is necessary for obtaining the smaller spot. If $f = 10$ cm, $\lambda = 500$ nm, and the aiming spot size is 50 μm (corresponding to 508 dpi spot size), the mirror diameter has to be >1 mm. The larger resonant frequency for faster scanning has the trade-off relation with higher resolution due to the increase of inertia (Urey *et al.* 1999a).

3.01.6.3 Confocal Microscope

Olympus Corp. uses two scanners in their confocal microscope, which illuminates the specimen with a small optical spot. The image is obtained by scanning a spot two-dimensionally. **Figure 63** shows the optical setup. One micromirror is for vertical and the

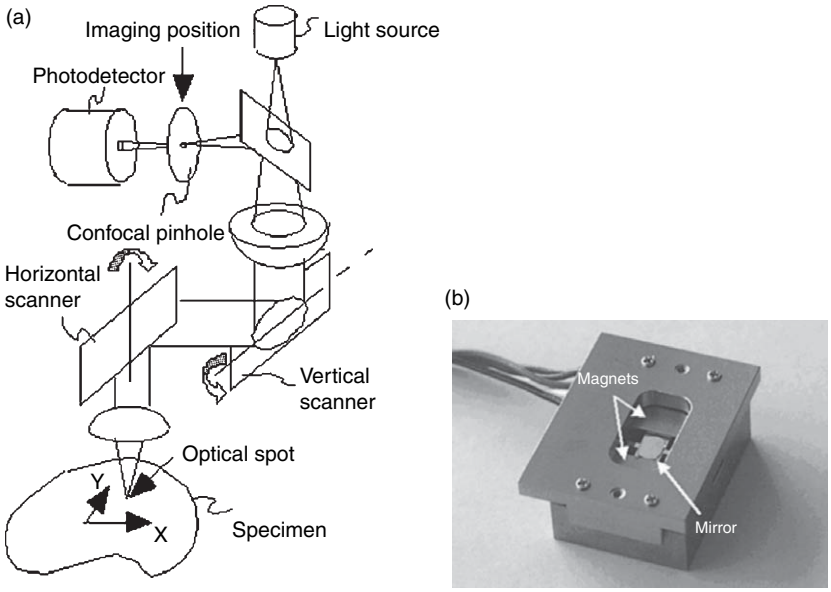


Figure 63 (a) Optical setup of confocal microscope and (b) photograph of the assembled scanner using micromirror. (Source: Miyajima H, Murakami K, Katashiro M 2004 MEMS optical scanners for microscopes. *IEEE J. Selected Topics in Quantum Electronics*, **10(3)**, 514–27.)

other is for horizontal scanning. The horizontal scanning is the fast scan. The micromirror is designed so that it can be installed into the existing product. **Table 5** shows the requirements for the horizontal scanner. The mirror size is >4.2 mm × 3.0 mm because the millimeter order optical beam size is used in the conventional microscope. The scan angle should be interchangeable between 2.1° and 16° to cover one to six times the optical zoom. Although higher resonant frequency is better for fast image acquisition, the scanning speed of ~4 kHz is still a moderate task.

A mirror thickness of 300 μm is required for maintaining its flatness against the residual stress of the coil films (7-μm-thick Cu-electroplated driving coil) on the mirror plate and dynamic deformation (described later). The development of a horizontal scanner (fast scanning) starts from the electromagnetic scanner using polyimide hingers (Miyajima *et al.* 2001). Later, the polyimide hinge is replaced by the c-Si hinge (Miyajima *et al.* 2003). The trade-off of the higher resonant frequency and sufficient scan angle (inversely proportional to the hinge stiffness k_θ in eqn [27]) is solved with the higher Q factor obtained by the c-Si hinge. The energy loss can be reduced. The scanner is equipped in the confocal microscope OLS1200 (renamed from OLS1100 as the micromirror is installed).

Table 5 Specifications of horizontal scanner for the confocal microscope

Mirror size (mm)	>4.2 × 3.0
Resonant frequency (Hz)	3900–4100
Scan angle (deg)	2.1–16 (changeable)
Scan angle fluctuation (%)	<0.1
Mirror flatness (nm peak-to-valley)	<244
Mirror reflectivity (%)	>85 ($\lambda = 488-633$ nm)

Source: Miyajima H, Asaoka N, Isokawa T, Ogata M, Aoki Y, **Imai** M, Fujimori O, Katashiro M, Matsumoto K 2003 A MEMS electromagnetic optical scanner for a commercial confocal laser scanning microscope. *J. Microelectromechanical Systems* **12(3)**, 243–51.

3.01.6.4 Laser Display

The laser display is a natural idea as an application of the micromirror and has been researched widely. The bright light spot is two-dimensionally scanned on a screen to create an image. Raster scanning systems combined with a modulating light source generates a 2D image. The advantages of this architecture include the absence of pixel boundaries, the ability to synchronize the system to a wide ratio of video input signals, and the ability to reformat image aspect ratios. **Table 6** lists the requirements for various video formats. The horizontal line rate is for fast scan. This value somewhat depends on the optical system supposed (Urey 2004). The vertical mirror motion (slow scan) is required to function as a linear ramp. The vertical drive must provide a sufficient torque to allow nonresonant actuation. Note that the resolution of SVC micromirror developed by Conant *et al.* is still 356 pixels, satisfying only the quarter video graphics array (QVGA) format. High-quality extended graphics array (XGA) or high-definition television (HDTV) displays require a higher resolution. The resolution and higher line rate are related to the trade-off. The resolution can be increased only by increasing the size of the mirror (assuming flat mirror surface). As seen from eqn [27], the resonant frequency f_r is proportional to $L^{-3/2}$. θ_{max} is the maximum mechanical scan angle amplitude (from center to peak).

The number of resolvable spots limited by diffraction depends on the relation $\theta_{max}D$, which is the product of scan angle and mirror size. The mirror size D is for overall size (e.g., diameter). This is a defining parameter for resolution in scanning displays. Eqn [20] can be rewritten as follows:

$$\theta_{max} 2a = \frac{NC\lambda}{2} \qquad [28]$$

This is the static relation. For the dynamic operation, the overall scanning angle can be used increasing the number of resolution by double (Urey 1999).

$$N = \frac{4\theta_{max} D}{C\lambda} \qquad [29]$$

Table 6 Requirements to scanners for realizing various display formats

Display format	*QVGA*	*VGA*	*SVGA*	*XGA*	*SXGA*	*UXGA*	*HDTV*
N_h, horizontal resolution	320	640	800	1024	1280	1600	1920
N_v, vertical resolution	240	480	600	768	1024	1200	1080
$\theta_{max}D$ (deg mm)	3.9	7.8	9.7	12.4	15.6	19.4	23.3
f_s, horizontal scanner frequency (kHz)	8	16	20	25.6	34	40	36

Source: Urey H 2004 MEMS scanners for display and imaging applications. *Proc. SPIE* **5604**, 218–29.

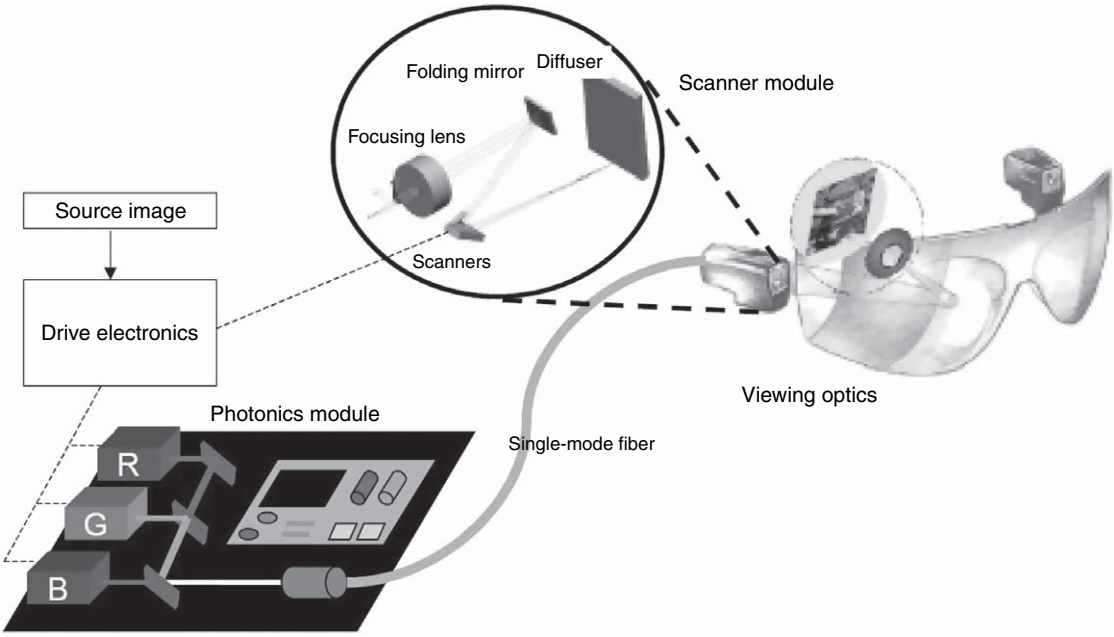


Figure 64 Conceptual drawing of a head-worn display. (Source: Urey H, Nestorovic N, Ng B, Gross A 1999b Optics designs and system MTF for laser scanning displays. *Proc. SPIE* **3689**, 238–48.)

The microscanner is considered to have the potential of opening new markets with regard to some new systems, where the miniaturization advantage matches with the application. An example is a head-worn display where the light beams are scanned directly over the viewer's vision. **Figure 64** shows the conceptual drawing. The beam from the light source is focused onto an intermediate image plane. The position of this focus point is determined by the angle of the scanning mirror, which is placed between the collection optics and the intermediate image plane. Scanning the mirror two-dimensionally, an image is created onto the intermediate image plane. This image is relayed onto the viewer's eye using a set of relay optics. The lens of the eye then focuses the image onto the retina. The net result is that the light from the light source is focused onto the retina of the viewer's eye.

Such virtual retinal display has been explored at the human interface technology laboratory of the University of Washington and Microvision Inc. is currentlyattempting to commercialize it by eliminating the bulky power-hungry panel monitor. The system suits the low-power requirements of mobile applications. One or two micromirrors are required to produce the entire image. Tiny semiconductor lasers for RGB (red, green, and blue) colors are used mixing colors. The display hardware is miniaturized and fits onto an eyeglass frame. This approach has a see-through capability, i.e., the displayed image can be superimposed on a bright external scene. This is advantageous for cockpit displays and in industrial and medical applications. According to Microvision's report, the systems are moving into several industries, including automotive service, to help service technicians keep track of the huge and ever-changing reams of repair data and display it precisely where and when they need it – in the service bay, while they are working on a car. This first-generation system was introduced to auto dealers in early 2004. Technicians may perform repair procedures in less time. The system is built around a lightweight display mounted on a baseball cap or visor. In the current version, a wireless computer with a touch pad control is worn on the belt.

Figure 65 shows the 2005 version of the biaxial microscanner with a mechanical layout (Sprague *et al.* 2005). The electromagnetic driving is adopted. The biaxial scanner architecture eliminates the need for a second scanner and makes the device more compact. The center mirror does not have a coil for escaping the stress, which degrades the mirror flatness.

The biaxial operation is based on superimposing the drive torques for the two scanning directions and apply this torque at 45° relative to the two scan axes, as shown in **Figure 66**. It is possible to excite both scan axes with only one applied drive mechanism. The

Figure 65 The biaxial microscanner. (Photo courtesy: Microvision Inc. http://www.microvision.com.)

scanner mechanism acts as a mechanical filter, separating the superimposed drive torque into its component parts. The frequency response characteristics of the mirror separate the torque components into their respective modal motions. The resonant response modes used are shown in **Figure 67**. Mode 1 is the vertical resonant mode (**Figure 67(a)**) and mode 6 is the horizontal mirror resonant mode (**Figure 67(b)**). Mode 6 includes a counterrotation of the gimbal ring with respect to the mirror. The high Q of mode 6 amplifies the 20-kHz motion to produce high angular deflection. This mode does not respond significantly to the low-frequency components for the vertical scan, so the resulting scan motion is pure sinusoidal motion at the resonant frequency. The frequency components of the applied torque are selected to excite the horizontal mirror resonance (20 kHz). The vertical axis is driven nonresonantly in a sawtooth pattern at the desired display frame rate (e.g., 60 Hz).

3.01.6.5 Digital Micromirror Devices

DMD was invented in 1987 by Dr. L Hornbeck of Texas Instruments (Van Kesel *et al.* 1998). Compared to the scanner driving a single mirror, DMD adopts the method using a large number of micromirrors (integration in number). The number of resolvable spots is decided by the number of mirrors in the array. The factor, which decides the resolution, is different from the case of the laser display. **Figure 68(a)** shows the schematic drawing for explaining image formation. One reflective mirror corresponds to one pixel. Each image of the mirror is projected on the screen using the lens. **Figure 68(b)** shows the case of the laser display. By preparing mirrors that are as many as the number of pixels for the image, the task of each mirror decreases.

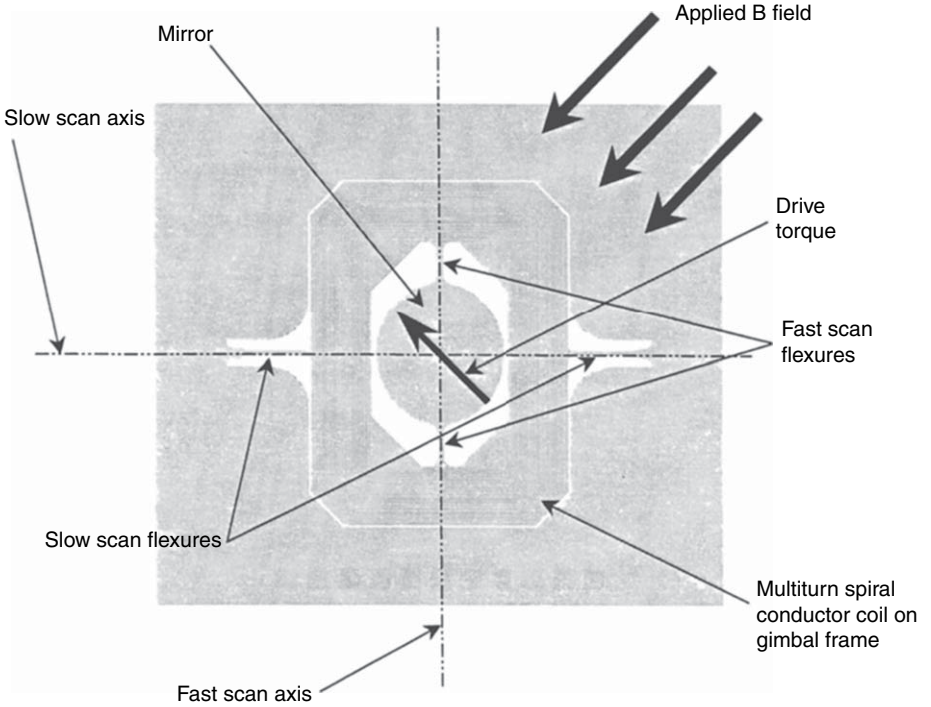


Figure 66 Schematic drawing for explaining the bi-axial operation based on superimposing drive torques for the two scanning directions. (Source: Sprague R, Montague T, Brown D 2005 Bi-axial magnetic drive for scanned beam display mirrors. *Proc. SPIE* **5721**, 1–13.)

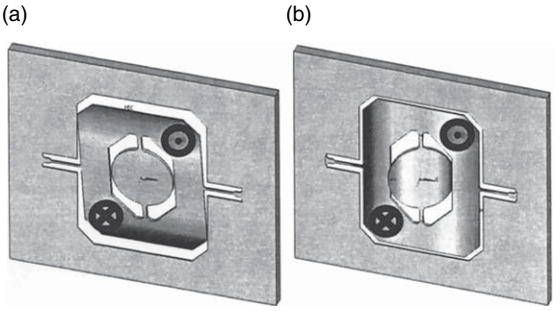


Figure 67 Modal responses of the biaxial microscanner against the torque at 45° relative to two scan axes. (Source: Sprague R, Montague T, Brown D 2005 Bi-axial magnetic drive for scanned beam display mirrors. *Proc. SPIE* **5721**, 1–13.)

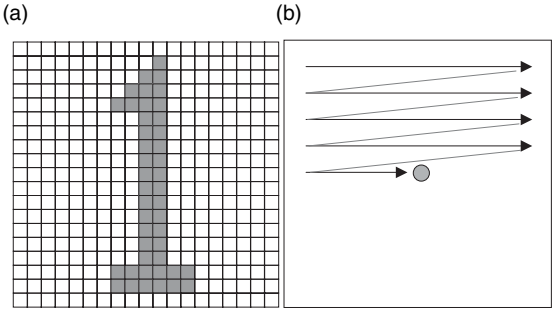


Figure 68 Schematic drawing for illustrating two methods of image generation on a screen using (a) pixel array and (b) raster scanning of the mirror.

The image generation proceeds in parallel. In 1996, 0.7SVGA 800 × 600 mirror array started to be shipped. 0.9SXGA 1280 × 1024, SXGA+ 1400 × 1050, and 1080p 1920 × 1080 mirror arrays were shipped in 2001, 2004, and 2005, respectively.

Figure 69 shows the schematic view of DMD illustrating the two mirrors (Rai-Choudhury 2000). The pitch of the micromirror array is 17 μm (decreased to 13.8 μm in early 2000). Micromirrors are mounted on the static random access memory (SRAM) circuits. This results in the integration of an optical element of the mirror and the electrical circuit of SRAM. The fabrication process is based on that for the established standard complementary metal oxide semiconductor (CMOS) circuit. The change of the process is minimized. A1 structural layers are deposited over organic sacrificial layers. The lower layer of CMOS electronics produces the electrostatic field, which acts on the yoke of the middle mechanical layer. The mirror can tilt along the hinge in the diagonal direction, through electrostatic attraction produced at the air gap under the mirror and on top of the CMOS. The driving method uses the

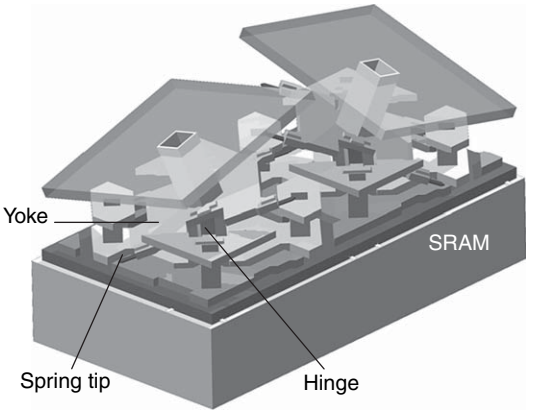


Figure 69 Cutaway view of Digital Micromirror Devices (DMD) showing two mirrors in opposite tilt states. (Photo courtesy: Texas Instruments.)

bounding from the touchdown situation. Mirrors are independently adjustable and can deflect through a tristable range of motion (fundamentally ±10° for the optical function and 0° for the reliability).

The on/off digital operation is the unique characteristic of this device. The use of this operation closely relates to the accuracy of the mirror rotation angle. **Figure 70** shows the optical setup for projecting the image. The illumination light is incident from the right side at around ∼20° against the normal axis of the device. The reflection light from the mirror is on/off controlled by the angle of the mirror. The mirror modulates the light by deflecting the beam in and out of the entrance pupil of the projection lens. Each individual mirror is used in a digital on/off mode. When the mirror is tilted clockwise (in **Figure 70**), the light is in the on state. The reflected light passes through a lens. The mirror image is projected on a screen. When the mirror is in the off state tilted counterclockwise, the light goes into a beam dump. The small mirror size contributes to the fast response decreasing the inertia. This enables time division multiplexing for generating the gray scale of the color. The digital light pulse becomes the analog mixing signal, as the human eye cannot detect the change faster than ∼60 Hz modulation. The response time of the micromirror is <5 μs (>200 kHz). Controlling the duty ratio of the on/off stages, gray scales can be created digitally. The accuracy of the mirror angle does not become an issue.

The color image is accomplished either by rotating the color wheel with one DMD chip (e.g., the

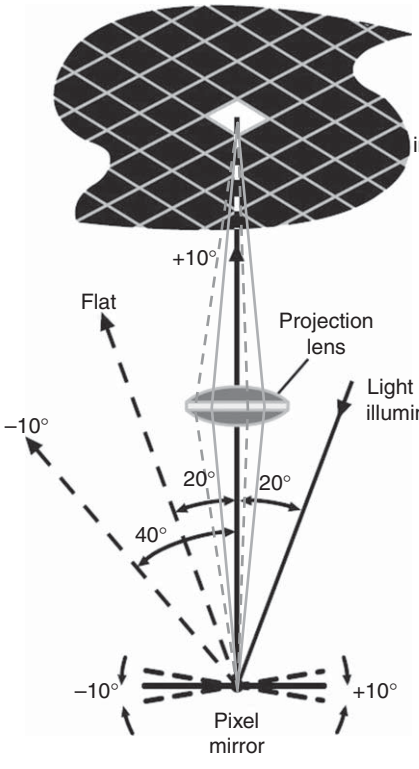


Figure 70 Schematic drawing for illustrating on/off switching action of the mirror. (Source: Van Kesel P F, Hornbeck L J, Meier. R E, Douglass M R 1998 A MEMS-based projection display. *Proc. IEEE* **86(8)**, 1687–704.)

portable system as shown in **Figure 71**) or by using dichroic mirrors and combining the prisms with three DMD chips for the RGB colors. Incoming video signals are turned into a digital code of binary (0 or 1) signals. DMD mirrors tilt accordingly. By passing the accurately modulated light through the projection lens with digital signal processing, a full color image is projected on to the screen. This time division multiplexing for each pixel permits digital creation of color images. The accuracy of the rotation angle of the mirror is greatly relaxed when compared to the analog operation.

In the literature of Texas Instruments, there are detailed discussions on hinge memory (Douglass 2003). **Figure 72** shows the image of micromirrors with hinge memory. Each mirror shows the minute different tilt angle. The first row at the bottom of the image is in the normal flat unbiased state. The device shown in **Figure 72** is still available and used normally. For decreasing hinge memory, a 0° (resting) position is required during the operation. The accuracy of the rotation angle is not guaranteed in each micromirror. Although the exact value of the accuracy of the tilt angle is not found from the documents, the fluctuation of the mirror angle will be a significant percentage of the total mirror rotation ($\pm 10°$ or $\pm 12°$). The tolerance for the accuracy can be balanced between the outside optical setup. When the total flux of the reflected light passes the lens, the intensity of the pixel image is the same and is independent of the rotational angle of the mirror. The dashed gray line in **Figure 70** shows the light path reflected from the mirror with small fluctuation of the image brightness. Unless the pupil clips the light, the image brightness is maintained. **Figure 73** illustrates the conceptual margin until the destruction. The optical setup can give the margin to the mirror rotation angle. Taking the advantage of digital processing, the mirror is open-controlled without monitoring the mirror rotation or the image intensity.

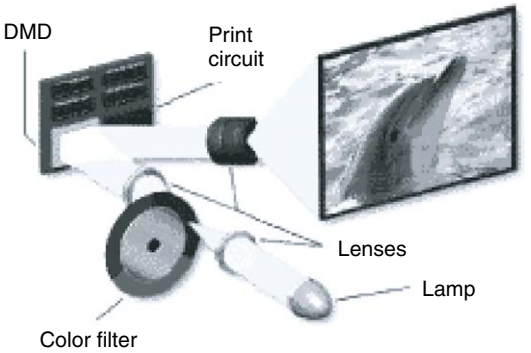


Figure 71 Optical setup of a front projection system using single Digital Micromirror Devices (DMD) chip and color filter wheel. (Image courtesy: Texas Instruments.)

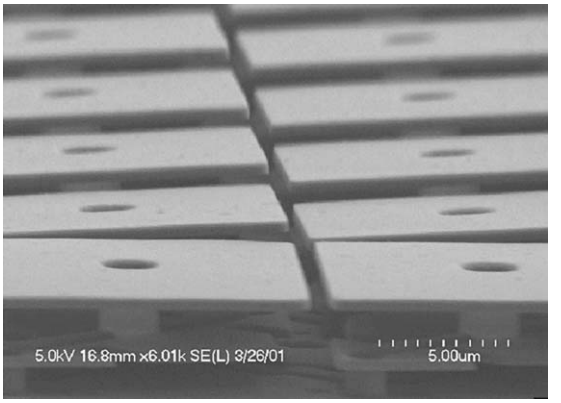


Figure 72 Scanning electron microscope (SEM) image of mirrors exhibiting hinge memory. (Photo courtesy: Texas Instruments.)

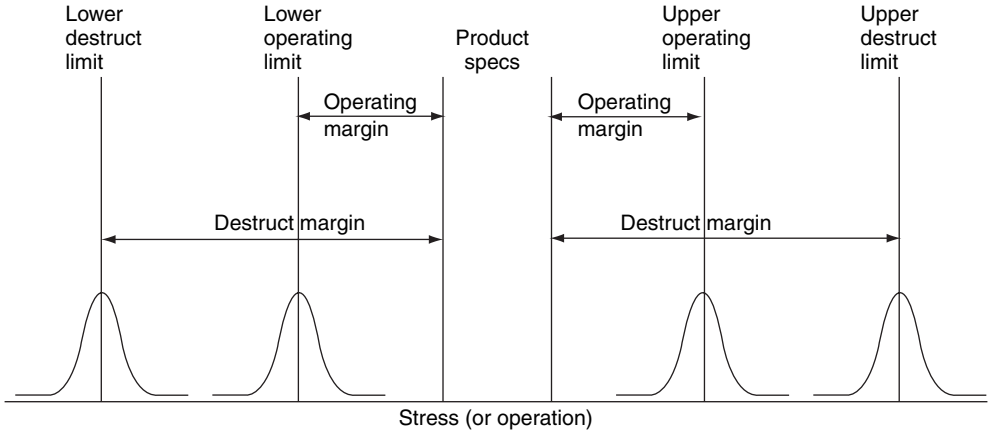


Figure 73 Illustration for explaining the stress (or operation) margin until the destruction. The operation margin relates to the allowable fluctuation of the mirror rotation angle. (Source: Douglass M R 2003 DMD reliability: A MEMS success story. *Proc. SPIE* **4980**, 1–11.)

Time division multiplexing contrast with the continuous additive color mixing adopted in the liquid crystal (LC) panel, which is the competitor in the projector market. A typical LC panel had a response time of 5–8 ms in 2006. Due to this relatively slow time response, LCs have to use the continuous additive color mixing method for generating color (spatial superposition). The preparation of the three RGB coloring segments for each image pixel becomes necessary. DMD can prepare one pixel using a single mirror.

Due to mechanical light modulation, the contrast can be larger than 1:1000. The LC display is under improvement for realizing the contrast 1:1000, but there is a trade-off between the response times. **Figure 74** shows the optical setups of the projector.

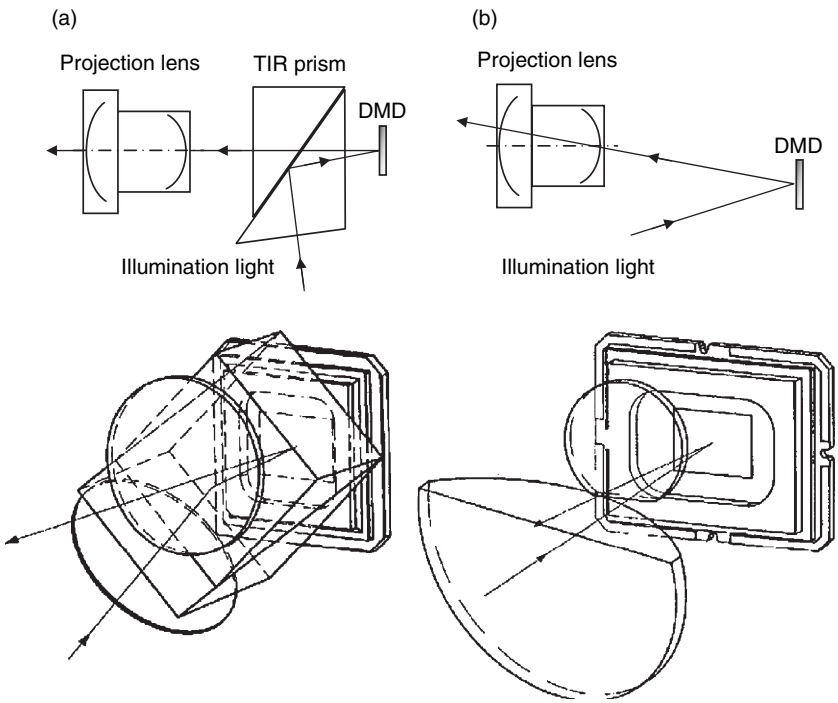


Figure 74 Schematic and perspective optical setup using (a) total internal reflection (TIR) and (b) off-axis illuminations to Digital Micromirror Devices (DMD) chip. (Image courtesy: Texas Instruments.)

The interface reflection has to be at least <0.1% for obtaining the contrast of 1:1000. The number of the optical elements is minimized for reducing the reflection at interfaces. The reflection stray light can be decreased. The biggest market is the lightweight portable projector and ~95% of projectors in the market fall within this category. Currently, the mainstream projection technology is the one-chip setup using the off-axis setup (as shown in **Figure 74(b)**). The space over the DMD element is crowded. When the lenses become larger, the projection lens cannot be set at the front of the DMD. This means that the surface plane of DMD cannot be set at the exact imaging plane of the projection lens. Using a total internal reflection (TIR) prism can be the solution (as shown in **Figure 74(a)**). However, the contrast has to be maintained by adopting high-quality antireflection coating. The theater system requiring high brightness uses this setup with three chips for high-end quality.

3.01.6.6 Grating Light Valve

The small and lightweight moving element is advantageous for increasing the resonant frequency as seen in the case of DMD. Grating light valve (GLV) takes further advantage of miniaturization. In this device, the moving element is a part of the mirror, decreasing the size. Around 1991, Professor D Bloom (Stanford University) considered the basic idea. The first paper appeared in 1992 (Solgaard *et al.* 1992). He started the venture company of Echelle in 1994. This company changed its name to Silicon Light Machine Inc. in 1996. GLV device has been considered to be a part of the display market.

GLV is a device that acts as a dynamic, tunable grating for varying the amount of laser light that is diffracted or reflected. **Figure 75** illustrates the principle showing the cross-sectional view. One reflecting mirror is divided into some ribbons, which is the basic element. The alternative ribbons can deflect down over small distances by controlling the magnitude of the electrostatic force between the ribbon and the electrode underneath. When ribbons do not deflect, being at the same height as shown in **Figure 75(a)**, the incident light reflects back. When the ribbons move up and down alternatively as shown in **Figure 75(b)**, the incident light diffracts. With the displacement of the ribbon, the diffraction angle is constant but the light intensity varies. Especially when the ribbon displacement is $\lambda/4$, the normal reflection beam is canceled due to the phase difference of $\lambda/2$ generated by the round trip propagation.

Figure 76 shows the schematic view. The GLV consists of a series of ribbons. The region corresponding to one image pixel is 25 μm in length with six ribbons. For the SXGA (1280 × 1024) display as an example, the linear diffraction grating array becomes 25.6 mm in the overall length.

Figure 77 shows the optical setup of the laser projection system using GLV chips (Perry 2004). Light beams of three-color (RGB) lasers are partially diffracted by GLVs. A Fourier (offner-type) filter distinguishes the diffracted light from the reflected light. The laser beams are combined to make up colors of pixels in the video image, for example 1080 pixels (corresponding to HDTV) at a time. The display system contains an external bulk optical scanning mirror (a bulk galvanometric scanner). A 1D linear array of GLV modulates pixel colors for a single column. Once an image is in a frame, this vertical line image is scanned across the screen to produce a complete image. **Figure 78** shows the schematic drawing illustrating the image formation. This is the combination of a 1D

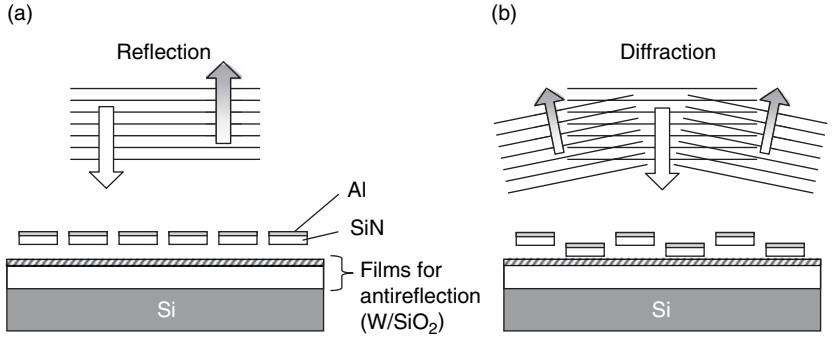


Figure 75 Grating light valve (GLV) ribbons at undeflect and deflected conditions for creating reflection and diffraction beams.

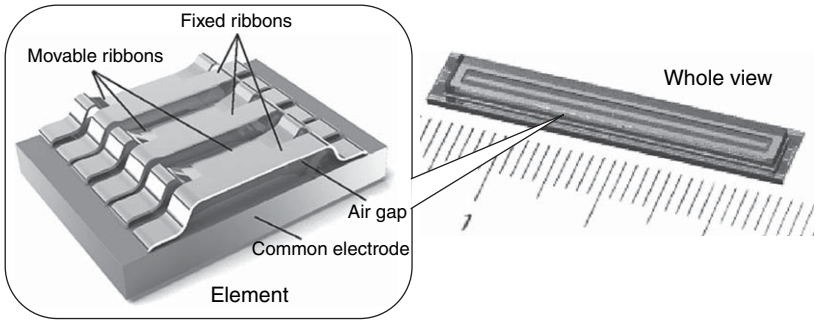


Figure 76 Schematic of whole and magnified element view of grating light valve (GLV). (Whole view courtesy: Sony. http://www.sony.co.jp/SonyInfo/News/Press_Archive/200206/02-023/.)

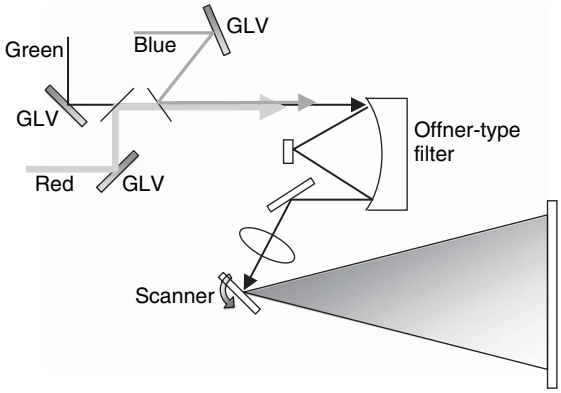


Figure 77 Optical setup of laser projector using three grating light valve (GLV) elements.

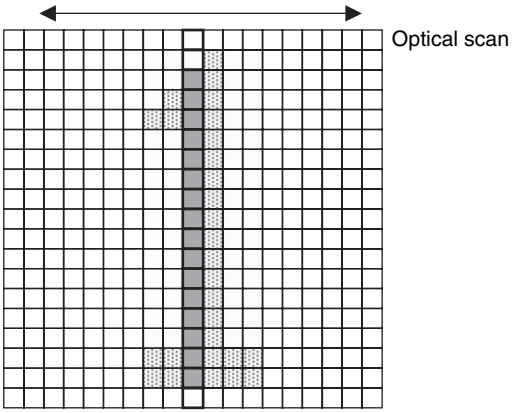


Figure 78 Schematic drawing for illustrating a scanning grating light valve (GLV) array architecture for creating an image.

discrete-pixel array and a 1D scanning system. This is the medium concept shown in **Figure 68(a)** and **(b)** for explaining DMD and laser display, respectively.

Figure 79 shows the schematic of the GLV element, which corresponds to a single image pixel. The fundamental element is the dual-supported parallel ribbon formed with SiN, coated with a reflective top metal layer. The metal layer is used also as an electrical conductor to realize electrostatic attraction to the bottom common electrode. The switching speed will be fastest in the optical MEMS devices. The first paper in 1992 mentions a 20-ns response, which may use the pull-in transition of the electrostatic actuator. The recent documents report a resonant frequency in the megahertz range (the exact value is tuned depending on the design). Before *c.* 1996, the ribbon was designed to contact the underneath electrode, as shown in **Figure 79(a)**. The reason could be for obtaining an accurate $\lambda/4$ displacement over a longer ribbon region. The contacting mechanism will involve the problem of charge up of the insulating material (SiN in this case) or the reliability issue. Now ribbons are designed to operate without contact as shown in **Figure 79(b)** (Senturia 2001). The incident light is focused slightly at around the center of the ribbon, where the displacement is considered to be equal. Current GLV devices are said to have diffraction efficiency near the theoretical maximum for square-well diffraction grating of 81% (into the ±first orders), fill factor of 95%, and top layer reflectivity of 91%, with an overall device efficiency of about 70%. This optical efficiency corresponds to an insertion loss of ~1.5 dB.

The diffraction pattern from GLV can be analyzed using the diffraction theory. Consider the array of rectangular slits (width c, length is not considered for the simplicity) having the pitch d as shown in **Figure 80**. The number of slits is N. The wave amplitude $u(x)$ is obtained as follows starting from eqn [1]. Because the shape of the slit is

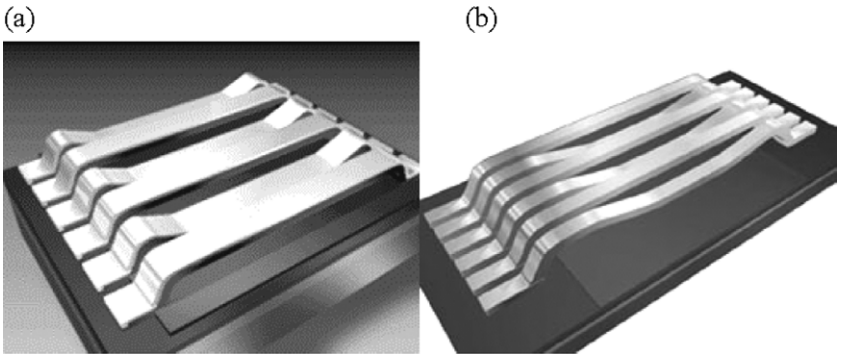


Figure 79 Grating light valve (GLV) ribbons corresponding a single-image pixel using (a) contact. (Source: Trisnadi J I, Carlisle C B, Monteverde R 2004 Overview and applications of grating-light-valve-based optical write engines for high-speed digital imaging. *Proc. SPIE* **5348**, 52–64.) and (b) noncontact scheme. (Source: Payne A, DeGroot W, Monteverde R, Amm D 2004 Enabling high-data-rate imaging applications with grating light valve technology. *Proc. SPIE* **5348**, 76–88.

the same, each slit generates the integration $C\int_{-c/2}^{c/2}\exp[\mathrm{i}\{(2\pi/\lambda)(\xi x/f)\}]\mathrm{d}\xi$. The difference is the relative phase depending on the position of the slit, which is multiplied to the integration. The total light intensity can be obtained as summation as follows:

$$u(x) = [1+\exp[\mathrm{i}\{(2\pi/\lambda)(dx/f)\}]+\exp[\mathrm{i}\{(4\pi/\lambda)(dx/f)\}]+\cdots+\exp[\mathrm{i}(N-1)\{(2\pi/\lambda)(dx/f)\}]]\,C\int_{-c/2}^{c/2}\exp[\mathrm{i}\{(2\pi/\lambda)(\xi x/f)\}]\,\mathrm{d}\xi$$
$$= \frac{-\exp[\mathrm{i}N\{(2\pi/\lambda)(dx/2f)\}](\exp[\mathrm{i}N\{(2\pi/\lambda)(dx/2f)\}]-\exp[\mathrm{i}N\{(2\pi/\lambda)(dx/2f)\}])}{-\exp[\mathrm{i}\{(2\pi/\lambda)(dx/2f)\}](\exp[\mathrm{i}\{(2\pi/\lambda)(dx/f)\}]-\exp[-\mathrm{i}\{(2\pi/\lambda)(dx/f)\}])}\,C\frac{\exp[\mathrm{i}\{(2\pi/\lambda)(cx/2f)\}]-\exp[-\mathrm{i}\{(2\pi/\lambda)(cx/2f)\}]}{\mathrm{i}\{(2\pi/\lambda)(x/f)\}} \quad [30]$$

The intensity from the slit array can be expressed as follows:

$$I(x) = N^2 I_0\left[\frac{\sin N\frac{2\pi}{\lambda}\frac{dx}{2f}}{N\sin\frac{2\pi}{\lambda}\frac{dx}{2f}}\right]^2\left[\frac{\sin\frac{2\pi}{\lambda}\frac{cx}{2f}}{\frac{2\pi}{\lambda}\frac{cx}{2f}}\right]^2 \quad [31]$$

where I_0 is the intensity at the center of the diffraction pattern from a single slit having width c. There is the factor having the shape of the sinc function:

$$\left[\frac{\sin\frac{2\pi}{\lambda}\frac{cx}{2f}}{\frac{2\pi}{\lambda}\frac{cx}{2f}}\right]^2$$

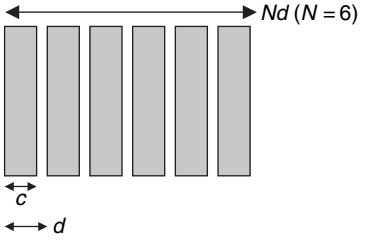


Figure 80 Schematic drawing for illustrating the grating light valve (GLV) ribbons that will interfere with the laser beam.

This is the same factor shown in eqn [2] obtained as the diffraction pattern from the single rectangular mirror. The grating pattern is considered as the superposition of the diffraction pattern of a single slit and the factor

$$\left[\frac{\sin N\frac{2\pi}{\lambda}\frac{dx}{2f}}{N\sin\frac{2\pi}{\lambda}\frac{dx}{2f}}\right]^2$$

which comes from the array.

Figure 81 shows the diffraction pattern expressed by eqn [31]. The ribbon condition $d=3c$ is maintained. Three patterns are shown changing $N=4$, 6, and 12. As for the display device, the ribbon number for one image pixel is first designed to be 4, and later increased to 6. The illuminated region is the center of the ribbon focusing the laser beam slightly. There are peaks at the condition of the numerator corresponding to $\sin N((2\pi/\lambda)(dx/2f))=\pm 1$. Fundamentally, this generates small peaks named subsidiary maxima. With them, there are large peaks named principal maxima. The position of the principal maxima is independent of N but decided by the condition of the denominator

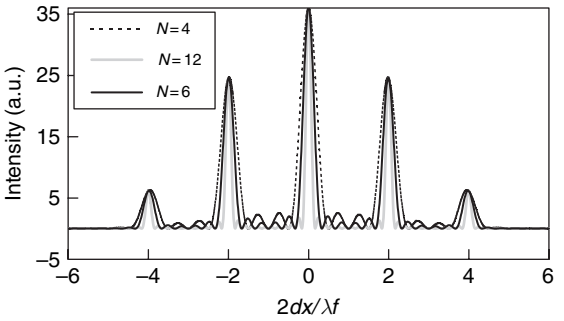


Figure 81 Diffraction pattern from the grating. The grating ribbons are supposed to have the relation of $d = 3c$. Three curves correspond to $N = 4$, 6, and 12.

$\sin((2\pi/\lambda)(dx/2f)) = 0$. The position is given by the following equation:

$$x_m = \frac{m\lambda}{d} f \quad (m = 0,\ 1,\ 2,\ 3, \cdots) \qquad [32]$$

The number of subsidiary maxima between the principal maxima is $N-2$. The peak height of the subsidiary maxima decreases with the increase of N. Because the principal maxima (especially $\pm$ first order) are used, some loss of contrast is incurred due to the existence of the subsidiary maxima at the same position. This loss can be decreased with the increase of the number N of ribbons for each image pixel. There is a trade-off between increasing the number of image pixels and obtaining the higher contrast, because the increase of N means preparing more GLV ribbons.

With the increase of N, the full-width at half-maximum of the principal maxima decreases. The width defined by the length Δx from the peak center to the neighboring 0 position is expressed as follows from eqn [31].

$$\Delta x = \frac{\lambda f}{Nd} \qquad [33]$$

The denominator Nd corresponds with the width of the reflecting area. The wider the reflecting area Nd, the sharper the diffraction spot. This is fundamentally the same as expressed by eqn [7] showing the diffraction limit. As can be seen in **Figure 81**, the gaps between ribbons do not appear inside the peak spot. This characteristic is unique compared to the standard micromirrors. The peak intensity changes depending on the fill factor of the reflective area. This factor is included in I_0 in eqn [31]. The basic diffraction pattern does not change.

Figure 82 shows the illustration for explaining two different ribbon heights. With the displacement of the

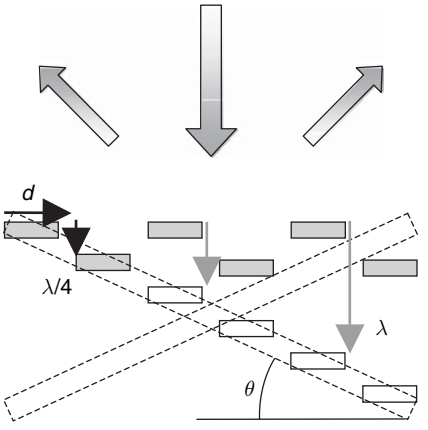


Figure 82 Schematic drawing for illustrating the role of grating light valve (GLV) ribbons.

ribbons, the relative phase of the wave reflected from the neighboring ribbon changes. When the GLV ribbons deflect downward exactly by $\lambda/4$ of the incident light, the diffraction spot intensity becomes maximum. In this case, the mathematical model becomes as follows. The phase difference between neighboring ribbons becomes $\lambda/2$ or π. As the phase 2π can be added without disturbing the phase condition, the series of integration can be expressed as follows instead of expressing it in terms of eqn [30]:

$$\begin{aligned} u(x) = 1 &+ \exp[\mathrm{i}\{(2\pi/\lambda)(\mathrm{d}x/f) + \pi\}] \\ &+ \exp[\mathrm{i}2\{(2\pi/\lambda)(\mathrm{d}x/f)\} + \pi] + \cdots \\ &+ \exp[\mathrm{i}(N-1)\{(2\pi/\lambda)(\mathrm{d}x/f) \\ &+ \pi\}]\, C \int_{-c/2}^{c/2} \times \exp[\mathrm{i}\{(2\pi/\lambda)(\xi x/f)\}]\ \mathrm{d}\xi \end{aligned} \qquad [34]$$

$$I(x) = N^2 I_0 \left[\frac{\sin N\left(\frac{2\pi}{\lambda}\frac{dx}{2f} + \frac{\pi}{2}\right)}{N \sin\left(\frac{2\pi}{\lambda}\frac{dx}{2f} + \frac{\pi}{2}\right)}\right]^2 \left[\frac{\sin \frac{2\pi}{\lambda}\frac{cx}{2f}}{\frac{2\pi}{\lambda}\frac{cx}{2f}}\right]^2 \qquad [35]$$

Figure 83 compares the diffraction patterns calculated from eqn [35] (gray curve) and eqn [31] (black curve). The black curve is when the ribbons are at the

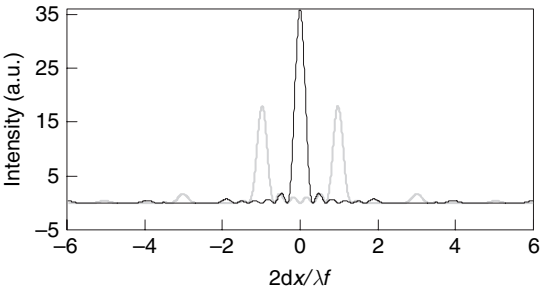


Figure 83 Calculated diffraction pattern from grating light valve (GLV) based on eqn [35].

same height. The ribbon pitch d is 4.17 μm (25 μm element size according to the actual device). The number of ribbons N is 6. I_0 is supposed to be 1. The ribbon gap that is supposed to be 10% of d setting is 0.417 μm. Due to the narrow gap, the diffraction peak that should appear at $2dx/\lambda f = \pm 2$ is small. The parallel ribbons can be considered as a single mirror. When the ribbon deflects downward by $\lambda/4$, the diffraction pattern will change to the gray curve. The center peak disappears by the destructive interference. Instead, the constructive interferences occur at the condition where the denominator becomes zero, i.e., $\sin((2\pi/\lambda)(cx/2f)+(\pi/2)) = 0$. Due to the phase offset $\pi/2$, the peak position shifts. As the result, the center principal peak is mainly divided to two peaks at the position expressed as follows:

$$\frac{x}{f} = \pm\frac{\lambda}{2d} \qquad [36]$$

This is the modulation of the light beam generated by the small displacement of ribbons.

Consider the condition when the ribbon deflects by $\lambda/4$. As the phase difference of λ (as a total of round trip) is same with the phase 0 for the light, the ribbon array can be considered as a single inclined mirror plane as shown by the dotted rectangles. The two inclining directions can be assigned due to the arbitrary nature of the phase. This inclined angle can be expressed as follows:

$$\tan\theta = \pm\frac{\lambda}{4d} \qquad [37]$$

Remembering that the optical reflection angle becomes double compared to the mirror rotation angle, this equation becomes equivalent to eqn [36]. The principal spots from GLV can be considered as the reflection from the equivalent tilt mirrors. In this meaning, GLV can be considered as a kind of micromirror.

The light powers carried by the zeroth and ± first orders are dominant and nearly complementary as seen from **Figure 83**. There are two possible optical setups using GLV (Trisnadi *et al.* 2004). One uses the reflected beam (zeroth order) as shown in **Figure 84(a)**. The other uses diffracted beams (± first orders) as shown in **Figure 84(b)**. The location where the orders are optimally separated is at the Fourier transform plane. The spatial filter is placed and discriminates reflected and diffracted beams. This Fourier filter is analogous to the Schlieren filter. Because a slit that path zeroth-order beams has a finite width, some diffracted peaks (corresponding to two subsidiary maxima under the zeroth-order principal maximum in **Figure 83**) cannot be filtered out. From the theoretical curve, the contrast becomes about 30 for the six ribbons (Tamaki *et al.* 2004). Use of more ribbon pairs generates less diffracted energy in the vicinity of the zeroth-order beam, thus giving higher contrast as seen from **Figure 81**. The computer-to-plate (CTP) system modulating the high-power laser beam uses the reflected beam (Tamaki *et al.* 2004). Twelve ribbons are used for one image pixel. The contrast is ~130. The laser beams are used to write images onto plates for the offset printing based on the thermal reaction activated by the high energy. The display application uses the optical setup shown in **Figure 84(b)**. A contrast larger than 1000:1 has been reported. The light scattering is also necessary to be minimized. The real device may include the technique of the profile control of the ribbon (like blazing).

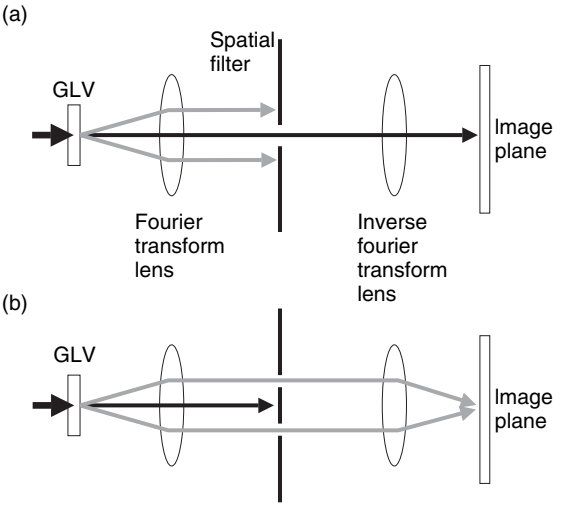


Figure 84 Imaging optics using the zeroth- and first-order beams.

Although both GLV and DMD are well-known devices as spatial light modulators, their features are contrasting. GLV works with a laser beam having a single wavelength λ. DMD works with white light. When the laser beam is used with DMD, the diffractions or the speckles will disturb the image. The photocurrent can disturb the SRAM operation. GLV can handle relatively high-power density because of the absence of IC underneath the ribbons. GLV is controlled in analog, so the displacement accuracy will be stricter than that of DMD using the digital on/off control.

3.01.6.7 Variable Optical Attenuator

The micromirror is applied to the telecommunication network. From the initial stage of the optic-fiber network, the device having the fixed attenuation value is used. With the progress of the WDM system (from *c.* 1996), the device that can control the attenuation (or insertion loss) according to the electrical control signal is required. This is the variable optical attenuator (VOA). A large number of VOAs become necessary because transmission powers of multiple channels of different wavelengths are individually controlled at the wavelength multiplexing and demultiplexing nodes. Many kinds of principles have been considered including the mechanical (conventionally the assembly of a prism or a mirror driven by a solenoid coil or a motor), electromechanically operated Fabry–Perot interferometers, and the magneto-optical methods. VOA is basically the device to be included in the larger systems, and a smaller size is required.

There are two main uses of VOAs. One is leveling the light intensity having different signal wavelengths, and another is leveling the output signals for amplifying the intensity of many photonic signals in erbium-doped fiber amplifier (EDFA) setting a constant amplifying ratio, even when the number of signals or the intensity changes. This need is because the telecommunication system performance becomes lower when the amplifying ratio changes. When optical add drop multiplexing (OADX) and optical cross connect (OXC) are introduced in the optical network, the optical signal intensity can change depending on its route. This nonuniform signal intensity has to be leveled. For using VOAs for each wavelength in WDM system, the arrayed VOA is preferred.

The specification to VOA is somewhat different depending on the system to be applied. The main specifications are

(1) insersion loss (<2 dB),
(2) dynamic range (>15 dB),
(3) polarization-dependent loss (<0.2 dB),
(4) wavelength dependence of attenuation,
(5) reflection (<−45 dB),
(6) driving speed,
(7) resolution of attenuation,
(8) hysteresis, and
(9) power consumption.

Specifications (3) and (4) become important when the VOA deals with many wavelength signals. The mechanical method using the micromirror can realize the little dependence on the wavelength. Specification (6) becomes important when the VOA is required to follow the transient response of EDFA.

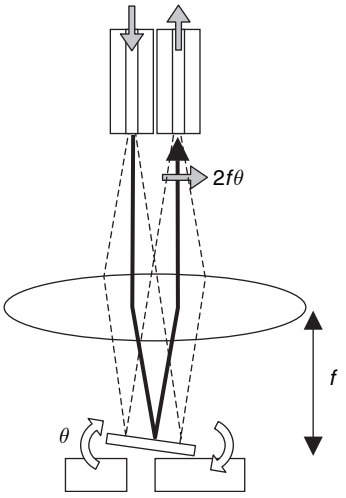


Figure 85 Schematic drawing of optical setup of a variable optical attenuator (VOA) using the micromirror adopted in the Santec Corporation.

Figure 85 shows a schematic drawing of the optical setup of VOA using the micromirror adopted in the Santec corporation (Isamoto *et al.* 2004). The setup is symmetrical to the center axis of the lens. Because the output fiber is at the symmetric position and the mirror is at the focus position, the output light beam is normally incident on the output fiber end with the same beam condition at the input fiber end. The mirror rotation generates the shift of the beam spot, which is the main reason of the attenuation. The large focus length f increases the displacement. When the mirror rotation angle is 0.3° and the focus length is 1.8 mm, the shift displacement is 9.4 μm. Because the core diameter of the fiber is ∼10 μm, the attenuation or the coupling ratio is highly sensitive to the minute displacement of the beam spot.

The attenuation can be calculated based on the coupling of the Gaussian beam with the output optical fiber. When the incident beam profile exactly matches with the mode of the output fiber, the attenuation η can be expressed as follows:

$$\eta = \kappa \exp\left[-\kappa \frac{x_0^2}{\omega^2}\right] \qquad [38]$$

where κ is set to 1 assuming the exact matching of the mode profile, ω is the radius of the Gaussian beam set to 5.5 μm, and x_0 is the shift having the relation of $2f\theta$

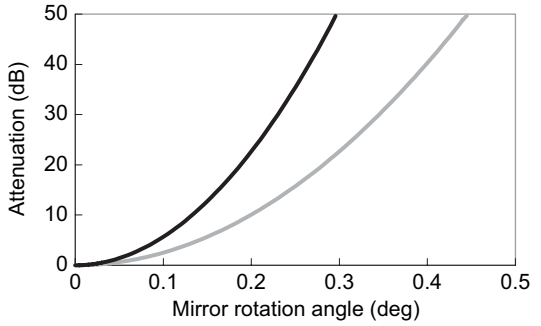


Figure 86 Calculated optical attenuation as a function of the mirror rotation angle.

with the focus length f and the mirror rotation angle θ. **Figure 86** shows the relation expressed by eqn [38]. Black and gray curves are for the focus lengths f of 1.8 and 1.2 mm. f is reported to be 1.8 mm. The mirror angle is controlled up to 0.4°.

With the outside optical setup, a large attenuation can be obtained with a small mirror rotation angle. Therefore, the parallel plate electrostatic torsion mirror can be used. Because the VOA does not need full-time scanning, dynamic deformation is not a problem. **Figure 87** shows the attenuation curve as a function of the driving voltage. A low-voltage operation (DC 5 V or less) is realized. **Figure 88** shows the schematic view of the mirror. A circular mirror with electrostatic parallel plates is suspended

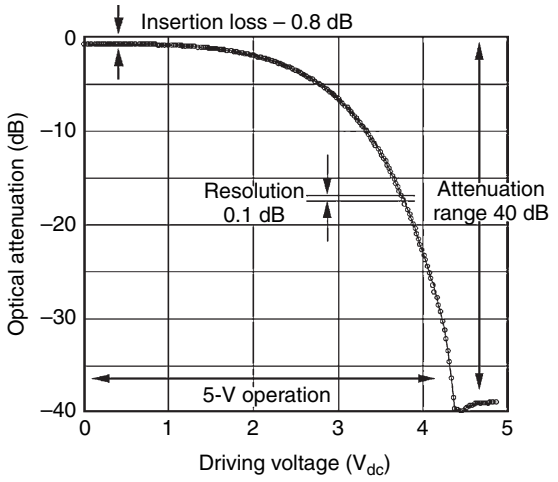


Figure 87 Typical attenuation curve as a function of the driving DC voltage of the micromirror. (Source: Isamoto K, Kato K, Morosawa A, Chong C, Fujita H, Toshiyoshi H 2004 A 5-V operated MEMS variable optical attenuator by SOI bulk micromachining. *IEEE J. Sel. Top. Quant. Electron.* **10(3)**, 570–8.)

with torsion bars. The structure is made of a SOI wafer. The right-side wing of the suspended mirror and the underlying substrate are electrodes of the gap-closing electrostatic actuator. The suspended mirror is biased. The underlying substrate is grounded a 2-μm air gap. The electrostatic attraction torque rotates the actuator plates with a small angle. There is a large through-hole under the mirror to prevent the mirror from contacting the underlying substrate. **Table 7** lists the performance. Because the incident angle to the micromirror is nearly normal even when the mirror is rotated, there is little polarization dependence.

Because shifting of the lens or mirror after packaging becomes the main cause for the degradation in performance, the positions of fibers or lens have to be highly stable. Hence, glue cannot be used for the assembly. The minute angular fluctuation caused by the outside force and the temperature change have to be avoided. A smaller lens holder and a metal package are adopted. **Figure 89** shows the micromirror mounted on the package base. The package size is 5.6 mm in diameter and 2.3 mm in length. The standard for the passive telecom device of Telcordia GR-1221-CORE requires testing with high acceleration. This standard does not take into account moveable elements like micromirror. The requirement for the mechanical shock is, for example, 500 G in 1 ms, five times per direction for six directions. Protecting the package with the suspensions or cushions is difficult, because the VOA device is required to be small. The shock resistance is satisfied by the technology of the metal packaging as well as the micromirror design.

3.01.6.8 Optical Switch

Some devices that have already been described (e.g., **Figures 19(b)** and **29**) are for optical switching. Around 2000, special attention was given to telecommunication applications. The large-scale OXC may be a representative application. **Table 8** shows the list of required specifications for the optical switch. Depending on the applications, the required port size and insertion loss are different. The optical route can be changed directly without conversion from optical to electrical signals. This switching method does not depend on the protocol, signal bit rates, and also wavelength. OXC can be used without disturbing the change of the above specifications. The main challenge of OXC is the realization of low loss in the large-scale system. The 2D switch setup as shown

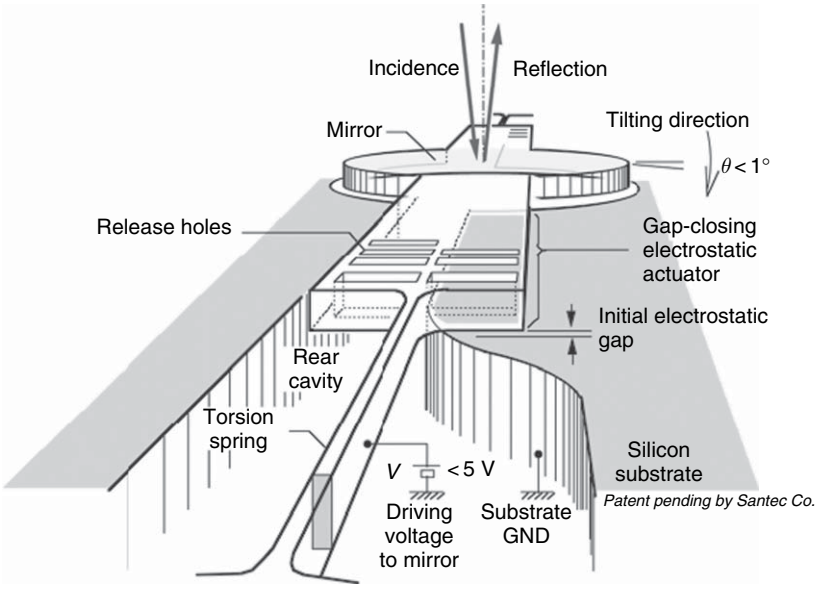


Figure 88 Schematic drawing of the micromirror used in the variable optical attenuator (VOA). (Source: Isamoto K, Kato K, Morosawa A, Chong C, Fujita H, Toshiyoshi H 2004 A 5-V operated MEMS variable optical attenuator by SOI bulk micromachining. *IEEE J. Sel. Top. Quant. Electron.* **10(3)**, 570–8.)

Table 7 Performance of the VOA using micromirror

Parameter	*Value*
Insersion loss (dB)	<0.8 at 0 V
Attenuation range (dB)	40, Blocking: >45
Polarization-dependent loss (dB)	<0.2 (at 20 dB attenuation)
Polarization mode dispersion (ps)	<0.1
Response time (ms)	<5
Driving voltage (V)	DC 5
Power consumption (mW)	<10 max

Source: Isamoto K, Kato K, Morosawa A, Chong C, Fujita H, Toshiyoshi H 2004 A 5-V operated MEMS variable optical attenuator by SOI bulk micromachining. *IEEE J. Sel. Top. Quant. Electron.* **10(3)**, 570–8.

in Figure 29 requires N^2 micromirrors for realizing $N \times N$ OXC. It is said that $\sim 32 \times 32$ scale will be the practical limit. Generally, a higher number of reflections increases the insertion loss. The free-space-type 3D MEMS beam-steering architecture is considered as the most promising candidate. The necessary number of micromirrors is $2N$ for realizing an $N \times N$ OXC.

Figure 90 illustrates the progress in the research on large-port-count OXC. To manage the flow of diverse optical signals, which was accelerated by the explosion of Internet data transport, petabit per second class optical switching capabilities were needed

Figure 89 Fabricated micromirror mounted on a TO-3 package. (Photo courtesy: Santec Corporation.)

Table 8 Applications and requirements of optical switch

Applications	*Port size*	*Switching speed (ms)*	*Insertion loss (dB)*
Protection Switching	1×2, 2×2	50	1
Metro	$\sim 64 \times 64$	50	7
Intraoffice	$\sim 256 \times 256$	50	7
Long-haul	$\sim 256 \times 256$	50	7–12

Source: Courtesy of John Bowers, Dept. of Electrical and Computer Engineering, University of California, Santa Barbara.

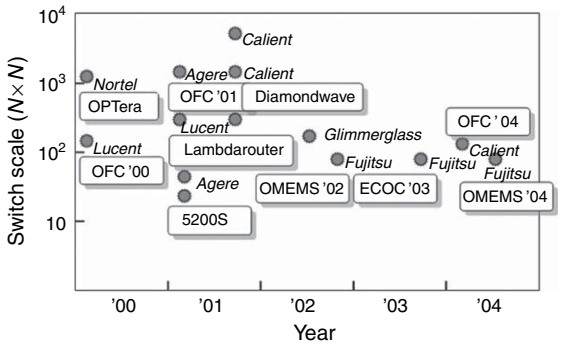


Figure 90 Progress of large-port-count scalable optical cross connect (OXC) switches for the market. (Source: Yano M, Yamagishi F, Tsuda T 2005 IEEE optical MEMS for photonic switching – Compact and stable optical crossconnect switches for simple, fast, and flexible wavelength applications in recent photonic networks. *J. Sel. Top. Quant. Electron.* **11(2)**, 383–94.)

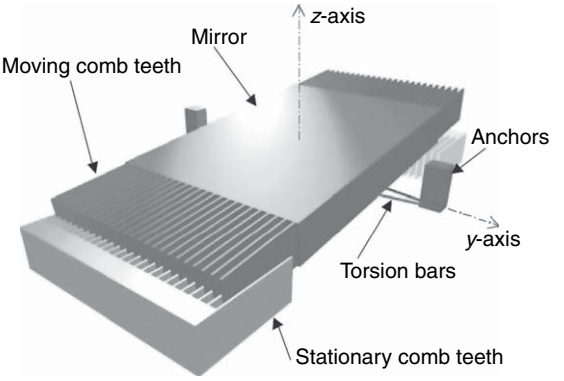


Figure 91 Schematic of single-axis type comb-driven micromirror developed at Fujitsu Laboratories Ltd. (Source: Tsuboi O, Mizuno Y, Koma N, Soneda H, Okuda H, Ueda S, Sawaki I, Yamagishi F 2002 A rotational comb-driven micromirror with a large deflection angle and low drive voltage. *Proc. 15th IEEE Int. Conf. on Micro Electro Mechanical Systems*, Las Vegas, Nevada, USA, pp. 532–5.

in the core optical networks. The large-port-count OXC of classes exceeding 4000 × 4000 were considered. Later, system concerns shifted to metropolitan area networks, because new network services have rapidly proliferated (especially those involving mobile access applications) and data traffic congestion in such regions is a growing problem. OXC of 100 × 100 class are being developed. In 2005, Calient Networks and Fujitsu Laboratories Ltd. were competing. In deployed optical networks, the acceptable full cross connect time among all ports is <5 ms for OXC at each network node, taking into account the permitted overall system failure time of 50 ms.

Figure 91 shows a schematic drawing of the micromirror developed at Fujitsu Laboratories Ltd. (Yano *et al.* 2005). The single-axis version is shown for simplicity. The vertical comb-drive actuator has a high aspect structure with thin comb teeth and a narrow air gap for generating a large electrostatic force. **Figure 92** shows images of the device. The comb-teeth pitch is 20 μm, the thickness is 5 μm, and the air gap is 5 μm. A 100-μm-thick Si layer on an SOI substrate is processed. The micromirror has two pairs of torsion bars realizing the rotation along the two axes. The mirror diameter is 650 μm. The torsion bar used is 5 μm thick and 3 μm wide. These narrow gaps require the torsion bars to provide stable tilt motions. A V-shaped torsion bar as shown in **Figure 91** is adopted. The V-shaped torsion bars have larger stiffness against the imbalanced forces causing the pull-in instability. Taking this advantage, the thick vertical combs can generate a larger driving force enabling both larger rotational angles and faster mirror motion.

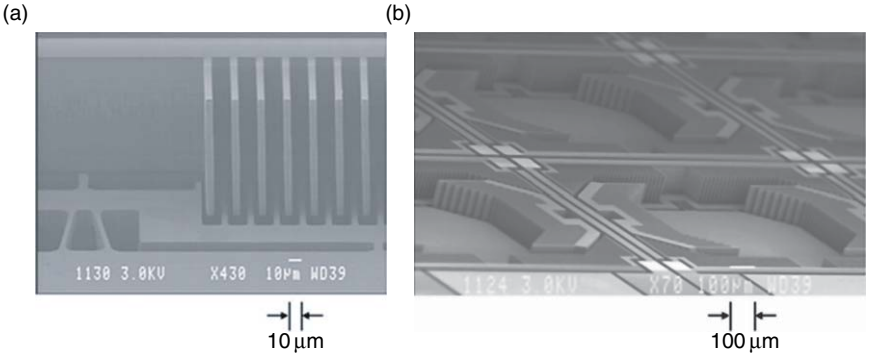


Figure 92 (a) Scanning electron microscope (SEM) view of the built-in comb-teeth structures. (b) Overall layout view of micromirrors and actuators. (Source: Yano M, Yamagishi F, Tsuda T 2005 IEEE optical MEMS for photonic switching – Compact and stable optical crossconnect switches for simple, fast, and flexible wavelength applications in recent photonic networks. *J. Sel. Top. Quant. Electron.* **11(2)**, 383–94.)

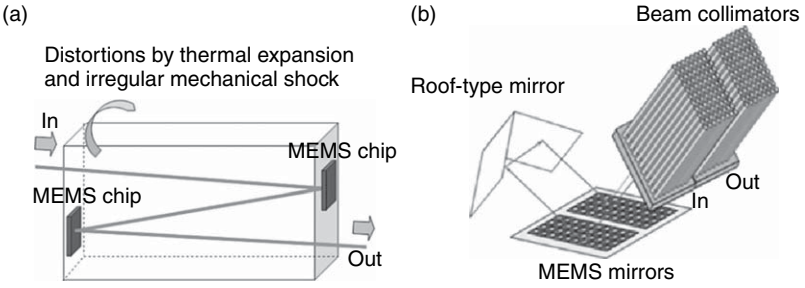


Figure 93 Optical configurations. (a) Rectangular frame used in the conventional housing. (b) Folded configuration with a retroreflector. (Source: Yano M, Yamagishi F, Tsuda T 2005 IEEE optical MEMS for photonic switching – Compact and stable optical crossconnect switches for simple, fast, and flexible wavelength applications in recent photonic networks. *J. Sel. Top. Quant. Electron.* **11(2)**, 383–94; Milanovic V, Last M, Pister K S J 2001 Torsional micromirrors with lateral actuators. *Proc. 11th Int. Conf. Solid-State Sensors and Actuators*, pp. 1298–301.)

The optical setup is the symmetric confocal optics. Through the lens in front of the fiber bundles, the beam waists coincide at the center of the optical setup. This compensates the optical divergence. The centrally located beam waist can minimize the mirror and the lens size. **Figure 93(a)** shows a rectangular frame used in conventional housing. The optical beam is reflected twice by the micromirrors. The optical path is Z-shaped. Fujitsu adopted a cubic frame as shown in **Figure 93(b)**. The mechanical stiffness increases and the mirror chips can be set parallel to the base plane. This is advantageous for the assembly. Notice that the p-polarized beam at the micromirrors is incident on the roof-type mirrors as the s-polarized beam. The polarization dependence is averaged. The optical condition is almost the same within the small rotation angle of $\sim 5°$.

Figure 94 shows the optical setup for explaining the characteristics using flat and roof-type mirrors. In case of the flat mirror, the maximum angle required for the micromirror can be expressed as follows:

$$\theta_{\max 1} \approx \frac{D}{l_1} \qquad [39]$$

where D is the total length of the mirror array chip and will be $\sim$11 mm for a 10×10 matrix if the spacing between micromirrors is 1.1 mm. l_1 is the distance between the device and the mirror. This maximum angle is decided by the condition for steering the beam connecting corner-to-corner mirrors. In case of the roof-type mirror, the maximum angle can be expressed as follows:

$$\theta_{\max 1} \approx \frac{D}{\sqrt{2}s + 2l_2} \qquad [40]$$

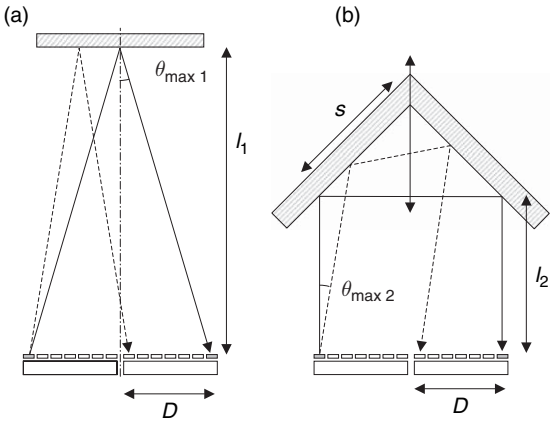


Figure 94 Schematic drawing of the optical configuration using (a) rectangular frame and (b) roof-type mirror for illustrating the parallel shift effect of the roof-type mirror.

The distances l_2 between the device and the mirror are relatively dominant. By the factor 2, the mirror rotation angle can be decreased or the shorter optical path length can be allowed, realizing a more compact switch fabric. This basically occurs from the parallel shift effect of the roof-type mirror. The normal beam incidence on the micromirror is allowed at the rest condition. The flat mirror requires the mirror rotation of $\theta_{\max 1}/2$ for connecting to the opposite corner. The reported total path length $(\sqrt{2}s + 2l_2)$ is $\sim$110 mm.

Figure 95 shows another example of an optical setup for a 238×238 OXC (named LambdaRouter) researched at Lucent Technologies (Aksyuk *et al.* 2003). The basic setup is a rectangular frame. At the center, a Fourier lens is placed for setting the beam waist at each micromirror. Two micromirror arrays (shown in **Figure 19(b)**) are placed at the left and right sides. Due to the function of the Fourier lens,

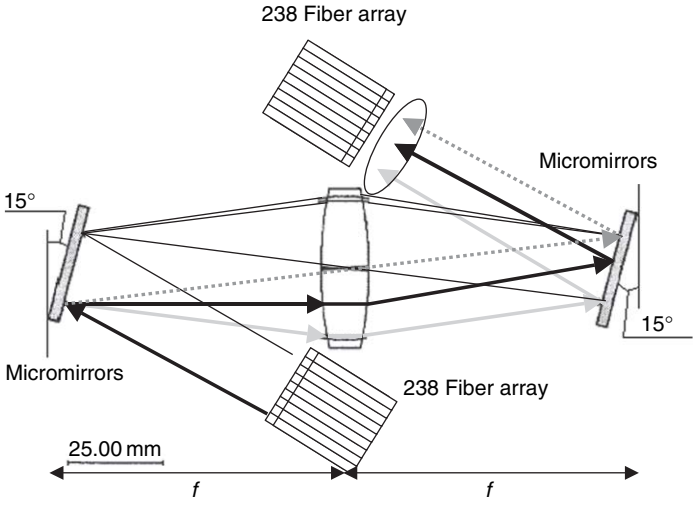


Figure 95 Optical configuration of the optical cross connect (OXC) using the Fourier lens. Beam paths are shown.

the reflection beam from the micromirror at the rest position will focus on the center mirror in the second micromirror array placed at the right side. The black curve shows the beam path. The required task of mirror rotation angle becomes the same for the arrayed mirrors. When the mirror rotates in the ± direction, the reflection beam shifts to the ± direction from the center as shown by gray curves. The Fourier lens can minimize the task of the micromirrors. It should be noted that the spot size of the beam is 372 μm on the micromirror, and that the mirror size is 600 μm. The additional size of the mirror can be the margin for the risk of clipping of the beam. The difficulty of the assembling can be alleviated.

3.01.7 Researches for Higher Performances

The micromirrors share some of the fundamental requirements. In this section, researches for higher performances of the micromirror are discussed.

3.01.7.1 Large Rotation Angle with Low-Voltage Driving

One research effort is realizing the larger rotation angle with the lower-voltage driving especially for the electrostatic actuator. The established system usually works with 5, 12, or 24 V DC sources. The higher driving voltage requires the additional power source and the driving circuit with the higher resistance. Fundamentally, there are two methods for decreasing the driving voltage. One is increasing the driving force, and another is decreasing the spring constant. Adopting both the methods, Hah *et al.* (2004b) report the poly-Si micromirror fabricated by surface micromachining (Sandia's ultraplanar multilevel MEMS technology-V, SUMMiT-V). **Figure 96** shows the schematic structure and the device image. The vertical comb-drive actuators and torsion springs are placed underneath the mirror (137 μm × 120 μm). **Figure 97** shows the mirror rotation angle. For the device with 0.5-μm finger spacing, a scan angle of 5.9° is obtained at 6 V bias. The finger length is 30 μm. The spring is 1 μm in width and thickness and 55 μm in length on each side. Considering the design enabling the plus and minus angles, the mirror rotation angle is 11.8° as the total. The maximum angle is limited by vertical pull-in instability due to the residual parallel plate capacitance between the mirror and the fixed fingers. As the finger gap spacing increases, the maximum continuous angle decreases because of the earlier occurrence of pull-in. The resonant frequency of the mirror ranges from 3.4 to 8.1 kHz.

In general, the torsion bar is frequently designed as the serpentine spring. The folded bar will be softer not only for rotation, but also for vertical and in-plane displacement and rotation. The resultant mirror motion depends on the balance of the spring constants for many modes of the mirror motion. The spring constant for the mirror rotation should be minimized while maintaining the rigidity in other motions. The serpentine spring helps the design, but often does not completely solve the problem.

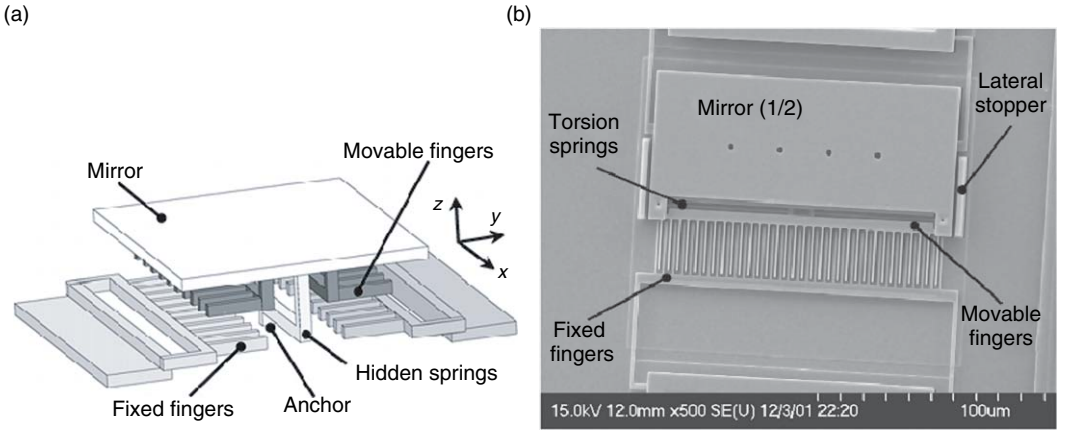


Figure 96 (a) Schematic view and (b) scanning electron microscope (SEM) image of the micromirror with hidden vertical comb drives and springs. The lower half of the micromirror was removed to reveal the underlying comb structures. (Source: Hah D, Huang S T, Tsai J, Toshiyoshi H, Wu M C 2004b Low-voltage, large-scan angle MEMS analog micromirror arrays with hidden vertical comb-drive actuators. *J. Microelectromech. Syst.* **13**, 279–89.)

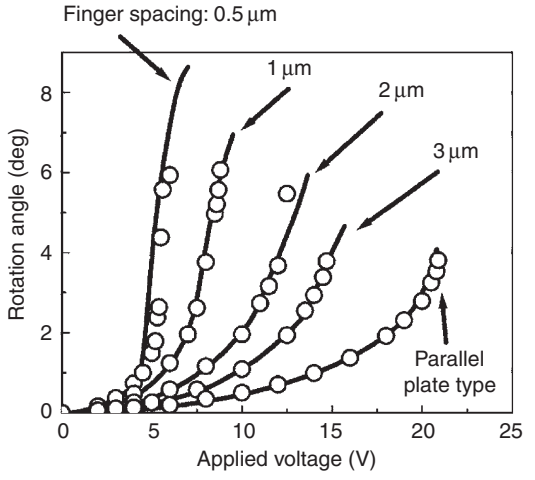


Figure 97 Measured (marks) and calculated (curves) DC rotation angle of the micromirror. The parameter is the finger spacing. (Source: Hah D, Huang S T, Tsai J, Toshiyoshi H, Wu M C 2004b Low-voltage, large-scan angle MEMS analog micromirror arrays with hidden vertical comb-drive actuators. *J. Microelectromech. Syst.* **13**, 279–89.)

The thin-film torsion bar is proposed for decreasing the rotational spring constant (Sasaki *et al.* 2006b). The tension is included inside for increasing the rigidity against the unwanted mirror motion, which will be the limiting factor of the maximum rotation angle. The torsional spring constant k_θ is expressed by eqn [26]. Remember the supposed relation of $w > t$. When thin films are considered, t is the thickness. The factor t^3 shows that decreasing the value t is effective for decreasing k_θ. Considering the deposited thin film as the structural material, the thickness of a few hundreds of nanometers can be obtained. Based on eqn [26], the spring constant is 1.4×10^{-11} N m rad^{-1} by setting the values of G, l, w, and t as 80 GPa, 200, 4, and 0.3 μm, respectively.

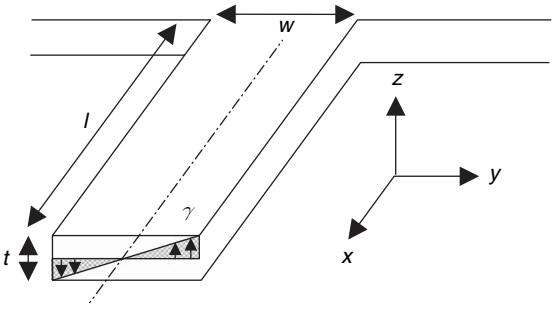
Consider the effect of tension included in the torsion bar. The tensile stress σ_0 inside the SiN film grown by low-pressure chemical vapor deposition can be 760 MPa, the tension $T = \sigma_0 wt$ is 910 μN for a 4 μm × 0.3 μm cross section. This value is large compared to the electrostatic driving force in normal cases. Because the thin-film torsion bar is the softest in the vertical direction, the vertical spring constant k_z is most important for the stability of the mirror motion. k_z is expressed as follows (Senturia 2001):

$$k_z \approx \frac{1}{\sum_{n=1,\ \text{odd}}^{\infty} \frac{1}{l} \frac{2}{\frac{Ewt^3}{12} k_n^4 + \sigma_0 wt k_n^2}}, \quad k_n = \frac{n\pi}{l} \quad (n = 1,\ 3,\ 5 \cdots) \qquad [41]$$

The first and the second terms correspond to the factors due to the elasticity E (290 GPa) and the stress σ_0. k_z are estimated to be 0.016 and 19 N m^{-1} when σ_0 is 0 and 760 MPa, respectively. The tension is dominant for deciding the spring constant. This large spring constant suppresses the vertical displacement of the mirror. The increase of the torsional spring constant $\Delta k_{\theta\ \text{stretch}}$ generated by the beam stretching can be estimated as follows:

$$\Delta k_{\theta\ \text{stretch}} \approx \frac{1}{8l} \left\{ \frac{\sigma_0 (w^3 t + wt^3)}{3} + \frac{E}{2l^2} \left(\frac{w^5 t}{10} + \frac{w^3 t^3}{9} + \frac{wt^5}{10} \right) \theta^2 \right\} \qquad [42]$$

The first term corresponds to stretching against the stress σ_0. The second nonlinear part corresponds to the elastic stretching. The dominant part is the first term having the value of $3.1 \times 10^{-12}\,\mathrm{Nm\,rad^{-1}}$ when the mirror rotation is not large (<6.6 rad). This is 22% of the spring constant estimated from eqn [26]. The tension is fundamentally perpendicular to the rotational displacement of the material inside the torsion bar. The increase of the spring constant is minimized.



Deduction of eqn [42] of $\Delta k_{\theta\ \mathrm{stretch}}$ using energy method: The transverse displacement $\omega(x, y, z)$ of any point generated by the mirror rotation inside the torsion bar can be expressed as follows:

$$\omega \approx \frac{\theta x}{l}\sqrt{y^2 + z^2} \qquad [43]$$

where θ is the rotation of the torsion bar in radian. With the help of the calculation mentioned in Senturia (2001, Chapter 10), the axial strain ε_x can be obtained.

$$\varepsilon_x = \frac{1}{2}\left(\frac{d\omega}{dx}\right)^2 = \frac{\theta^2}{2l^2}\left(y^2 + z^2\right) \qquad [44]$$

The strain energy can be found as follows:

$$\begin{aligned} W &= \frac{1}{2}\int \sigma_x \varepsilon_x \,\mathrm{d}V = \frac{1}{2}\int (\sigma_0 + E\varepsilon_x)\varepsilon_x \,\mathrm{d}V \\ &= \frac{1}{2}\left(\sigma_0 \frac{\theta^2}{24l}\left(w^3 t + w t^3\right)\right. \\ &\quad \left. + E\frac{\theta^4}{32l^3}\left(\frac{w^5 t}{10} + \frac{w^3 t^3}{9} + \frac{w t^5}{10}\right)\right) \end{aligned} \qquad [45]$$

The work done by the external load is $T\theta$, where T is the torque. The total potential energy is

$$U = W - T\theta \qquad [46]$$

Taking the derivative with respect to θ and setting to zero yields T.

$$T = T\frac{1}{2}\left(\frac{\sigma_0\left(w^3 t + w t^3\right)}{12l}\theta + \frac{E}{8l^3}\left(\frac{w^5 t}{10} + \frac{w^3 t^3}{9} + \frac{w t^5}{10}\right)\theta^3\right) \qquad [47]$$

From this relation, eqn [42] is obtained.

Figure 98 shows a fabricated micromirror. The structures of the thin-film torsion bar and the vertical comb-drive actuator are realized by combining isotropic and anisotropic Si plasma etching. The magnified image shows the torsion bar and the vertical comb. The vertical comb is generated by the downward deflection of the cantilever generated by the compress stress in an SiO_2 film.

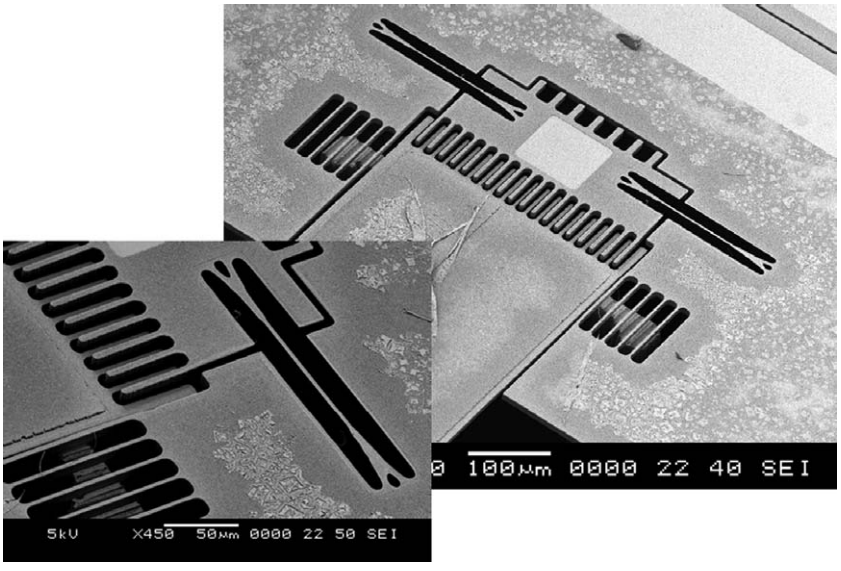


Figure 98 Fabricated micromirror and the magnified image of the torsion bar and the vertical comb.

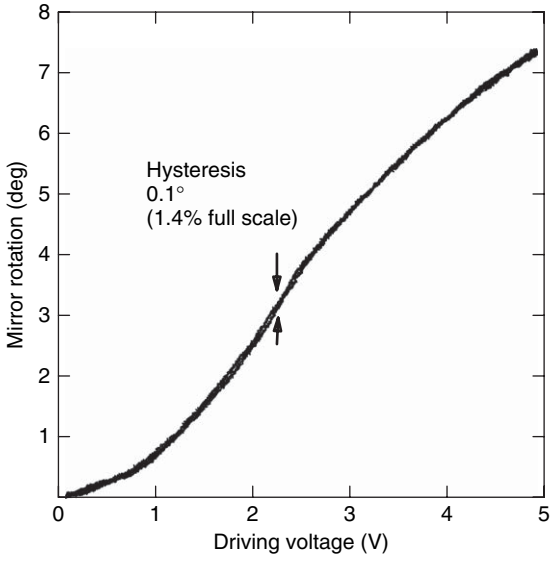


Figure 99 Mirror rotation as a function of the driving voltage.

Figure 99 shows the mirror rotation angle as a function of the driving voltage. The curves change from the concave up to down with increasing driving voltage. The thickness of the comb finger is 10 μm. The rotation angle reaches 7.3° at 5 V. The comb gap is 4 μm in the lateral direction. The curve is smooth showing a nearly linear relation. Data are for a round trip. The hysteresis (defined by the maximum difference of the rotation angle at the same driving voltage in the round trip motion) is 0.1°. This corresponds to 1.4% of the full scale. The resonance frequency of the mirror is about 670 Hz. The stress values are monitored using a strain gauge. Assuming the Young's modulus E of an SiN film to be 290 GPa, the estimated stress is 780 MPa.

3.01.7.2 Accurate Rotation Angle Control

In general, the applications require the mirror rotation angle, and the servo control becomes indispensable. As described, the galvanometric scanner realizes the high accuracy of the mirror rotation angle by combining with the rotation angle sensor. Consider a simple example of the laser printer of 600 dpi supposing the optical arm of 10 cm, the neighboring dot will be generated by the 0.012° mirror rotation. The accuracy will be on the order of 0.001°. The simple open-loop control of the miromirror cannot guarantee accuracy at this level. Remember that some microactuators have hysteresis in their motion (e.g., piezoelectric actuation). The feedback control of the rotation angle based on a sensor is desired. OXC uses the feedback control monitoring the light intensity signal, coupled with the output fiber. With the port switching, the combinations between the output port and the controlling mirrors change. The system becomes complex because the combination of the signal and the feedback loop changes from time to time. Since the driving circuits have a minute difference in their performance, the controlling method becomes complicated. Integrating the sensor inside the micromirror device is attractive for making the system simple and reliable. The same sensor signal and the same driving circuit can be used for controlling the same micromirror.

When the micromirror uses the electromagnetic actuator for full-time scanning application, open voltage based on the inverse electromagnetic effect can be used for sensing the rotation angle. Figure 100 shows the earlier version of the scanner developed at Olympus Corp. The sensing coil is prepared with the driving coil (Miyajima *et al.* 2001). Even when the micromirror does not have a sensing coil, the time division method can be used for detecting the open voltage. In a short period, the coil on the micromirror is driven with enough power, and then switched to monitor the current in another period for detecting the scanning amplitude. Sensing rotation angle using

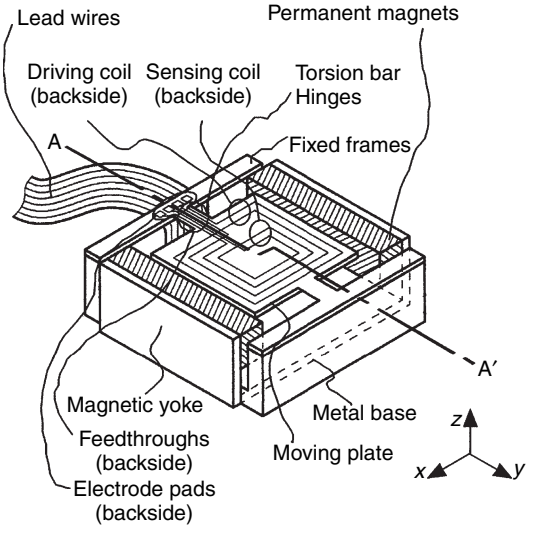


Figure 100 Schematic drawing of electromagnetic scanner with sensing coil. (Source: Miyajima H, Asaoka N, Arima M, Minamoto Y, Murakami K, Tokuda K, Matsumoto K 2001 A durable, shock-resistant electromagnetic optical scanner with polyimide-based hinges. *J. Microelectromech. Syst.* **10(3)**, 418–24.)

electromagnetic coupling at a high-frequency signal between the driving coil and detecting coil has been tried from an early stage. The detecting coil is prepared on the glass basement (Asada *et al.* 1994). A high-frequency (1 MHz) current is superimposed on the driving current, so that this does not disturb the mirror motion. Via the mutual inductance between the driving coil and the detection coil, a open voltage is measured at the fixed detection coil placed underneath. This signal has a linear relation to the rotation angle of the mirror.

There are some patents. Onix Microsystems Inc. proposes detecting the capacitance between electrodes, which will change with the mirror motion (Behrang and Satinderpall 2001). Texas instruments Inc. proposes assembling the lasers and photodetectors at the backside of the micromirrors inside the device package (Orcutt *et al.* 2001). The tails of the reflected beam profile are monitored. By comparing the signal magnitudes, the mirror rotation angle can be measured. Assembling elements with high accuracy will be necessary.

Kallweit and Zappe (2006) propose the use of a part of the incident light beam for sensing the mirror rotation. A transmission grating (period: 2 μm) is prepared in a small part of the micromirror. On the backside of the micromirror, a position-sensitive detector (PSD) is assembled for detecting the transmitted light beam through the grating. **Figure 101** shows the setup. The device is composed of three layers: a Si mirror layer on the top, then a Pyrex contact layer using indium–tin oxide electrodes, and a PSD at the bottom. The major part of the light power is reflected at the mirror. A small part of the incident light is transmitted through the grating. This light beam generates a tilt-angle-dependent diffraction pattern on a PSD.

Considering the case where the incident laser beam is orthogonal to the initial position of the mirror and the grating, the diffraction angle γ has the following relation with the mirror rotation angle β:

$$\gamma(\beta) = \sin^{-1}\left(\frac{m\lambda}{\Lambda} - \sin\beta\right) \pm \beta \qquad [48]$$

where Λ is the grating pitch and $m\ (= \pm 1, \pm 2, \cdots)$ is the diffraction order. The zeroth-order beam stays at the center. Knowing the wavelength of the laser light λ, the period of the grating Λ and the distance between the mirror and the PSD allows the determination of the mirror tilt angle β by measuring the position of the incident light on the PSD (related to γ). **Figure 102** shows the relation setting the value $\lambda = 650$ nm and $\Lambda = 2$ μm. The positive and negative diffraction maxima do not shift symmetrically upon the mirror rotation angle. The demonstrated resolution of the angular sensing is ~0.03° (confirmed angular resolution of 0.066° with the reference error of 0.04°) for a wavelength of 650 nm and a spacing of 7.90 mm between the mirror and the underlying PSD.

Another sensing principle is the piezoresistive method (Yalcinkaya *et al* 2006). The strain introduced in c-Si changes the conduction and valence band structures. The resultant carrier density or mobility change is observed as a bulk resistivity change. Bourouina *et al.* (2002) reported a magnetostrictive actuator integrated with a piezoresistive sensor. Taking advantage of the simplicity of the magnetostrictive microactuator, two Wheastone bridges are incorporated near the clamping

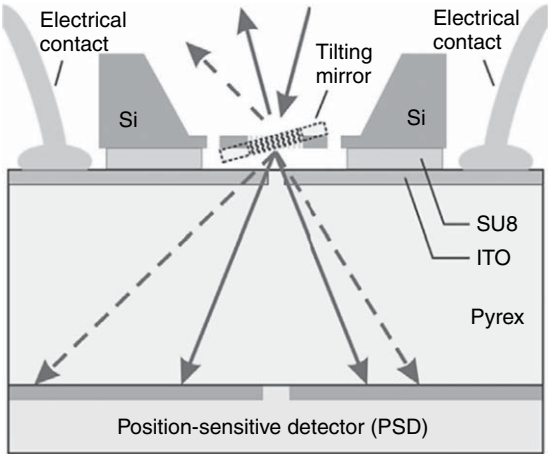


Figure 101 Schematic of the optical tilt sensing mechanism using the mirror with transmission-type grating. (Source: Kallweit D, Zappe H 2006 Fabrication of bulk-Si micromirrors with an integrated tilt sensing mechanism. *J. Micromech. Microeng*. **16**, 463–9.)

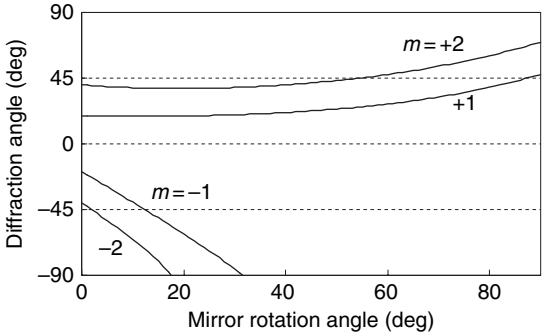


Figure 102 Relationship between the diffraction angle γ and the mirror tilt angle β of the ±first and ±second diffraction orders. The diffraction angle γ is measured relative to the surface normal of the untilted (horizontal) grating.

base of the cantilever for sensing bending and twisting motions of the mirror corresponding to the two-axis scanning. This scanner uses the normal gauge. The resistance change is generated by the longitudinal piezoresistance effect under normal stress.

Another micromirror integrated with a piezoresistive rotation angle sensor uses the shear piezoresistance effect (Sasaki *et al.* 2006a). The shear gauge measures the transverse voltage generated by the shear piezoresistance effect. The mirror rotation generates shear stress inside the torsion bar. Figure 103 shows the basic structure of the sensor. The sensor consists of an element and two pairs of terminals. One pair is for flowing the sensing current I_s. The other is for detecting the signal of the transverse voltage. The signal V_{out} is expressed as follows:

$$V_{out} = \frac{fW\pi_S\tau}{L} V_S \quad [49]$$

where f is a geometric factor that depends on the shape of the gauge. W and L are the width and length of the sensing element, respectively. τ is the shear stress, which is proportional to the mirror rotation angle. V_s is the voltage for supplying the sensing current I_s. π_s is the shear piezoresistance coefficient. This value depends on the doping type, conductivity, and orientation of c-Si. The sensor is prepared from p-type Si. In the device Si layer, a pn junction is prepared. The top surface is the p-type region, inside which the sensing current is confined. The inset of Figure 103 illustrates the considerable distribution of shear stress inside the torsion bar. The shear stress will be maximum at the top surface.

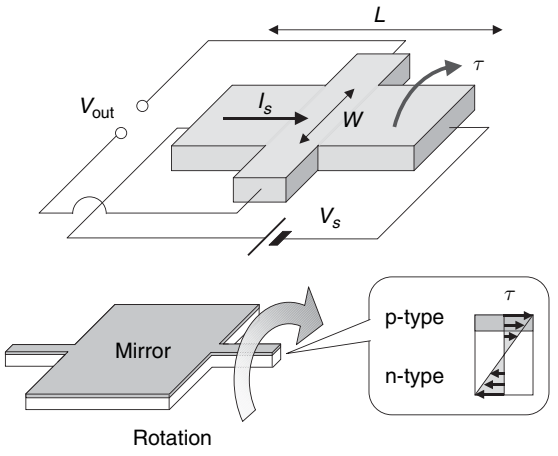


Figure 103 Basic structure of transverse voltage-type gauge for detecting shear stress inside torsion bar for supporting rotating mirror. The inset shows the distribution of the shear stress inside the torsion bar.

Figure 104 shows an embodiment of the combination of a micromirror with a rotation angle sensor. The mirror has through holes, as the study focuses on a proof-of-principle demonstration of the rotation angle sensor simplifying the process. Two sensor structures are included in the left and right torsion bars. The sensing current I_s flows along the torsion bar. The sensor voltage signal is measured through the meandering Si structures without using the metal wiring. These meandering structures have a rotational spring constant that is ~1/10 of that of the torsion bar. The piezoresistance coefficient is known to depend on the crystal orientation. The maximum sensitivity is obtained by aligning the torsion bar along the <100> direction.

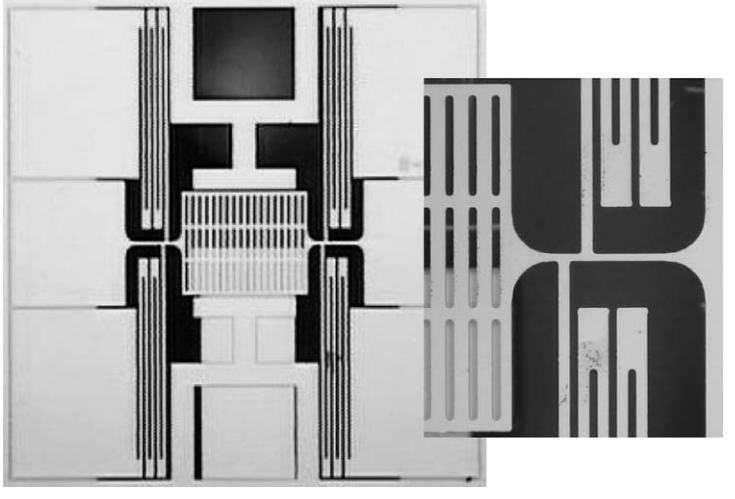


Figure 104 One of fabricated mirror devices. The enlarged image shows the sensor in the torsion bar.

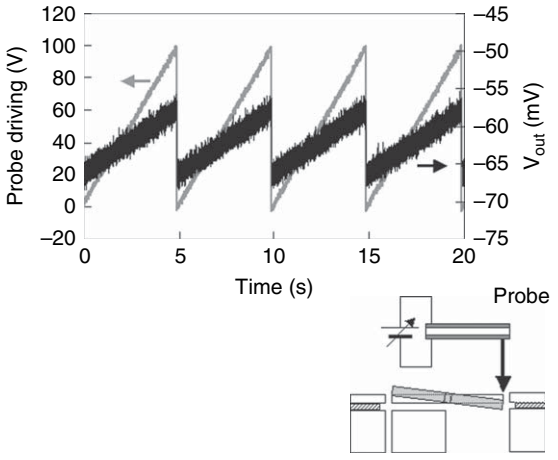


Figure 105 Sensor signal (black) obtained from sensor along <100> direction when mirror is rotated while being pushed by piezoelectric actuator driven by sawtooth wave (gray). The inset shows the experimental setup.

Figure 105 shows a typical signal. The DC sensor signal V_{out} is directly measured using an oscilloscope (input impedance: 1 MΩ) without signal amplification. To increase the sensor signal, a relatively large rotation angle is mechanically introduced by pushing the mirror side using the probe driven by the unimorph piezoelectric actuator, as shown in the inset. The mirror is rotated following the sawtooth wave shown by the gray color. The peak-to-valley of the driving voltage corresponds to 10.2° mirror rotation. The black curve is the sensor signal showing an amplitude of about 10 mV. This signal shows the same sawtooth waveform indicating the linear relation with the mirror rotation angle. Taking advantage of the orientation dependence of the piezoelectric coefficent, the torsion bar along the <100> direction can be sensitive to the mirror rotation but insensitive to the shift motion.

3.01.7.3 Lightweight Flat Mirror

The base micromirror structure can be prepared from the Si structure. The Si layer has little stress (it is known to have slight compress stress) in the SOI wafer. As for the poly-Si film used in surface micromachining, the film stress is known to decrease to nearly zero when poly-Si film (on the Si substrate) is annealed at high temperature (e.g., >1100°C, 30 min) (Guckel *et al.* 1988). However, the micromirror usually needs the metal coating for high reflectivity. In general, the deposited film has stress. The stress

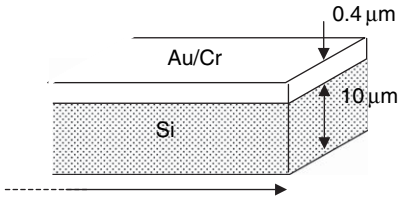


Figure 106 Schematic drawing of the cross section of the micromirror.

control becomes important for decreasing the bending of the structure.

As the simplest example, consider the two-layer structure. Figure 106 shows the cross-sectional view of the model. The substrate layer is assumed to have no stress. The structure is supposed to be a long beam having the same width. The bending moment M, which generates the deflection, is expressed as the sum of two factors.

$$M = \frac{t_f + t_s}{2}\left\{\frac{\sigma}{E_p\left(\frac{1}{E_f w_f} + \frac{1}{E_s w_s}\right)} + \frac{(\alpha_s - \alpha_f)\Delta T}{\frac{1}{E_f w_f} + \frac{1}{E_s w_s}}\right\} \quad [50]$$

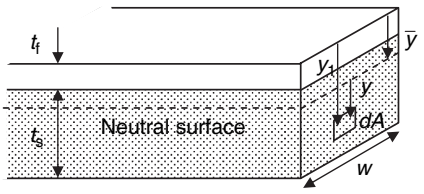
Subscripts f and s show the top film and substrate layer. σ is the intrinsic stress inside the film, E is the Young's modulus, ΔT is the temperature change, w and t are the width and thickness of the layer, respectively, and α is the coefficient of thermal expansion. When the temperature changes, the cantilever deflects generating the moment. This is the mechanism used in the thermal actuator. For making the following discussion simpler, assume ΔT to be negligible. When the beam (or mirror) length is long compared to its deflection, the deflection of the beam can be calculated using the following equation:

$$\frac{d^2 y}{dx^2} = -\frac{M}{EI} = -\frac{1}{R} \quad [51]$$

$$y = -\int\int \frac{M}{EI}\,dx\,dx + C_1 x + C_2 \quad [52]$$

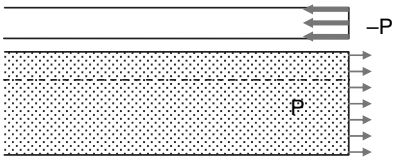
where R is the radius of curvature of the beam and I is the moment of inertia counting the neutral surface.

The position of the neutral surface $\bar{y}$ is decided from the balance of stress requiring the condition that the sum of the stress is 0 over the cross section of the beam. The position y is usually measured from the top surface.



$$\bar{y} = \frac{\sum_i E_i \int_{\text{Area } i} y_i \, dA_i}{\sum_i E_i A_i} = \frac{1}{2} \cdot \frac{E_f t_f^2 + E_s \{(t_f + t_s)^2 - t_f^2\}}{E_f t_f + E_s t_s} \quad [53]$$

The virtual film force P is introduced into the layers. For example, the film under the tensile stress is considered to be fixed to the substrate being pulled with P. In the center, the substrate layer is considered to be fixed to the film being pushed with P. The moment M generated by the force P can be calculated as follows:



$$M = M_f + M_s = w \int_{-\bar{y}}^{t_f - \bar{y}} \frac{-P}{w t_f} y \, dy + w \int_{t_f - \bar{y}}^{t_f + t_s - \bar{y}} \frac{P}{w t_s} y \, dy$$
$$= \frac{1}{2}(t_f + t_s) P \quad [54]$$

The value P can be obtained by the condition of continuity of the stain at the interface. This can be expressed as follows:

$$\frac{\sigma}{E_f} + \alpha_f \Delta T + \frac{P}{E_f w t_f} + \frac{y_{\text{interface}}}{R} = \alpha_s \Delta T - \frac{P}{E_s w t_s} + \frac{y_{\text{interface}}}{R} \quad [55]$$

From eqns [54] and [55], the value of the moment expressed by eqns [50] is obtained.The radius of curvature R can be expressed using eqn [51].

$$R = \frac{\sum_i E_i I_i}{M} \quad [56]$$

The moment of inertia I is counted form the neutral surface. So, the contributions from film layer I_f and substrate layer I_s can be expressed as follows:

$$I_f = w \int_{-\bar{y}}^{t_f - \bar{y}} y^2 \, dy = \frac{w}{3} \{(t_f - \bar{y})^3 - (-\bar{y})^3\} \quad [57]$$

$$I_s = w \int_{t_f - \bar{y}}^{t_s + t_f - \bar{y}} y^2 \, dy = \frac{w}{3} \{(t_s + t_f - \bar{y})^3 - (t_f - \bar{y})^3\} \quad [58]$$

Using the above values, the radius of curvature is obtained as follows:

$$R = \frac{E_f \frac{w}{3} \{(t_f - \bar{y})^3 + \bar{y}^3\} + E_s \frac{w}{3} \{(t_s + t_f - \bar{y})^3 - (t_f - \bar{y})^3\}}{M} \quad [59]$$

For the calculation in this chapter, the values in **Table 9** are used.

Table 9 lists reported values of residual stress (Lammel *et al.* 2002). Au/Cr film is for increasing the reflectivity. With the deposition condition, the film stress may be further decreased. Au is frequently used as it is resistant against HF etching. Au/Cr film generates tensile stress. The maximum bending $y_{\max}$ of peak-to-valley deflection is the difference of the deflection from the center to the edges.

$$y_{\max} = \frac{L^2}{8R} \quad [60]$$

where L is the length of the beam (or mirror). Assuming a base Si layer thickness of 10 μm, the bending profile is calculated. **Figure 107** shows the result. The values of E are the ones averaged for different materials and crystal orientations. Because the beam has a uniform bending moment M and a radius of curvature R, the larger mirror generates the larger deflection depending on the factor of L^2. One method for decreasing $y_{\max}$ is dividing one mirror into segments, making it into a mirror array. The gray curve shows such an example setting with the same value for R. Remember that the allowable deformation of the mirror is expressed by eqn [17]. The tolerance for the stress control and the available material will be wider in the segmented mirror. As for

Table 9 Reported values of residual stress of films

	σ (Mpa)	ρ (kg m^{-3})	E (Gpa)	Note
Au/Cr	325	–	265	$t = 0.4$ μm (Au 0.03 μm, Cr 0.37 μm), average of Au 78 GPa, Cr 280 GPa
Si	0	2.33×10^3	162	Average of <111>188 GPa, <100>130 GPa, <110>169 GPa

Source: Data of σ of Au/Cr: Lammel G, Schweizer S, Renaud P 2002 *Optical Microscanners and Microspectrometers using Thermal Biomorph Actuators*, Kluwer Academic Publishers, Boston, Ch. 2.

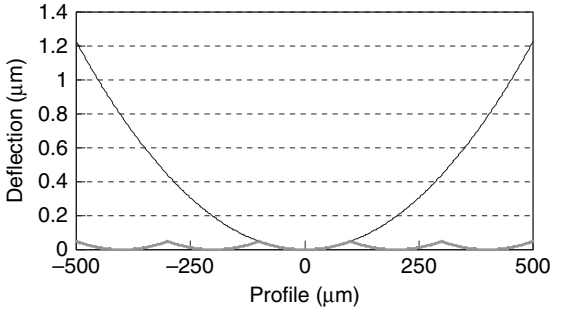


Figure 107 Calculated deflection of the mirror. Black curve shows single mirror. Gray curve shows the profile of the segmented mirrors. The radius of curvature is the same.

the image, the mirror array architecture includes sharp transitions between pixels.

Figure 108(a) shows an example of a small micromirror array for the wavelength add-drop switching developed by Lucent Technologies (Ford *et al.* 1999). The focused light beam is incident on this mirror. The mirror has 3 nm of Cr, followed by 50 nm of Au onto the poly-Si layer. This composition is decided taking into consideration the lower stress. **Figure 108(b)** shows the mirror surface profile. The top is a gray map, indicating the two cross sections below. The curvatures of the cross sections indicate that the stress-induced sag is 20 nm across the full 57 μm size. This is $<\lambda/20$ for the optical communication wavelength of 1.55 μm.

If a single large micromirror is indispensable for the application, the controllable parameter is the substrate thickness t_s in eqn [48]. **Figure 109** shows the three curves changing the substrate thickness. The profile of $t_s = 3.5$ μm is an example of when the micromirror is prepared by the surface micromachining (the radius of curvature is ~4.3 mm). It is clear that the thin Si film easily deflects over $\lambda/4$. For suppressing the deflection $<\lambda/4$, assuming the wavelength to be that of visible light, a substrate thickness of at least ~30 μm will be necessary. The VOA mirror (diameter of 950 μm) from Santec Corporation has a Si thickness of 30 μm. The micromirror from Fujitsu has a Si thickness of 100 μm with the radius of curvature being <30 nm.

Another method for compensating the mirror deflection is depositing additional coatings for canceling the deflection. There is a combination for achieving optical flatness as well as high reflectivity simultaneously (Cao *et al.* 2001). This curvature compensation is possible with a single additional coating layer or with multiple numbers of coating layers. **Figure 110** shows an example before

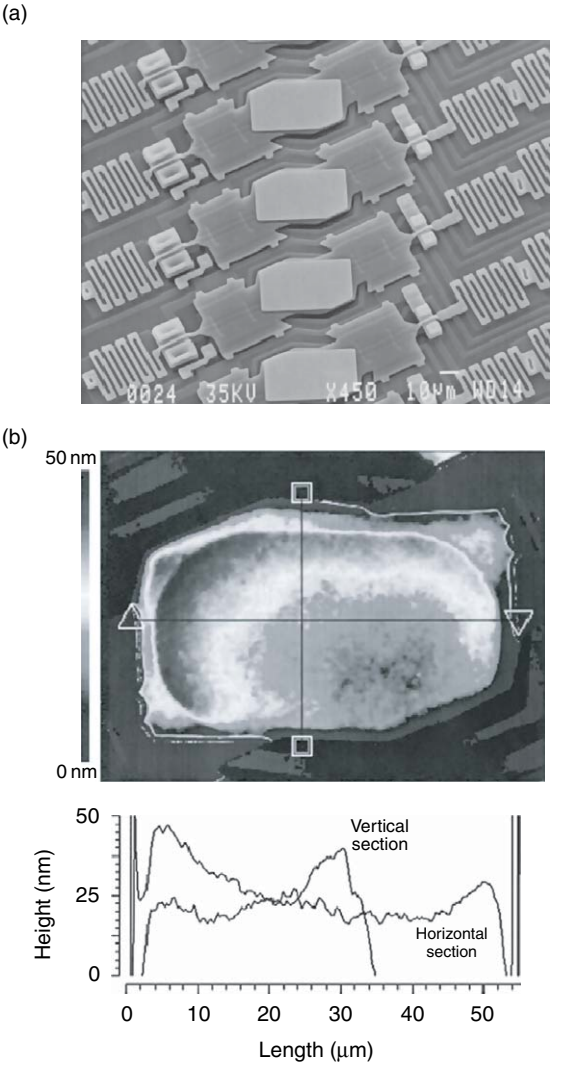


Figure 108 (a) Scanning electron microscope (SEM) image of the gold-coated tilt-mirror array for wavelength add-drop switching prepared by poly-Si surface micromachining. (b) Surface profile of mirror. (Source: Ford J E, Aksyuk V A, Bishop D J, Walker J A 1999 Wavelength add-drop switching using tilting micromirrors. *J. Lightwave Technol.* **17(5)**, 904–11.)

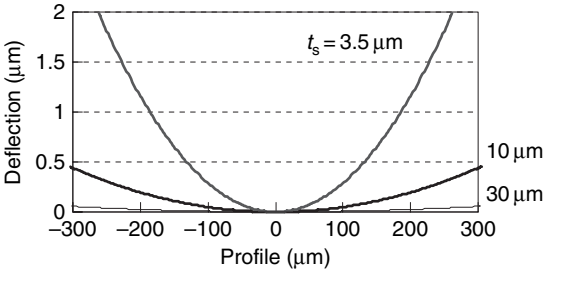


Figure 109 Calculated deflection of the gold-coated micromirror changing the base Si layer thickness.

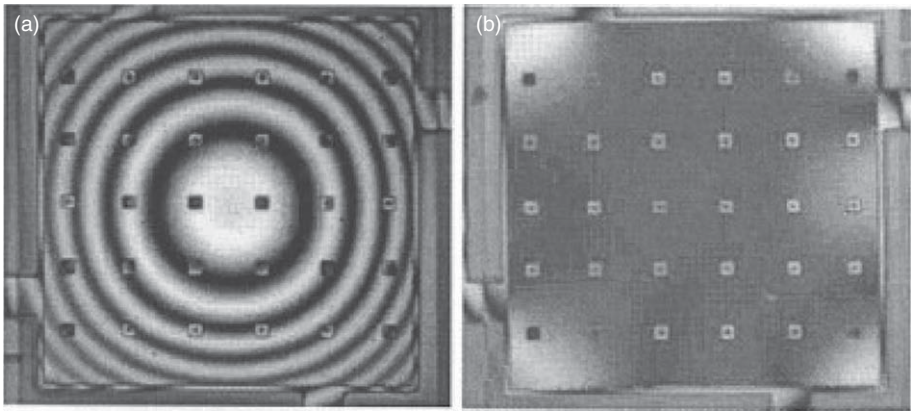


Figure 110 Interferometric images of micromirror (a) before and (b) after the flattening with multilayer deposition of dielectric films. (Source: Cao K, Liu W, Talghader J J 2001 Curvature compensation in micromirrors with high-reflectivity optical coatings. *J. Microelectromech. Syst.* **10(3)**, 409–17.)

(**Figure 110(a)**) and after (**Figure 110(b)**) flattening of the micromirror. Plasma-enhanced chemical vapor deposition is used for coating seven layers of additional SiO_2 and SiN films. The mirror is a $200\,\mu m \times 200\,\mu m \times 3.5\,\mu m$ poly-Si plate with gold film. The images correspond to zero and 175-nm dielectric thickness of top SiO_2 film, respectively. The center-to-edge deflection of the mirror before flattening is 2 μm while that of the flattened mirror is ~45 nm. It must be taken into account that the metal film begins to behave plastically at strains on the order of tenths of a percent. This behavior causes the curvature of a mirror to deviate from the simple prediction. The result shown in **Figure 110** uses the strain of the metal found empirically versus the thickness of dielectric film. As seen from the simple case of eqns [50] and [51], the radius of curvature of the mirror has the sensitivity against the temperature. The technique using the compensation coatings is also applicable for obtaining thermally invariant mirrors (Liu and Talghader 2002).

Although a thicker Si layer is a simple practical solution, there is a trade-off with the resonant frequency. The resonant frequency decreases due to the increase of inertia as expressed by eqn [27]. The flat mirror with the low inertia is ideal. The lightweight flat mirror is prepared by depositing thin films with tensile stress. **Figure 111** shows an embodiment of the

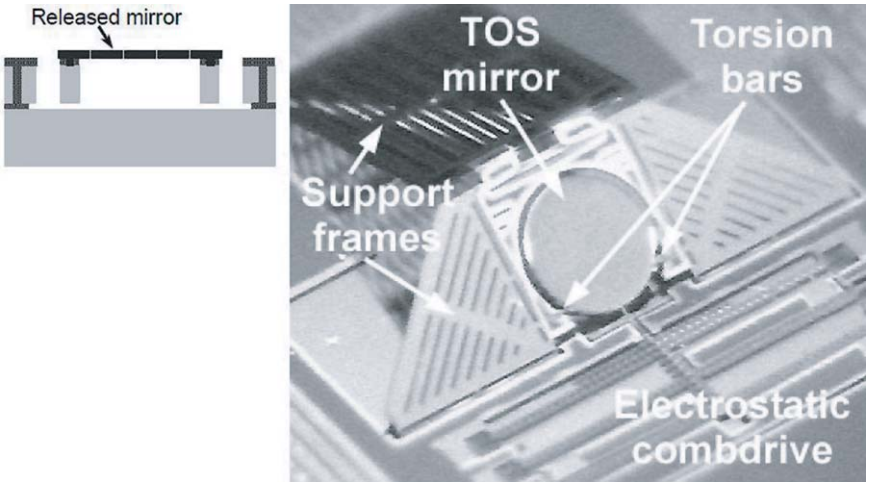


Figure 111 A tensile optical surface micromirror device after release and assembly. The mirror diameter is 550 m. (Source: Nee J T, Conant R A, Lau K Y, Muller R S 2000 Stretched-film micromirrors for improved optical flatness. *Proc. IEEE 13th Annu. Int. Conf. Micro-Electro-Mechanical Systems*, Miyazaki, Japan, pp. 704–9.)

reinforced mirror microstructure reported (Nee *et al.* 2000). A poly-Si membrane is stretched across a rigid c-Si cylinder preparing a drum structure. The tensile stress increases the stiffness of the mirror without increasing its mass significantly. The tensile film gives a flatter surface compared to that given by a standard poly-Si film. The central portion of the mirror is relatively flat (<500 nm deformation over 70% of the cross section).

The mirror with the structure of a honeycomb core is prepared for decreasing the inertia maintaining the solid faceplate as shown in **Figure 112** (Patterson *et al.* 2000). The honeycomb structure consists of c-Si hexagonal cells with sides of 100 μm and wall thickness of 10 μm. These honeycombs are prepared using the timed DRIE from the Si layer of the SOI wafer such that a thin layer remains for forming the mirror face. The two SOI wafers are fusion bonded. **Figure 113** shows the image of the fabricated mirrors. The bonded micromirrors with honeycomb structures (450 μm × 450 μm) show the radius of curvature of >1 m, which is as large as that of the solid mirror.

Timed DRIE can be applied to the SOI micromirrors using a simpler method. **Figure 114** shows two micromirrors (Milanovic *et al.* 2001) described recently. **Figure 114(a)** is the image from the backside of the mirror surface. **Figure 114(b)** is the image from the frontside of the mirror. The mass reduction is carried out from the backside.

3.01.8 Summary

The representative micromirror devices are described focusing on the optical performance. The potential of the micromirror can be seen from the established works and the progress. The micromirror is an advantageous tool for controlling light with high contrast in a small space. The commercial success of DMD from Texas Instruments is well known. This does not mean that DMD has the quality of satisfying

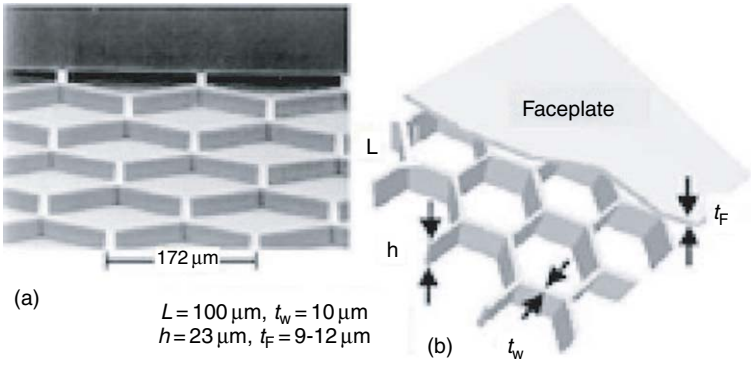


Figure 112 (a) Scanning electron micrograph (SEM) and (b) schematic drawing of micromirror with microhoneycomb structures. (Source: Patterson P R, Su G-D J, Toshiyoshi H, Wu M C 2000 A MEMS 2-D scanner with bonded single-crystalline honeycomb micromirror. *Proc. Int. Workshop Solid-State Sensors and Actuators,* Hilton Head, SC, USA, pp. 17–18.)

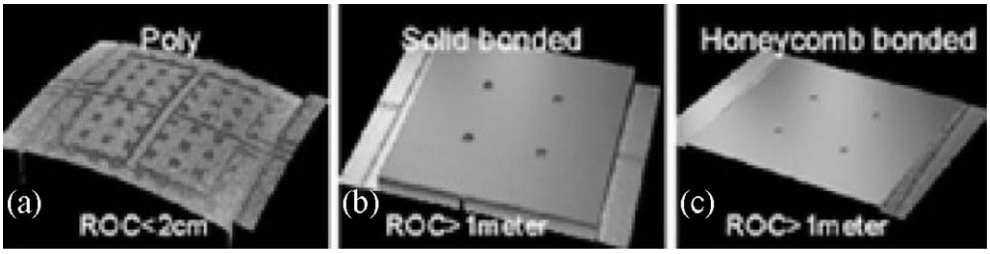


Figure 113 Perspective images of (a) poly-Si, (b) solid bonded, and (c) honeycomb bonded micromirrors measured by interferometer. Radius of curvature is included. (Source: Patterson P R, Su G-D J, Toshiyoshi H, Wu M C 2000 A MEMS 2-D scanner with bonded single-crystalline honeycomb micromirror. *Proc. Int. Workshop Solid-State Sensors and Actuators,* Hilton Head, SC, USA, pp. 17–18.)

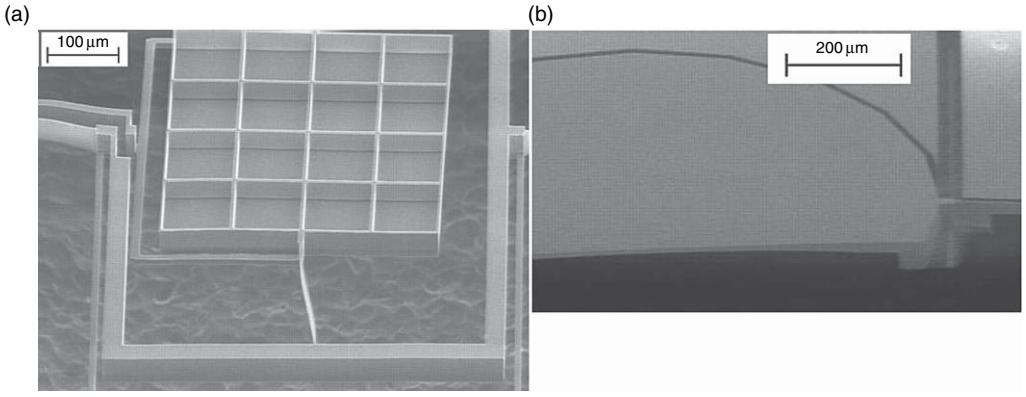


Figure 114 Micromirrors treated with the partial etching for the mass reduction. (Source: Milanovic V, Last M, Pister K S J 2001 Torsional micromirrors with lateral actuators. *Proc. 11th Int. Conf. Solid-State Sensors and Actuators*, pp. 1298–301.)

all the specifications required in general optical applications. Digital light processing is a method for avoiding the disadvantage of the inaccuracy of the rotation angle. Because the display is the application of a man–machine interface, the fast modulation that cannot be detected by the human eye can be allowed. If the device is supposed to work in the core of machines combining with other systems, the required performance of the micromirror will be rather strict. The OXC in the optical network is one such example. A higher accuracy of the mirror rotation angle and higher mirror quality are required. In case of the 238×238 OXC and the optical setup reported from Lucent Technologies, the accuracy of the micromirror rotation angle is estimated to be 0.04° for realizing the reported insertion loss of 0.15 dB (Aksyuk *et al.* 2003). This is the high quality for the micromirror, but the bulk galvanometric scanner guarantees the higher accuracy as shown in **Table 4**. The micromirror is aided with the tolerance generated by the optical setup. OXCs from Calient Networks or Fujitsu Laboratories are now being used in telecommunication networks. VOA from Santec realizes the large attenuation with the long optical arm increasing the change generated by the small mirror rotation angle. GLV is being used in the next generation CTP machines. The parallel light modulator that can deal with the high-power laser beam matches the market. There is still plenty of room for further development. If a higher performance is realized in the micromirror itself, a wide range of applications is possible. In 2006, the mirror array for maskless lithography and the micromirror for laser display were under research. Maskless lithography requires displacement accuracy at the subwavelength level. This has to be satisfied with a large number of mirror arrays. The laser display requires a higher scanning rate with the flat mirror surface having a large diameter. These are projects that are near commercialization. Integrations with ICs, sensors, or other optical elements are well-recognized directions of the technology. These are the longer-term perspectives. The performance of the micromirror will be stabilized realizing the advantages of higher reliability and functionality, and the lower assembly cost as in the case of ICs.

As described, there are many complications in designing and developing micromirrors. The balance of the performances depends on the applications. Nowadays, the developments regarding devices or systems are carried out rather independently. The systematic organization of the problems from the established works is important. In this chapter, the classification of some common problems is discussed. The research based on the systematic organization of the technology will clear the future direction.

References

Akiyama T, Shono K 1993 Controlled stepwise motion in polysilicon microstructures. *J. Microelectromech. Syst.* **2(3)**, 106–10

Akiyama T, Collard D, Fujita H 1997 Scratch drive actuator with mechanical links for self-assembly of three-dimensional MEMS. *J. Microelectromech. Syst.* **6(1)**, 10–17

Aksyuk V A, Pardo F, Bolle C A, Arney S, Giles C R, Bishop D J 2000 Lucent Microstar micromirror array technology for large optical crossconnects. *Proc. SPIE* **4178**, 320–4

Aksyuk V A, Arney S, Basavanhally N R, Bishop D J, Bolle C A, Chang C C, Frahm R, Gasparyan A, Gates J V, George R, Giles C R, Kim J, Kolodner P R, Lee T M, Neilson D T, Nijander C, Nuzman C J, Paczkowski M, Papazian A R,

Pardo F, Ramsey D A, Ryf R, Scotti R E, Shea H, Simon M E 2003 238 × 238 micromechanical optical cross connect. *IEEE Photonics Techonol. Lett.* **15**, 587–9

Anthon D, Berger J D, Drake J, Dutta S, Fennema A, Grade J D, Hrinya S, Ilkov F, Jerman H, King D, Lee H, Tselikov A, Yasumura K 2002 External cavity diode lasers tuned with silicon MEMS. *Technical Digest, Optical Fiber Communication Conference,* Paper TUO7, pp. 97–8

Asada N, Matsuki H, Minami K, Esashi M 1994 Silicon micromachined two-dimensional galvano optical scanner. *IEEE Trans. Magn.* **30(6)**, 4647–9

Bartels R, Backus S, Zeek E, Misoguti L, Vdovin G, Christov I P, Murnane M M, Kapteyn H C 2000 Shaped-pulse optimization of coherent emission of high-harmonic soft X-rays. *Nature* **406**, 164–6

Behrang B, Satinderpall P 2001 *US Pat.* 6 629 461

Bell D J, Lu T J, Fleck N A, Spearing S M 2005 MEMS actuators and sensors: Observations on their performance and selection for purpose. *J. Micromech. Microeng.* **15**, S153–64

Bifano T G, Mali R K, Dorton J K, Perreault J, Vandelli N, Horenstein M N, Castanon D A 1997 Continuous-membrane surface-micromachined silicon deformable mirror. *Opt. Eng.* **36**, 1354–60

Bourouina T, Lebrasseur E, Reyne G, Debray A, Fujita H, Ludwig A, Quandt E, Muro H, Oki T, Asaoka A 2002 Integration of two degree-of-freedom magnetostrictive actuation and piezoresistive detection: Application to a two-dimensional optical scanner. *J. Microelectromech. Syst.* **11**, 355–61

Bruce E, Riezenman M J 2002 Tunable lasers. *IEEE Spectrum.* **39(2)**, 35–39

Bustillo J M, Howe R T, Muller R S 1998 Surface micromachining for microelectromechanical systems. *Proc. IEEE* **86(8)**, 1552–74

Cao K, Liu W, Talghader J J 2001 Curvature compensation in micromirrors with high-reflectivity optical coatings. *J. Microelectromech. Syst.* **10(3)**, 409–17

Conant R A, Nee J T, Lau K Y, Muller R S 2000 A flat high-frequency scanning micromirror. *Proc. Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, June 4–8, pp. 6–9

Doble N, Yoon G, Chen L, Bierden P, Singer B, Olivier S, Williams D R 2002 Use of a microelectromechanical mirror for adaptive optics in the human eye. *Opt. Lett.* **27**, 1537–9

Douglass M R 2003 DMD reliability: A MEMS success story. *Proc. SPIE* **4980**, 1–11

Ezoe Y, Koshiishi M, Mita M, Mitsuda K, Hoshino A, Ishisaki Y, Yang Z, Takano T, Maeda R 2006 Micro pore X-ray optics using anisotropic wet etching of (110) silicon wafer. *Appl. Opt.,* **45(35)**, 8932–8

Ford J E, Aksyuk V A, Bishop D J, Walker J A 1999 Wavelength add-drop switching using tilting micromirrors. *J. Lightwave Technol.* **17(5)**, 904–11

Greywall D S, Pai C-S, Oh S-H, Chang C-P, Marom D M, Busch P A, Cirelli R A, Taylor J A, Klemens F P, Sorsch T W, Bower J E, Lai W Y-C, Soh H T 2003 Monolithic fringe-field-activated crystalline silicon tilting-mirror devices. *J. Microelectromech. Syst.* **12(5)**, 702–7

Guckel H, Burns D W, Visser C C G, Tilmans H A C, Deroo D 1988 Fine-grained polysilicon films with built-in tensile strain. *IEEE Trans. Electron Devices* **35(6)**, 800–1

Guldberg J, Nathanson H C, Balthis D L, Jensen A S 1975 An aluminum/SiO_2/silicon-on-sapphire light valve matrix for projection displays. *Appl. Phys. Lett.* **26(7)**, 391–3

Hagelin P M, Solgaard O 1999 Optical raster-scanning displays based on surface-micromachined polysilicon mirrors. *IEEE J. Sel. Top. Quant. Electron.* **5(1)**, 67–74

Hah D, Patterson P R, Nguyen H D, Toshiyoshi H, Wu M C 2004a Theory and experiments of angular vertical comb-drive actuators for scanning micromirrors. *IEEE J. Sel. Top. Quant. Electron.* **10(3)**, 505–13

Hah D, Huang S T, Tsai J, Toshiyoshi H, Wu M C 2004b Low-voltage, large-scan angle MEMS analog micromirror arrays with hidden vertical comb-drive actuators. *J. Microelectromech. Syst.* **13**, 279–89

Hart M R, Conant R A, Lau K Y, Muller R S 1999 Time-resolved measurement of optical MEMS using stroboscopic interferometry. *Proc. 10th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '99)*, Sendai, Japan, pp. 470–3

Hecht E 2002 *Optics*. Person Education Inc., San Francisco, CA, Chap. 10

Isamoto K, Kato K, Morosawa A, Chong C, Fujita H, Toshiyoshi H 2004 A 5-V operated MEMS variable optical attenuator by SOI bulk micromachining. *IEEE J. Sel. Top. Quant. Electron.* **10(3)**, 570–8

Ishimori M, Sasaki M, Hane K 2007 Micromirror with parallel two rotation axes for external cavity diode laser. *IEICE Trans. Electron.,* **E90-C(1)**, 72–7

Jain A, Qu H, Todd S, Xie H 2005 A thermal bimorph micromirror with large bi-directional and vertical actuation. *Sens. Actuators A Phys.* **122**, 9–15

Jerman H, Grade J D 2002 A mechanically-balanced, DRIE rotary actuator for a high-power tunable laser. *Proc. Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, USA, pp. 7–10

Johnson D, Westerman R, Lai S 2005 *US Pat.* 6 924 235

Kallweit D, Zappe H 2006 Fabrication of bulk-Si micromirrors with an integrated tilt sensing mechanism. *J. Micromech. Microeng.* **16**, 463–9

Kiang M-H, Solgaard O, Lau K Y, Muller R S 1998a Polysilicon optical microscanners for laser scanning displays. *Sens. Actuators A Phys.* **70(1–2)**, 195–9

Kiang M-H, Solgaard O, Lau K Y, Muller R S 1998b Electrostatic combdrive-actuated micromirrors for laser-beam scanning and positioning. *J. Microelectromech. Syst.* **7(1)**, 27–37

Kobayashi T, Ichiki M, Tsaur J, Maeda R 2005a Effect of multi-coating process on the orientation and microstructure of lead zirconate titanate (PZT) thin films derived by chemical solution deposition. *Thin Solid Films* **489**, 74–8

Kobayashi T, Tsaur J, Maeda R 2005b Fabrication of optical micro scanner driven by PZT actuators. *Jpn. J. Appl. Phys.* **44(9B)**, 7078–82

Krishnamoorthy U, Li K, Yu K, Lee D, Heritage J P, Solgaard O 2002 Dual-mode micromirrors for optical phased array applications. *Sens. Actuators A Phys.* **97–98**, 21–6

Lammel G, Schweizer S, Renaud P 2002 *Optical Microscanners and Microspectrometers Using Thermal Bimorph Actuators*. Kluwer Academic Publishers, Boston

Lazar J, Cip O, Ružiča B 2004 The design of a compact and tunable extended-cavity semiconductor laser. *Meas. Sci. Technol.* **15**, N6–9

Lee D, Krishnamoorthy U, Yu K, Solgaard O 2004 Single-crystalline silicon micromirrors actuated by self-aligned vertical electrostatic combdrives with piston-motion and rotation capability. *Sens. Actuators A Phys.* **114**, 423–8

Lee M-C M, Yao J, Wu M C 2005 Silicon profile transformation and sidewall roughness reduction using hydrogen annealing. *Proc. 18th IEEE Int. Conf. Micro Electro Mechanical Systems*, Miami Beach, FL, USA, pp. 596–9

Lin L Y, Goldstein E L, Tkach R W 1999 Free-space micromachined optical switches for optical networking. *IEEE J. Sel. Top. Quant. Electron.* **5(1)**, 4–9

Liu W, Talghader J J 2002 Thermally invariant dielectric coatings for micromirrors. *Appl. Opt.* **41(16)**, 3285–93

Ljungblad U, Martinsson H, Sandstrom T 2005 Phase shifted addressing using a spatial light modulator. *Microelectron. Eng.* **78–9**, 398–403

Lopez D, Simon M E, Pardo F, Aksyuk V, Klemens F, Cirelli R, Neilson D T, Shea H, Sorsch T, Ferry E, Nalamasu O, Gammel P L 2002 Monolithic MEMS optical switch with amplified out-of-plane angular motion. *Proc. IEEE/LEOS Int. Conf. Optical MEMS*, pp. 165–6

Mali R K, Bifano T, Vandelli N 1997 Development of MEMS deformable mirrors for phase modulation of light. *Opt. Eng.* **36(2)**, 542–8

Martinsson H, Sandstrom T, Bleeker A, Hintersteiner J D 2005 Current status of optical maskless lithography. *J. Microlith. Microfab. Microsyst* **4(1)**, 011003-1–011003-15

Marxer C, Thio C, Gretillat M-A, de Rooij N F, Battig R, Anthamatten O, Valk B, Vogel P 1997 Vertical mirrors fabricated by deep reactive ion etching for fiber-optic switching applications. *J. Microelectromech. Syst.* **6(3)**, 277–85

Marxer C, de Rooij N F 1999 Micro-opto-mechanical 2×2 switch for single-mode fibers based on plasma-etched silicon mirror and electrostatic actuation. *J. Lightwave Technol.* **17(1)**, 2–6

Milanovic V, Last M, Pister K S J 2001 Torsional micromirrors with lateral actuators. *Proc. 11th Int. Conf. Solid-State Sensors and Actuators*, Munich, Germany, pp. 1298–301

Milanovic V, Last M, Pister K S J 2003 Laterally actuated torsional micromirrors for large static deflection. *IEEE Photonics Technol. Lett.* **15(2)**, 245–7

Miyajima H, Asaoka N, Arima M, Minamoto Y, Murakami K, Tokuda K, Matsumoto K 2001 A durable, shock-resistant electromagnetic optical scanner with polyimide-based hinges. *J. Microelectromech. Syst.* **10(3)**, 418–24

Miyajima H, Asaoka N, Isokawa T, Ogata M, Aoki Y, Imai M, Fujimori O, Katashiro M, Matsumoto K 2003 A MEMS electromagnetic optical scanner for a commercial confocal laser scanning microscope. *J. Microelectromech. Syst.* **3(12)**, 243–51

Miyajima H, Murakami K, Katashiro M 2004 MEMS optical scanners for microscopes. *IEEE J. Selected Topics in Quantum Electronics* **10(3)**, 514–27

Muller R S, Lau K Y 1998 Surface-micromachined microoptical elements and systems. *Proc. IEEE* **86(8)**, 1705–20

Nee J T, Conant R A, Lau K Y, Muller R S 2000 Stretched-film micromirrors for improved optical flatness. *Proc. IEEE 13th Annu. Int. Conf. Micro-Electro-Mechanical Systems*, Miyazaki, Japan, pp. 704–9

Nee J T 2001 Hybrid surface-/bulk-micromachining processes for scanning micro-optical components. Ph.D. thesis, University of California, Berkeley

Noell W, Clerc P-A, Dellmann L, Guldimann B, Herzig H-P, Manzardo O, Marxer C R, Weible K J, Dändliker R, de Rooij N 2002 Applications of SOI-based optical MEMS. *IEEE J. Sel. Top. Quant. Electron.* **8**, 148–54

Orcutt J W, Keller R C, Melendez J L, Bartholomew D 2001 *US Pat.* 6 714 336

Park S E, Kwon T Y, Shin E, Lee H S 2003 A compact extended-cavity diode laser with a Littman configuration. *IEEE Trans. Instrum. Meas.* **2(52)**, 280–3

Patterson P R, Su G-D J, Toshiyoshi H, Wu M C 2000 A MEMS 2-D scanner with bonded single-crystalline honeycomb micromirror. *Proc. Int. Workshop Solid-State Sensors and Actuators*, Hilton Head Island, SC, USA, pp. 17–18

Payne A, DeGroot W, Monteverde R, Amm D 2004 Enabling high-data-rate imaging applications with grating light valve technology. *Proc. SPIE,* **5348**, 76–88

Perry T S 2004 Tomorrow's TV. *IEEE Spectrum* 38–41

Perreault J A, Bifano T G, Levine B M, Horenstein M N 2002 Adaptive optic correction using microelectromechanical deformable mirrors. *Opt. Eng.* **41(3)**, 561–6

Petersen K E 1977 Micromechanical light modulator array fabricated on silicon. *Appl. Phys. Lett.* **31(8)**, 521–3

Petersen K E 1980 Silicon torsional scanning mirror. *IBM J. Res. Develop.* **24(5)**, 631–7

Petersen K E 1982 Silicon as a mechanical material. *Proc. IEEE* **70(5)**, 420–57

Rai-Choudhury P 2000 *MEMS and MOEMS Technology and Applications*. SPIE Press, Bellingham, WA

Sasaki M, Miura K, Hane K, Minami K 2002 Tunable fiber Bragg grating combined with microactuator. *Jpn. J. Appl. Phys.* **41(6B, Part 1)**, 4356–61

Sasaki M, Fujii T, Hane K 2003a Anisotropic Si etching condition for preparing optically smooth surfaces. *Sens. Mater.* **15(2)**, 83–92

Sasaki M, Yamaguchi T, Song J H, Hane K, Hara M, Hori K 2003b Optical scanner on a three-dimensional microoptical bench. *J. Lightwave Technol.* **21(3)**, 602–8

Sasaki M, Tabata M, Haga T, Hane K 2006a Piezoresistive rotation angle sensor integrated in micromirror. *Jpn. J. Appl. Phys* **45(4B)**, 3789–93

Sasaki M, Yuki S, Hane K 2006b Large-rotation and low-voltage driving of micromirror realized by tense thin film torsion bar. *IEEE Photonics Technol. Lett.* **18(15)**, 1573–5

Schweizer S, Calmes S, Laudon M, Renauld Ph 1999 Thermally actuated optical microscanner with large angle and low consumption. *Sens. Actuators* **76**, 470–7

Schweizer S, Cousseau P, Lammel G, Calmes S, Renauld Ph 2000 Two-dimensional thermally actuated optical microprojector. *Sens. Actuators* **85**, 424–9

Senturia S D 2001 *Microsystem Design*. Springer Science+Business Media, New York, NY

Shao Y, Dickensheets D L, Himmer P 2004 3-D MOEMS mirror for laser beam pointing and focus control. *IEEE J. Sel. Top. Quant. Electron.* **10**, 528–35

Solgaard O, Sandejas F S A, Bloom D M 1992 Deformable grating optical modulator. *Opt. Lett.* **17**, 688–90

Sontheimer A B 2002 Digital micromirror device (DMD) hinge memory lifetime reliability modeling. IEEE 02CH37320. 40th Annual International Reliability Physics Symposium, Dallas, Texas, pp. 118–21

Sprague R, Montague T, Brown D 2005 Bi-axial magnetic drive for scanned beam display mirrors. *Proc. SPIE* **5721**, 1–13

Su G-D J, Fan L, Wu M C 1999 Surface-micromachined adaptive micromirror arrays with large strokes. *Proc. 10th Int. Conf. Solid-State Sensors and Actuators*, pp. 578–81

Tamakia E, Hashimoto Y, Leung O 2004 Computer-to-plate printing using the grating light valve device. *Proc. SPIE* **5348**, 89–97

Thomas R N, Guldberg J, Nathanson H C, Malmberg P R 1975 The mirror-matrix tube: A nonel light valve for projection displays. *IEEE Trans. Electron Devices* **ED-22(9)**, 765–75

Trisnadi J I, Carlisle C B, Monteverde R 2004 Overview and applications of grating-light-valve-based optical write engines for high-speed digital imaging. *Proc. SPIE* **5348**, 52–64

Tsuboi O, Mizuno Y, Koma N, Soneda H, Okuda H, Ueda S, Sawaki I, Yamagishi F 2002 A rotational comb-driven micromirror with a large deflection angle and low drive voltage. *Proc. 15th Int. Conf. on Micro Electro Mechanical Systems*, Las Vegas, NV, USA, pp. 532–5

Urey H 2004 MEMS scanners for display and imaging applications. *Proc. SPIE* **5604**, 218–29

Urey H, Wine D W, Lewis J R 1999a Scanner design and resolution tradeoffs for miniature scanning displays. *Proc. SPIE* **3636**, 60–8

Urey H, Nestorovic N, Ng B, Gross A 1999b Optics designs and system MTF for laser scanning displays. *Proc. SPIE* **3689**, 238–48

Van Kesel P F, Hornbeck L J, Meier R E, Douglass M R 1998 A MEMS-based projection display. *Proc. IEEE* **86(8)**, 1687–704

Vdovin G, Middelhoek S 1995 Deformable mirror display with continuous reflecting surface micromachined in silicon. *Proc. IEEE Workshop on Micro-Electro-Mechanical Systems*, Amsterdam, The Netherlands, pp. 61–5

Vdovin G, Sarro L 1994 Flexible reflecting membranes micromachined in silicon. *Semicond. Sci. Technol* **9**, 1570–2

Wagner B, Kreutzer M, Benecke W 1991 Electromagnetic microactuators with multiple degrees of freedom. *Proc. 6th Int. Conf. Solid-State Sensors, Actuators and Microsystems (Transducers '91)*, San Francisco, CA, USA, pp. 614–17

Yalcinkaya A D, Urey H, Brown D, Montague T, Sprague R 2006 Two-Axis electromagnetic Microscanner for high resolution displays. *J. Microelectromech. Syst.* **15(4)**, 786–794

Yamada K, Kuriyama T 1998 A novel asymmetric silicon micro-mirror for optical beam scanning display. *Proc. IEEE Workshop on Micro-Electro-Mechanical Systems*, Heidelberg, Germany, pp. 110–15

Yamamoto T, Yamaguchi J, Takeuchi N, Shimizu A, Higurashi E, Sawada R, Uenishi Y 2003 A three-dimensional MEMS optical switching module having 100 input and 100 output ports. *IEEE Photonics Technol. Lett.* **15(10)**, 1360–2

Yano M, Yamagishi F, Tsuda T 2005 IEEE optical MEMS for photonic switching – Compact and stable optical crossconnect switches for simple, fast, and flexible wavelength applications in recent photonic networks. *J. Sel. Top. Quant. Electron.* **11(2)**, 383–94

Yasuda Y, Akamatsu M, Tani M, Iijima T, Toshiyoshi H 2006 Piezoelectric 2D-optical micro scanners with PZT thick films. *Integrated Ferroelectrics* **80**, 341–53

Yeh J-L A, Jiang H, Tien N C 1999 Integrated polysilicon and DRIE bulk silicon micromachining for an electrostatic torsional actuator. *J. Microelectromech. Syst.* **48(4)**, 456–65

Biographies

Minoru Sasaki received the M.S. and Dr. Eng. degrees from Nagoya University in 1993 and 1995, respectively. In 1996, he was a Research Fellow of the Japan Society for the Promotion of Science. From 1997 to 2007, he worked as a member of the Department of Nanomechanics, Tohoku University. Since 2007, he has been a professor of the Department of Advanced Science and Technology, Toyota Technological Institute, and is currently engaged in the research on Optical MEMS including three-dimensional lithography.

Kazuhiro Hane received the M.S. and Dr. Eng. degrees from Nagoya University in 1980 and 1983, respectively. From 1983 to 1994, he worked as a member of the Department Electronic-Mechanical Engineering, Nagoya University. From 1985 to 1986, he was a guest researcher of National Research Council of Canada. Since 1994, he has been a professor of the Department of Nanomechanics, Tohoku University, and is currently engaged in the research on microsensors and opto-mechanical systems.

3.02 Diffractive and Defractive Micro-optics

Stefan Sinzinger, Technische Universität Ilmenau, Ilmenau, Germany

Nomenclature

$\tilde{\boldsymbol{A}}(\nu)$	Array factor of a diffraction grating	$\boldsymbol{d}$	Distance between the electrodes of a capacitor
$\boldsymbol{a}$	Extension of an image formed by the individual microlenses of a microlens array	$\delta(\boldsymbol{x})$	δ distribution
α	Parameter describing a GRIN profile	ε, ε_1, ε_2	Dielectric constants of different materials
$\alpha_{\mathbf{diff}}$, $\Theta_{\mathbf{d}}$	Diffraction angles	$\varepsilon_\perp$, $\varepsilon_{\|\|}$	Dielectric constants of different planes of polarization
$\boldsymbol{C}$	Capacitance	λ, λ_1, λ_2	Wavelength
$\boldsymbol{D}$	Diameter of a microlens	ν_x, ν_y	Spatial frequencies
Δ	Size of a pinhole	$\boldsymbol{F}$, $\boldsymbol{f}$	Focal lengths
		$\tilde{\boldsymbol{F}}(\nu)$	Form factor of a diffraction grating

$g(x)$	Complex transmission function of a diffraction grating
$\tilde{G}(\nu)$	Fourier transform (diffraction amplitudes) of a diffraction grating
γ_{sl}	Surface energy of a surface–liquid interface
h	Sag of a lens
k	Propagation vector of an optical wavefront
l, Δl	Physical path lengths
n, $n(\lambda)$, $n(x)$	Refractive index
NA	Numerical aperture
N_F	Fresnel number
p, $p_{microlens}$, $p_{pinhole}$	Period of a grating, pitch of a microlens, and pinhole
R, r	Distance vector, radius
r_C	Radius of curvature
t	Thickness of a diffraction grating
$T_{element}$, T_{prism}, T_{lens}	Complex transmission function of optical elements
u	Aperture angle
U	Voltage
$U(x, y, z)$	Complex amplitude of a wave field
V_0	Voltage applied to a liquid crystal microlens
V_{cyl}, V_{cap}	Volume of specific elements
w_{min}	Minimum feature size
w	Width of a grating bar
X_C	Complex resistance of a capacitor
X_R	Complex resistance of a resistor
x, y, z, x_0, y_0, z_0	Spatial coordinates
z_T	Talbot distance

Glossary

Active microlenses Microlenses with actively controlled optical performance

CGH Computer-Generated Hologram; diffraction grating designed numerically for the generation of a specific diffractive behavior

Chirped microlens array Array of microlenses with locally varying optical properties

Diffraction limit The best performance one can expect from optical elements or systems due to diffraction at the physical aperture; aberrations are small enough to be neglected

Diffractive lens Diffraction grating with spherical symmetry, designed to collimate or focus an incoming wavefront

DOE Diffractive Optical Element, a microstructured optical element designed for a specific performance in the diffraction plane

Electrowetting Change in the wetting angle of polar liquids by application of electric fields

FOV Field of View; the angular range in which an optical system provides good imaging

Fraunhofer diffraction Far-field diffraction; the relation between the object amplitude distribution and the diffracted wavefront is calculated by a Fourier transform

Fresnel diffraction Near-field diffraction; the relation between the object amplitude distribution and the diffracted wavefront is calculated by a Fresnel transform

Fresnel number Number of Fresnel rings within an aperture determined by the curvature of the wavefront inside the aperture

GRIN microlenses Gradient index microlenses

Huygens' principle Explanation of the phenomenon of diffraction by propagation of elementary spherical waves interfering to form the respective complex diffracted wavefronts

Kirchhoff diffraction integral Mathematical formulation of the Huygens' principle

Liquid microlens Microlens fabricated of liquid material enclosed by a flexible membrane-like material; by deformation of the container, the optical properties of the lens can be tuned

OPL Optical path length; physical path length scaled with the refractive index of the medium through which the light propagates

Photon sieve Specific type of diffractive lens consisting of a number of holes in a nontransparent screen

PSF Point Spread Function; distribution of the image of a single object point in the image plane

Reflow microlens Microlens fabricated by melting photoresist

SBP Space-Bandwidth Product; the number of resolvable points (pixels) in an optical image

Scalar diffraction theory Approximation to the rigorous diffraction by neglecting the vectorial character of light (polarization); diffraction in the scalar domain is described by the Kirchhoff diffraction integral

Talbot effect Self-imaging effect of optical signals with lateral periodicity; an optical wave field with lateral periodicity also shows longitudinal periodicity

TEA Thin Element Approximation; in the TEA the interaction of a wavefront with an object takes place in a single plane; multiple interactions are neglected

Thick diffraction grating grating described by Bragg reflection; holographic grating

Zero-order grating High-frequency diffraction grating (periods smaller than the wavelength) with only one propagating diffraction order (zeroth order); due to the small period all nonzero orders are evanescent

3.02.1 Diffractive Micro-Optics

One of the most prominent effects of wave propagation phenomena, i.e., in our case of light propagation, is diffraction. It is, for example, responsible for the deviation of the propagation direction from the geometrical expectations behind obstacles. The most straightforward way to understand diffraction is the so-called Huygens' principle. The famous Dutch mathematician and physicist Christian Huygens (1629–1695) explained diffraction phenomena by stating that every point of a wavefront (i.e., a plane of equal phase) can be considered as the origin of a spherical elementary wave. After independent propagation, the interference of these spherical waves forms subsequent wave fields (**Figure 1**). In combination with the periodicity of the wave and the laws of interference this principle leads to an understanding of light propagating not only into the geometrical shadow regions behind diffraction objects but also to the discrete propagation directions behind periodic objects called diffraction gratings.

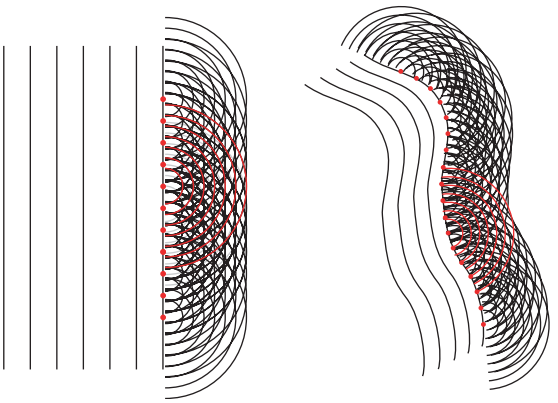


Figure 1 Illustration of the Huygens' principle. Diffraction is explained by the generation of spherical elementary waves propagating independently and interfering to form the diffracted wavefront.

3.02.1.1 Kirchhoff Diffraction Integral and Scalar Diffraction Theory

Mathematically, the Huygens' principle is formulated based on the Kirchhoff diffraction integral (Goodman 1996)

$$U(x, y, z) \propto \frac{\mathrm{i}}{\lambda} \iint U(x_0, y_0, z_0) \frac{\mathrm{e}^{\mathrm{i}|\underline{k}||\underline{r}|}}{|\underline{r}|} \cos\varepsilon \, \mathrm{d}x_0 \mathrm{d}y_0 \quad [1]$$

Here, $U(x, y, z)$ is the wave field at a distance $(z - z_0)$ behind the diffracting plane (x_0, y_0, z_0) with the wave field distribution $U(x_0, y_0, z_0)$; ε is the propagation angle. The term $(\exp(\mathrm{i}|\underline{k}||\underline{r}|))/|\underline{r}|$ represents the spherical wavefronts ($\underline{k}$ is the propagation vector of the wavefront $|\underline{k}| = (2\pi/\lambda)$) which are weighted through the complex amplitude of the incident wave field $U(x_0, y_0, z_0)$ and subsequently added (integrals). This diffraction integral can be derived rigorously by solving the wave equation and the resulting Debye–Sommerfeld diffraction integral, for example, through the method of stationary phase (Lohmann 2006, pp. 111–18).

3.02.1.1.1 Fresnel and Fraunhofer diffraction

The Kirchhoff diffraction integral is the foundation for the scalar diffraction theory where the vector character, that is, the polarization of the light wave, is neglected, and the distance between the plane of observation and the diffracting plane is considered to be much larger than the wavelength. Further approximations lead to the Fresnel and Fourier diffraction integrals, which define the regions of Fresnel and Fraunhofer diffraction. To this end, the distance

vector $|\underline{r}|$ is approximated through the following equation:

$$\begin{aligned}|\underline{r}| &= \sqrt{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2} \\ &\approx \underbrace{(z-z_0)\left[1 + \frac{1}{2}\left(\frac{x-x_0}{z-z_0}\right)^2 + \frac{1}{2}\left(\frac{y-y_0}{z-z_0}\right)^2\right]}_{\text{Fresnel approximation}} \\ &\approx \underbrace{(z-z_0)\left[1 + \frac{x^2-2xx_0}{2(z-z_0)^2} + \frac{y^2-2yy_0}{2(z-z_0)^2}\right]}_{\text{Fraunhofer approximation}} \end{aligned} \quad [2]$$

where it has been assumed in the Fraunhofer approximation that $x_0 << x$ and $y_0 << y$, i.e., the extension of the diffraction screen is much smaller than the extension of the observation plane, which is plausible for large distances from the diffraction screen.

The diffraction integrals resulting from the application of these approximations to the Kirchhoff integral are the Fresnel diffraction integral

$$\begin{aligned}U(x, y, z) \approx \frac{i}{\lambda}\frac{\exp(ik(z-z_0))}{z-z_0}\iint U(x_0, y_0, z_0) \\ \times \exp\left(\frac{i\pi}{\lambda(z-z_0)}[(x-x_0)^2 \right. \\ \left. + (y-y_0)^2]\right)dx_0dy_0 \end{aligned} \quad [3]$$

and the Fraunhofer diffraction integral

$$\begin{aligned}U(x, y, z) \approx \frac{i}{\lambda}\frac{\exp(ik(z-z_0))}{z-z_0} \\ \times \exp\left(\frac{i\pi(x^2+y^2)}{\lambda(z-z_0)}\right)\iint U(x_0, y_0, z_0) \\ \times \exp\left(2i\pi\frac{xx_0+yy_0}{\lambda(z-z_0)}\right)dx_0dy_0 \end{aligned} \quad [4]$$

One of the remarkable effects in Fresnel diffraction (eqn [3]) is the so-called Talbot self-imaging. This effect, which is sometimes referred to as Talbot effect directly, results from diffraction theory according to the Fresnel diffraction integral. As such, the term Talbot imaging or self-imaging is rather misleading. For optical elements that exhibit a lateral periodicity (e.g., gratings or microlens arrays), the diffracted field in the Fresnel approximation is periodic also in propagation direction. At specific distances, mz_T, the so-called Talbot distances, behind the periodic object with period p, the structure of the diffraction grating is reconstructed even without any imaging elements applied:

$$z_T = \frac{2p^2}{\lambda} \quad [5]$$

Generally speaking, according to the Talbot effect, a lateral periodicity in the wave field causes a longitudinal periodicity along the direction of propagation. This has important effects such as on the use of array of micro-optical components, for example, of microlens arrays, where it is quite difficult to find the correct focal plane among the periodically replicated planes, which occur due to the Talbot effect. Applications based on the Talbot effect can be found in interferometry (Lohmann 1971), optical interconnects (array illumination) (Lohmann 1988), as well as time domain filtering (Jahns and Lohmann 2004, Jahns *et al.* 2006).

In the case of Fraunhofer approximation (eqn [4]) the distribution in the diffraction plane results from a Fourier transform of the distribution in the diffracting screen.

While the Huygens' principle states that the interference of numerous spherical elementary waves forms the resulting diffracted wave, in the Fresnel approximation, these elementary waves are approximated as parabolic waves. Therefore, the Fresnel approximation is often called the parabolic approximation. If the distance to the observation screen is increased further, the former spherical waves can be approximated as sections of plane waves. Thus, the Fraunhofer approximation can be called plane wave approximation. Effectively, the wave field in the diffracting screen is disassembled into plane wave segments of a certain orientation, i.e., angle with respect to the optical axis. The corresponding orientation results in a specific location in the far-field observation plane. **Figure 2** illustrates this transition from the spherical elementary wave over the parabolic to the plane wave approximation.

3.02.1.1.2 The Fresnel number

Figure 2 illustrates that as the distance from the center of the elementary spherical wave increases, the curvature of the wavefront constantly decreases. Eventually, in the region of the Fraunhofer approximation, the observed section of the spherical wave can be considered a plane wave. This can also be illustrated when calculating the interference between the elementary wavefronts with a plane wave at different distances from the center. **Figure 2** shows that the number of interference fringes (i.e., the so-called Fresnel rings) in the observation aperture, which is a measure of the curvature of the wave, decreases with increasing distance. Thus, this number of fringes, which is called the Fresnel number N_F, can be used to define the diffraction regime according to the definition above. The distance between the fringes (black

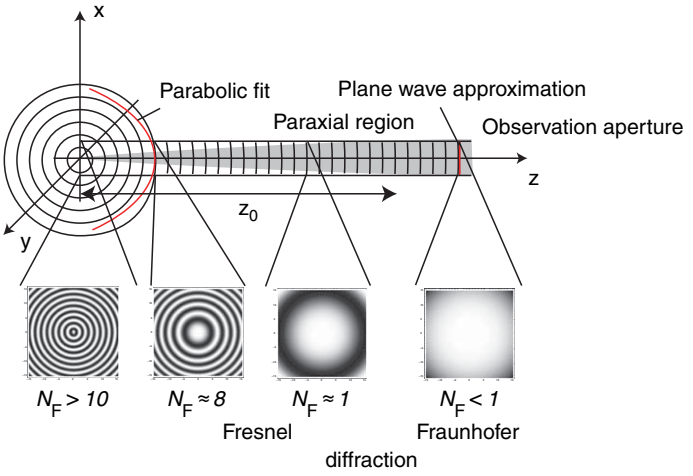


Figure 2 Illustration of the transition between the different diffraction regimes and the significance of the Fresnel number N_F.

and white) in the interference pattern is $\lambda/2$. Thus, in the general case, the Fresnel number N_F can be calculated from the optical path difference (OPD) between a spherical wave and a plane wave

$$N_F \frac{\lambda}{2} = z_0 \sqrt{\frac{x_{max}^2}{z_0^2} + \frac{y_{max}^2}{z_0^2} + 1} - z_0$$
$$\underbrace{\approx z_0 \left(1 + \frac{1}{2} \frac{x_{max}^2 + y_{max}^2}{z_0^2}\right)}_{\text{paraxial approximation}} - z_0$$
$$= \frac{1}{2} \frac{x_{max}^2 + y_{max}^2}{z_0} \Rightarrow N_F = \frac{x_{max}^2 + y_{max}^2}{\lambda z_0} \quad [6]$$

The Fresnel number can also be interpreted as the ratio of the focusing aperture angle (also called the numerical aperture (NA))

$$\sin u \approx \underbrace{u = \frac{x_{max}}{z_0}}_{\text{paraxial appr.}}$$

and the diffraction angle $\Theta_d = \lambda / x_{max}$:

$$N_F = \frac{\sin u}{\Theta_d} \quad [7]$$

or the ratio of the aperture diameter $2x_{max}$ and the diameter of the diffraction limited spot $2\lambda/\sin u$:

$$N_F = \frac{2x_{max}}{\frac{2\lambda}{\sin u}} \approx \frac{2x_{max}}{\frac{2\lambda z_0}{x_{max}}} = \frac{x_{max}^2}{\lambda z_0} \quad [8]$$

In these cases, we limited ourselves to the 1D situation with $y = 0$ without limitation of the generality.

In summary, the Fresnel number is an important measure indicating the ratio of geometrical and diffraction effects on the beam propagation. The Fresnel number is used in the definition of the diffraction regime according to Gross (2005, p. 532) as follows:

$N_F \leq 1 \Rightarrow$ Fraunhofer diffraction regime
$N_F = 1 - 10 \Rightarrow$ Fresnel diffraction regime
$N_F >> 1 \Rightarrow$ Geometrical optics region, diffraction negligible

3.02.1.2 Thin Diffractive Optical Elements

The diffraction integrals in eqns [3] and [4] allow the calculation of the wave propagation through homogenous media in the approximations discussed above. For a comprehensive analysis of the optical functionality of optical systems, we additionally need to discuss the description of optical components. The conceptually simplest consideration of optical elements is known as thin element approximation (TEA), sometimes also called Kirchhoff approximation. The TEA is applicable in scalar diffraction theory. In this case, optical elements are assumed to interact effectively with the incident wave field in only one single plane along the optical axis. Mathematically, the effect of the optical element is then calculated by a simple multiplication of the complex amplitude of the incident wave field with a complex amplitude distribution representing the optical component (Goodman 1996).

$$U(x, y, z + 0) = U(x, y, z - 0) T_{element}(x, y, z) \quad [9]$$

where $U(x, y, z + 0)$ represents the complex amplitude distribution of the wavefront immediately behind the optical element, $U(x, y, z - 0)$ denotes

the complex amplitude distribution in front of the optical element, and $T_{\text{element}}(x, y, z)$ represents the complex amplitude transmission of the optical element. The transmission function $T_{\text{element}}(x, y, z) = A(x, y)\exp(\mathrm{i}\varphi(x, y))$ thus influences the amplitude as well as the phase of the incident wave. For amplitude-only objects (e.g., amplitude gratings) consisting of a more or less absorbing layer, $T_{\text{element}}(x, y, z)$ is real valued. Phase-only functions, on the other hand, are characterized by $A(x, y) = 1$ and a function $\varphi(x, y)$, which defines the shape of the phase profile. Physically, the TEA is a crude idealization because for the modulation of both the amplitude and the phase, a finite physical thickness is required. A specific phase variation, for example, results from a wave traveling through a dielectric material with a specific shape, i.e., thickness variation. A prism, for example, is a wedge of transparent material (refractive index n), while for the lens a plano-convex spherical shape with a radius of curvature R needs to be considered (**Figure 3**). Thus, for the OPD between rays passing through the elements at different positions x, we find for the prism (surrounding medium $n_0 = 1$) (Lohmann 2006):

$$\begin{aligned}\mathrm{OPD}_{\text{prism}}(x) &= (n-1)D(x) = (n-1)x\tan\alpha \\ &\approx (n-1)x\alpha \Rightarrow \varphi_{\text{prism}}(x) \\ &= \frac{2\pi}{\lambda}(n-1)x\tan\alpha \end{aligned} \quad [10]$$

This consideration results in a TEA representation of a prism:

$$T_{\text{prism}} = \exp(2\pi\mathrm{i}(n-1)x\tan\alpha) \approx \exp(2\pi\mathrm{i}(n-1)x\alpha) \quad [11]$$

Similarly for a TEA representation of a spherical lens:

$$T_{\text{lens}} = \exp\left(-\pi\mathrm{i}\frac{(n-1)x^2}{\lambda R}\right) = \exp\left(\pi\mathrm{i}\frac{x^2}{\lambda f}\right), \quad \text{where } f = \frac{R}{(n-1)} \quad [12]$$

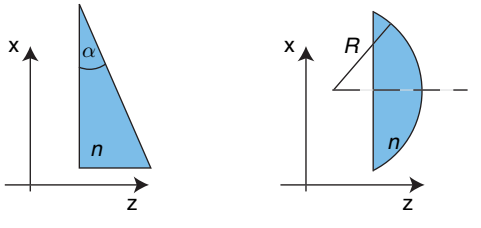


Figure 3 The shape of a prism and a plano-convex refractive lens.

3.02.1.3 Diffraction Gratings

The complex amplitude distribution of periodic diffractive elements (i.e., diffraction gratings) can be expanded in a Fourier series:

$$g(x) = g(x+p) \Rightarrow g(x) = \sum_{m=-\infty}^{+\infty} A_m \exp(2\pi\mathrm{i}m(x/p)), \quad \text{with } A_m = \frac{1}{p}\int_0^p g(x)\exp(-2\pi\mathrm{i}m(x/p))\mathrm{d}x \quad [13]$$

Due to their potentially high efficiency, phase gratings of varying complexity need to be considered. The most general form of a multilevel phase profile is shown in **Figure 4**.

If N is the number of phase levels per 2π phase depth, the transmission function of such a multilevel phase grating can be written mathematically as ($p = 1$) (Jahns 1994):

$$g(x) = \sum_{k=-\infty}^{+\infty} \exp\left(-2\pi\mathrm{i}\frac{k}{N}\right)\mathrm{rect}\left[\frac{x - \frac{k}{N} - \frac{1}{2N}}{\frac{1}{N}}\right] \quad [14]$$

The complex amplitude in the diffraction plane can be calculated by means of a Fourier transform (Fraunhofer diffraction) and is thus represented by the Fourier coefficients A_m as:

$$\begin{aligned} A_m &= \int_0^1 \sum_{k=0}^{N-1} \exp\left(-2\pi\mathrm{i}\frac{k}{N}\right)\mathrm{rect}\left[\frac{x - \frac{k}{N} - \frac{1}{2N}}{\frac{1}{N}}\right] \\ &\quad \times \exp(-2\pi\mathrm{i}mx)\mathrm{d}x \\ &= \exp\left(\mathrm{i}\pi\frac{m}{N}\right)\mathrm{sinc}\left(\frac{m}{N}\right)\frac{1}{N}\sum_{k=0}^{N-1}\exp\left(-2\pi\mathrm{i}\frac{k(m+1)}{N}\right) \\ &= \begin{cases} \exp\left(\mathrm{i}\pi\frac{m}{N}\right)\mathrm{sinc}\left(\frac{m}{N}\right) & \text{if } m = jN-1,\ j \text{ is an integer} \\ 0 & \text{else} \end{cases} \end{aligned} \quad [15]$$

The A_m represent the complex amplitudes of the diffraction orders of the grating where

$$\mathrm{rect}[x] = \begin{cases} 1, & \text{if } x < 0.5 \\ 0, & \text{else} \end{cases}$$

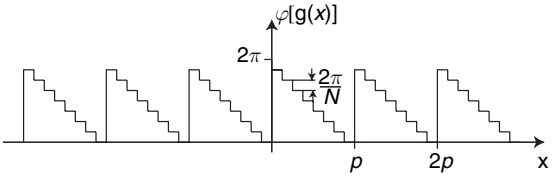


Figure 4 The shape (phase profile) of a multilevel diffraction grating. Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

is the rectangular function and

$$\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

The diffraction angles under which these diffraction orders of a thin diffraction grating appear can be calculated from the important relation:

$$\sin\alpha_{\mathrm{diff}} = m\frac{\lambda}{p} + \sin\alpha_{\mathrm{inc}}, \quad m = 0, 1, 2, 3, \cdots, \quad [16]$$

here, α_{diff} and α_{inc} are the diffraction angle and the angle of incidence of the light beam on the grating, respectively.

Thus, in the scalar diffraction theory (eqn [15]), we find that for a linear phase grating with only N phase steps, every Nth diffraction order has a nonzero diffraction amplitude. If only one diffraction order (here A^{-1}) is considered for the optical functionality, it is obvious that the efficiency of this diffraction amplitude can be significantly larger if the number of phase steps is increased. In other words, the number of undesired stray orders can be reduced significantly by increasing the number of phase steps N per 2π. This is illustrated in **Figure 5**.

More generally diffraction gratings consist of a possibly complex structure ($f(x)$), which is replicated periodically (period length p). Thus, the transmission function of a diffraction grating with Q periods can be written as:

$$g(x) = \sum_{n=-\frac{Q}{2}}^{\frac{Q}{2}} f(x+np) = f(x) * \sum_{n=-\frac{Q}{2}}^{\frac{Q}{2}} \delta(x+np),$$
$$\text{where } f(x) \neq 0 \text{ only in } \left[-\frac{p}{2}, \frac{p}{2}\right] \quad [17]$$

Here, $*$ denotes the convolution operation and $\delta(x)$ is the so-called δ distribution:

$$\delta(x) = \begin{cases} +\infty, & \text{if } x = 0 \\ 0, & \text{else} \end{cases} \quad \text{and} \quad \int \delta(x)\mathrm{d}x = 1 \quad [18]$$

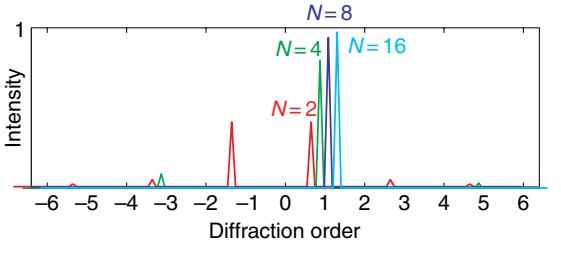


Figure 5 Diffraction amplitudes of phase gratings with varying numbers of phase levels (N).

The Fraunhofer diffraction distribution $\tilde{G}(\nu)$ of such a grating is calculated by a Fourier transform of $g(x)$

$$\tilde{G}(\nu) = \int g(x)\exp(-2\pi i\nu x)\mathrm{d}x = \tilde{F}(\nu)\tilde{A}(\nu),$$
$$\text{where } \tilde{A}(\nu) = \int \sum_{n=-\frac{Q}{2}}^{\frac{Q}{2}} \delta(x+np)\exp(-2\pi i\nu x)\mathrm{d}x$$
$$= \sum_{n=-\frac{Q}{2}}^{\frac{Q}{2}} \exp(-2\pi i\nu(x+np)) \quad [19]$$

In this case $\tilde{F}(\nu)$ is called the form factor since it exclusively depends on the shape $f(x)$ of the individual periods. $\tilde{F}(\nu)$ is calculated as the Fourier transform of $f(x)$. $\tilde{A}(\nu)$ is called the array factor, which has its origin in the periodicity of the diffraction grating. For a large number of periods Q, $\tilde{A}(\nu)$ again becomes an array of δ peaks with a spacing $1/p$. The smaller the number of effective periods, the wider is the extension of the individual diffraction orders, which is a typical phenomenon of multiple beam interference. Thus, the shape of the individual periods forms an envelope for a discrete array of diffraction orders generated due to the periodic structure of the grating (**Figure 6**). This is to say that in the diffraction plane the form factor $\tilde{F}(\nu)$ is sampled by the discrete array of δ peaks with distances proportional to $1/p$. If only a finite number of periods of the diffraction grating are illuminated, the sampling peaks can be described by a series of peaks with finite width, e.g., approximated by the sinc function $\sin(\pi x)/\pi x$ (Lohmann 2006).

3.02.1.4 Zero-order Gratings and Form Birefringence

According to eqn [16] a reduction in the size of the grating period leads to larger diffraction angles. Thus, the number of propagating diffraction orders (with $\sin(\alpha) < 90°$) and stray orders can be reduced with decreasing grating period (**Figure 7**). This leads to an increase in the diffraction efficiency. If the grating period is on the order of wavelength λ or smaller, even the diffraction angles of the first diffraction orders ($m = \pm 1$) reach 90°. As a result, even these first diffraction orders are converted into so-called evanescent waves. These evanescent waves are not propagating like conventional wavefronts but they are merely electromagnetic field distributions existing along the grating surface whose strengths decrease rapidly as the

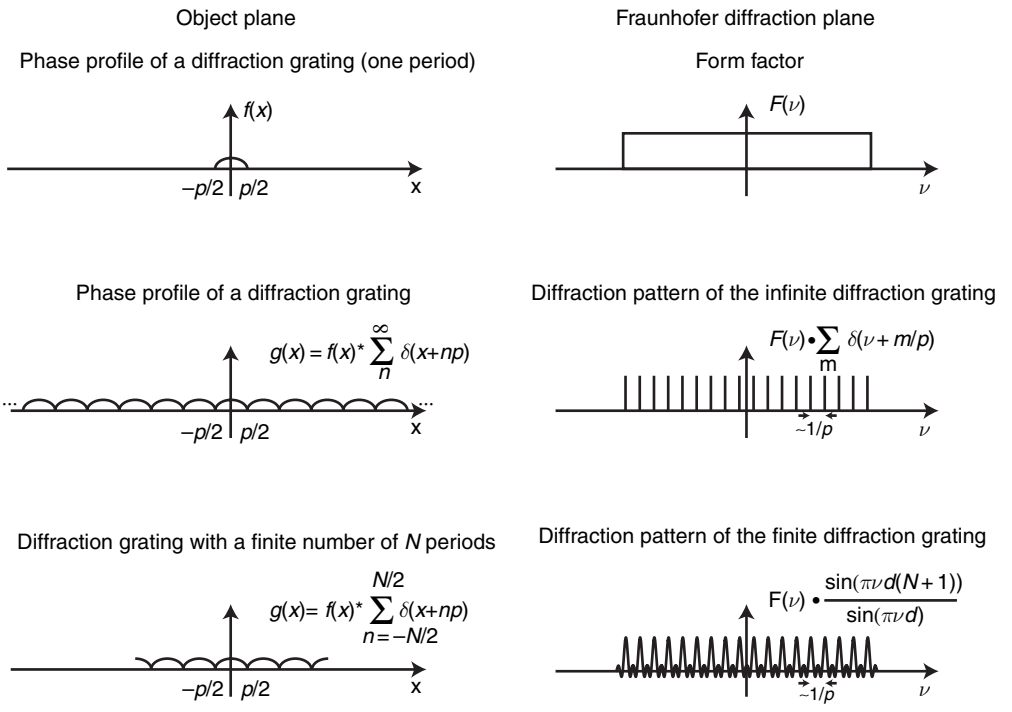


Figure 6 Illustration of the form factor and the array factor of a diffraction grating (here a microlens array).

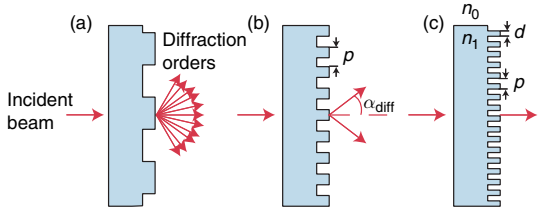


Figure 7 Diffraction orders of gratings with different period sizes. (a) $p >> \lambda$ (scalar theory); (b) $p \sim \lambda$ (rigorous diffraction theory); and (c) zero-order grating: $p < \lambda$ (effective medium theory).

distance from the grating surface increases (Lohmann 2006, pp. 163–8). Without elaborating on the extremely interesting physical background of these evanescent fields, we would like to point out their significance for near-field and waveguide optics (Saleh and Teich 1991). If the far field of such a high-frequency grating is considered, only the zeroth order can be observed as a propagating order. Even though the zeroth order is not deflected through the diffraction at the grating, the grating has an important influence on the propagation. Depending on the duty cycle w/p of the grating, i.e., the ratio between the grating period p and the width of the grating bar w, the propagating zeroth order experiences a refractive index whose value is between the value of the grating material and the surrounding medium. In the approximation of the effective medium theory, the propagating zeroth order experiences a dielectric medium with dielectric constants $\varepsilon_\perp$ and $\varepsilon_{||}$, or refractive indices $n = \sqrt{\varepsilon}$. Obviously, the optical properties depend on the orientation of the polarization with respect to the grating orientation (Born and Wolf 1980):

$$\varepsilon_{||} = \varepsilon_1 \frac{w}{p} + \varepsilon_2 \left(1 - \frac{w}{p}\right); \quad \varepsilon_\perp = \frac{\varepsilon_1 \varepsilon_2}{\varepsilon_2 \frac{w}{p} + \varepsilon_1 \frac{p-w}{p}} \quad [20]$$

This rather crude approximation shows the possibility to influence the optical properties of materials through subwavelength micro- or nanostructuring. It is possible to generate materials with specific refractive indices as well as birefringence (the so-called form birefringence) (**Figure 8**). This has become well known as the possibility to tailor optical properties through micro- and nanostructuring and leads to the field of photonic crystals (Joannopoulos and Meade 1995, Kraus and De La Rue 1999).

3.02.1.5 Thick Diffraction Gratings – Holographic Optical Elements

An alternative way of improving the diffraction efficiency of diffraction gratings is through the implementation of thick gratings, for example,

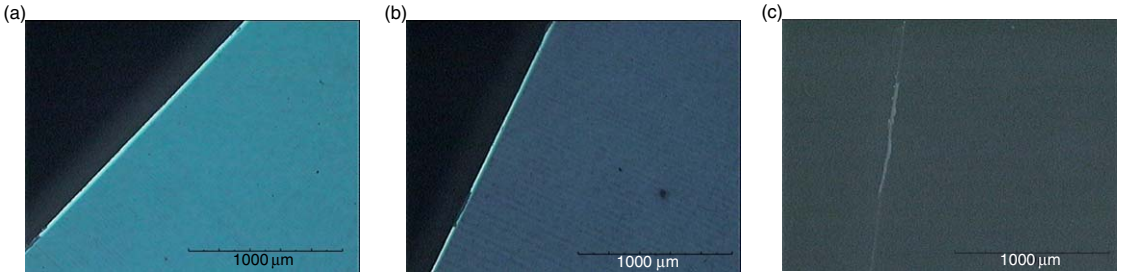


Figure 8 Anisotropy induced in isotropic materials through microstructuring: The pictures, taken from a polariscope, show the anisotropy of a diffraction grating (period 840 nm) replicated in poly(methyl methacrylate) (PMMA). The grating is introduced between two polarizers oriented perpendicular to each other. Then the grating is rotated around the optical axis. The pictures show the output when the grating structures are aligned at an angle of (a) 45°, (b) 60°, and (c) 80° relative to the axes of the polarizers. Due to the anisotropic properties of the grating (form birefringence) the plane of polarization of the incident wavefront is rotated depending on the orientation of the grating. Thus the portion of the light passing the analyzer depends on the grating orientation.

through holographic recording in optically thick photographic material. While in thin optical elements, the light is considered to interact with the elements just one single time, in thick optical elements multiple interactions with the element have an important influence on the optical functionality. As shown in **Figure 9**, due to the multiple interactions with the thick periodic structure, the Bragg effect results in an efficient additional interference filter, which is combined with the thin element. This Bragg filter is designed such that it selects precisely one of the diffraction orders, which would be generated by a thin grating of the same period. The Bragg grating can be considered as a stack of alternating layers with high and low refractive indices. The thickness of each of these layers is $p/2$, if p is the period of the holographic interference grating. By calculating the path difference Δl between two rays reflected at the high and low refractive index layer, respectively, the condition for a ray to be reflected at such a stack layer can be derived, i.e., to interfere constructively at the respective angle:

$$\Delta l_{\text{opt}} = 2\frac{p}{2}\sin\alpha_{\text{stack}} = \frac{\lambda}{2} \Rightarrow p = \frac{\lambda}{2\sin\alpha} \qquad [21]$$

Here, an additional phase shift of π, taking place when the beam is reflected at the higher refractive index layer, was taken into account (the refractive indices of the different layers were neglected). It is interesting to point out that an identical relation results from the calculation of the period of the sinusoidal interference fringes resulting for the interference of two wavefronts propagating at a relative angle of α. Thus, we understand that once these interference fringes are recorded in a holographic material and are illuminated by either of these wavefronts, the second wavefront is automatically reconstructed as a diffraction pattern (Solymar and Cooke 1981). This is the fundamental concept of diffraction. In the context of diffraction gratings, it is important to note that as multiple beam interference occurs at these optically thick Bragg gratings, very efficient diffraction gratings can be fabricated. For a detailed analysis of thick holographic gratings, rigorous solutions to the diffraction problem in periodic media need to be found (Petit 1980). One of the successful approaches to this type of problem is the so-called coupled wave theory (Knop 1978).

The transition from the case of Bragg reflection at thick gratings to the approximation of thin diffraction gratings that generate a significant number of diffraction orders can also be described with rigorous diffraction theory. This is possible for the assumptions that the light beam is incident parallel to the

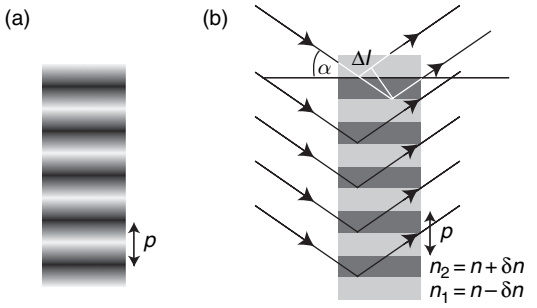


Figure 9 (a) Thick volume gratings generated by a holographic interference grating; (b) Bragg reflection at the multilayer stack modeling the holographic grating. Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

Bragg reflecting layers (i.e., $\alpha = 0$ in **Figure 9** and thus the Bragg condition is clearly violated) and that the variation of the electric field in y-direction can be neglected. In this case, the well-known grating equation for directions of the diffraction orders can be derived:

$$\sin \alpha_{\text{diff}} = m\frac{\lambda}{p}, \; m \text{ is an integer} \qquad [22]$$

This regime is called Raman–Nath diffraction named after the Indian scientists who suggested the theoretical approach for the diffraction of light at acousto-optical waves (Solymar and Cooke 1981).

This Bragg filtering effect being based on the principle of multiple beam interference becomes more and more selective when more layers of the multilayer stack are effectively involved in the propagation of light through the element. Thus, the effect of thickness also gradually becomes more important with larger diffraction angles (since in this case the number of interfering beams increases due to the effectively increasing thickness of the stack) and the larger the physical thickness of the element. There is no strict limit as to when scalar diffraction theory is valid and when the gratings can be approximated as thick Bragg gratings. As a rule of thumb the ratio between the physical thickness t and the grating period p can be used for approximation of the validity of the approximations (Caulfield 1979):

$$\underbrace{\left(\frac{t}{p}\right)_{\text{thick}} > 10}_{\text{Bragg approximation}} \cdots \underbrace{1 > \left(\frac{t}{p}\right)_{\text{thin}}}_{\text{scalar approximation}} \qquad [23]$$

In the region in between those two extreme cases of thick and thin diffraction gratings, none of the two approximations is valid to a sufficient precision. In this case, rigorous methods for the calculation of the diffraction amplitudes have to be applied. Various methods for a rigorous solution to the diffraction problem have been suggested, which we cannot elaborate on in this context. It is important to note that scalar theory does correctly predict the propagation directions of the diffracted beams. However, for a precise analysis of the diffraction amplitudes, a rigorous calculation becomes necessary in many cases.

3.02.1.6 Specific Diffraction Gratings and Applications

The biggest potential applications of diffractive optical elements (DOEs) lies in the possibility to fabricate very compact and lightweight components and systems with complex functionality. This becomes possible because of the periodicity of the optical wave field, which allows one to reduce the thickness of the diffractive components to an optical thickness of one wavelength. With a subsequent quantization step, the structure of a binary or multilevel phase grating can be derived directly from this consideration (**Figure 10**). Despite this very clear and straightforward transition from the conventional refractive component to the corresponding diffractive element, there are very significant consequences on functional principle as well as on the fabrication process.

The physical parameter that determines the optical functionality of refractive components is the shape of the refractive surface profile. This shape eventually determines the angle of incidence of the beam onto the refracting boundaries and thus, together with the refractive indices, the refraction angles. This leads to extremely challenging fabrication techniques applying 3D profiling and polishing with optical surface quality. In DOEs, on the other hand, the fabrication technology is reduced to the fabrication of binary absorbing elements (for amplitude gratings) or etching steps (for phase gratings), which can be fabricated with extremely high precision through lithographic fabrication technologies. The optical functionality of the DOE is exclusively achieved through the positioning of these small binary diffracting phase shifting segments. Thus, in diffractive optics, no complex surface profiles need to be fabricated. This has a further consequence on the complexity of the optical functionality, which can be realized using diffractive elements. All DOEs consist of the same inherent structure, i.e., binary

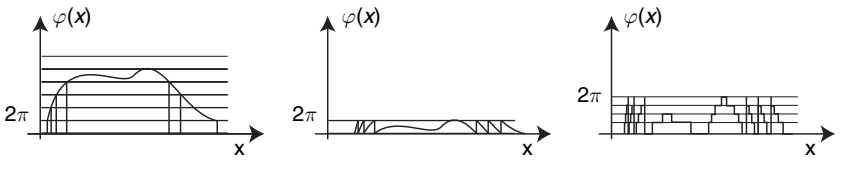


Figure 10 Illustration of the quantization of a refractive phase profile: reduction of the element thickness due to the periodic wave field; quantization of the analog sections into a step profile.

microstructures, which is independent of the optical functionality. Therefore, various functionalities such as beam splitting and focusing can be combined within a single diffractive element without negative effects on the fabrication processes. However, the quantization needs to be adjusted properly, which might eventually lead to very small feature sizes. In summary, the challenges of refractive optics fabrication, i.e., the precise 3D profiling, in diffractive optics are shifted to the precise positioning of micro- or nanoscaled diffracting features.

Many of the most important applications of DOEs can be listed under the headline multiple beam splitting or beam shaping. Recently, the term wavefront engineering has been coined for this wide field of applications. Generally speaking, the purpose of any optical element, no matter if prism, lens, or DOE, is beam shaping, i.e., to redistribute the incident wave field to a desired complex amplitude distribution in a specific plane behind the element (**Figure 11**). This is possible by influencing the complex amplitude of the incident wave field, e.g., by manipulating the amplitude and the phase of the wavefront. Fundamentally, with the ability to calculate the propagation of a specific wave field, for example, in the Fraunhofer or Fresnel approximation, it is rather simple to inversely propagate the desired complex amplitude distribution into the element plane. Then, with suitable modeling of the beam shaping element, for example, in the TEA, it is straightforward to calculate the necessary structure of the DOE from the knowledge of the incident wave field. Due to the challenging fabrication process of classical refractive elements, various beam shaping elements and thus also various beam shaping tasks in refractive optics are limited, for example, to focusing or deflection. More complex beam shaping tasks are possible with various individual elements, e.g., anamorphotic imaging with different magnifications in x- and y-directions. With the invention of DOEs, originally coined as computer generated holograms (CGHs), and the development of the technological basis for their fabrication, the variety of achievable tasks has increased dramatically. Nowadays, almost any physically realistic light distribution can be generated using DOEs. Fundamentally, it is possible to generate any complex amplitude distribution in a specific plane by influencing the amplitude as well as the phase within one plane. The only optical elements that allow such a complex manipulation of an incident wavefront are holograms. However, from the technological perspective, there are limitations to this concept (Sinzinger and Jahns 2003).

3.02.1.7 Design of Diffractive Optical Elements

Obviously, a phase only modulation, for example, using diffractive elements fabricated through lithographic surface patterning, has the advantage of high transmission since no absorption takes place in the element. During the design process of these diffractive phase elements, the task is thus to optimize the phase distribution in order to generate a desired amplitude distribution in the output plane (Soifer *et al* 1997). One of the most widely used approaches for this task is the iterative Fourier transform algorithm illustrated in **Figure 12**. In this algorithm, the phase profile of the diffractive element is adjusted iteratively to improve the approximation of the diffraction pattern generated in the Fraunhofer diffraction plane. It is thus applicable for the optimization of the far-field diffraction pattern of DOEs. The fundamental algorithm has been adopted from the reconstruction of image phase information from knowledge of the intensity distribution of the diffraction pattern in electron microscopy (Fienup 1982, Gerchberg and Saxton 1972). The algorithm starts with either a predesigned or a random phase distribution. The far-field diffraction pattern for this distribution is calculated by means of a Fourier transform and is compared to the desired output. Very often, the goal of the design process is a specific

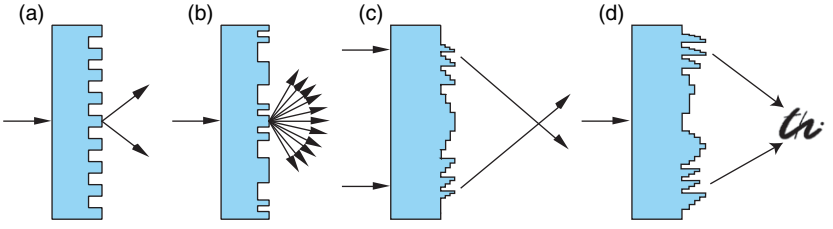


Figure 11 Functionality of diffractive optical elements. (a) 1 × 2 beam splitter; (b) multiple beam splitter; (c) diffractive lens; and (d) arbitrary beam splitting (beam shaping).

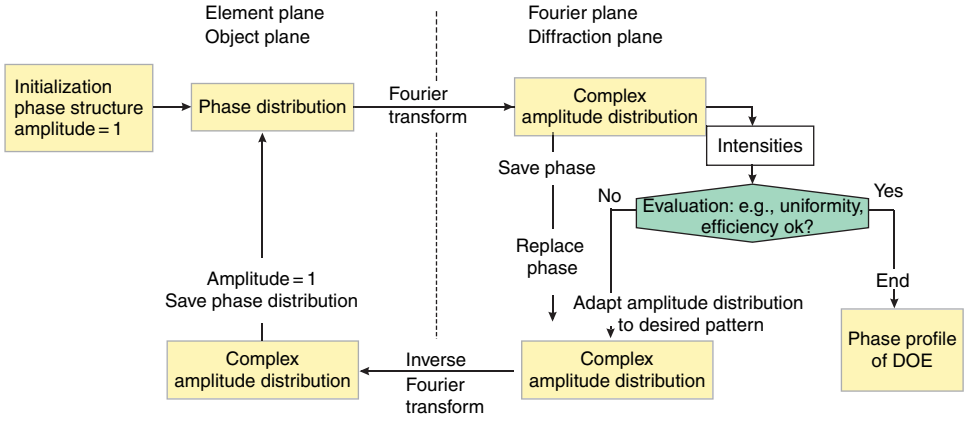


Figure 12 Schematic of the iterative algorithm for the design of phase-only diffractive optical elements (DOEs).

amplitude distribution, while the phases of the diffraction amplitudes are free design parameters. Thus, the next step in the algorithm is to force the amplitudes to the desired values while keeping the phase information as calculated. After an inverse Fourier transformation, we are back in the plane of the DOE. Here, for efficiency reasons, we require the element to be a pure phase element. We thus adjust the amplitudes of all element pixels to be unity. After another Fourier transform we find a new diffraction pattern. This iteration continues until a phase distribution is found, which generates a diffraction pattern that approximates the desired output amplitude distribution sufficiently well. The advantage of this iterative design procedure is its inherent simplicity and the flexibility with respect to the desired output distributions. The algorithm is often applied to the design of beam splitter gratings, in which case it is not restricted to symmetrical spot arrays or equal intensity diffraction orders (Bryngdahl and Wyrowski 1990). It can thus be applied to the design of DOEs for the generation of raster images (Gruber 2001). It is also possible to design diffraction orders with predetermined phases if a sufficient number of free design parameters are available. The algorithm can also be adapted for diffraction regimes other than Fraunhofer diffraction, e.g., for the design of DOEs generating specific distributions in the Fresnel regime (Dorsch *et al.* 1994). In this case, however, the limitations as to the design freedom have to be considered thoroughly.

In conventional diffractive elements, it is only possible to modulate either the phase or the amplitude. Truly complex filtering is thus possible only with two optical elements separated by a specific distance (Bartelt 1984; Sinzinger and Jahns 2003, pp. 273–8). Most flexibility is achieved if the two elements are positioned in the optical systems such that their relationship is defined through far-field diffraction. However, flexible beam shaping with two DOEs has also been demonstrated in the Fresnel relationship (**Figure 13**).

Diffractive optical elements designed in the way outlined above have been applied for lateral, axial, and temporal or spectral shaping of wave fields (Sinzinger and Jahns 2003, Chap. 11). From a very general perspective almost any application of DOEs can be categorized in this respect. Nevertheless, we want to mention briefly some more specific applications.

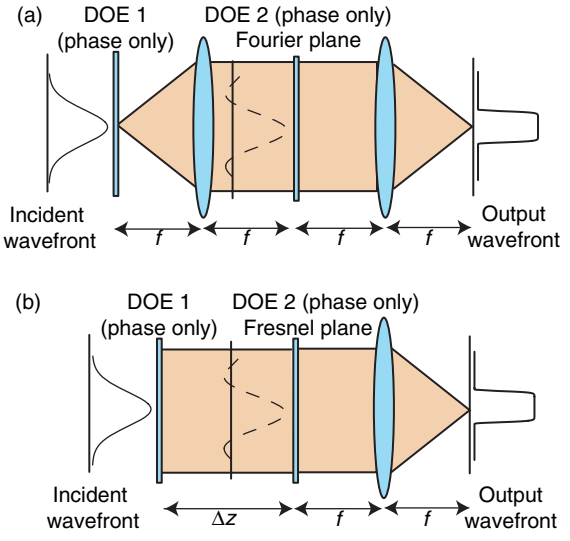


Figure 13 The concept of complex beam shaping by using two phase-only diffractive optical elements (DOEs) in corresponding planes in the optical system. (a) Fourier plane relation and (b) Fresnel diffraction relation between the DOEs. Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

Computer-generated holograms, for example, have been invented for so-called spatial filtering, which is useful, for example, in feature extraction, character recognition, or optical correlation (Vander Lugt 1992). The basics of optical spatial filtering are best illustrated in the example of the 4f configuration schematically shown in **Figure 14**. In this application, a specifically designed DOE is located in the Fraunhofer plane of an incident wave field, which, for example, is formed by a uniformly illuminated object. Through complex or phase only filtering in the Fraunhofer plane it is possible to select specific features of the object, which are then imaged in the observation plane after an additional Fraunhofer diffraction step. Since a multiplication of two functions in the Fourier plane corresponds to the convolution of the two functions, the fundamentals of the approach are understood immediately. The basic concepts of spatial filtering go back to the 1960s (Goodman 1996, Lohmann 2006, Weaver and Goodman 1966). Traditionally this approach has been applied to analog signal processing as well as for the implementation of associative memories and neural network systems (Cassasent 1978, Horner 1987). In combination with modern design and fabrication technologies they have been adjusted and are applied, for example, in optical cryptography, optical document security, etc. (Javidi 2006, Javidi and Horner 1994, Sinzinger 2002).

DOEs are specifically useful components for spectral analysis, which is traditionally the most important area of the application of diffraction gratings. This is due to the strong wavelength selectivity of the diffraction angles illustrated through eqn [16]. Consequently in the Fraunhofer diffraction plane of a linear grating (period p), the wavelength spectrum of the incident wave field is separated laterally. The basic configuration of a grating spectrometer is shown in **Figure 15**. The lateral separation Δx of the mth diffraction orders for different wavelengths λ_1 and λ_2 is calculated from:

$$\Delta x = \frac{mf}{p}(\lambda_1 - \lambda_2) = \frac{mf}{p}\Delta\lambda \qquad [24]$$

Taking into account the size of the lateral resolution element $\delta x \approx (\lambda/\mathrm{NA})$, we find that the number of grating periods N illuminated in such a configuration determines the spectral resolution of the system:

$$\frac{\lambda}{\Delta\lambda} = mN \qquad [25]$$

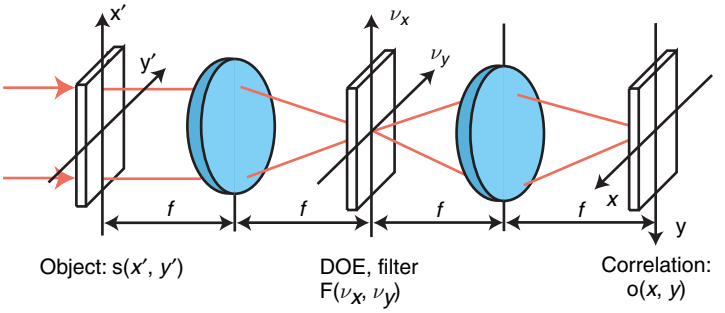


Figure 14 Schematic of the basic optical system for correlation and convolution (4f-configuration). Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

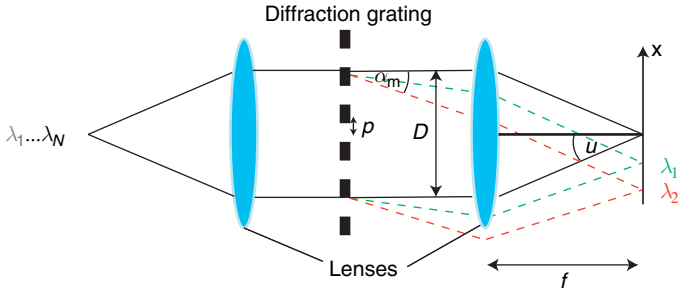


Figure 15 Schematic of a grating spectrometer for spectral analysis of the first stage toward spectral or temporal beam shaping.

High-resolution spectral systems are, for example, used in communication technology for wavelength division (de-)multiplexing, in sensing for sensors in the security as well as in environmental monitoring area. Specifically in sensor systems, the desire to build highly integrated compact systems currently provides the challenges.

For spectral or temporal beam shaping, the configuration shown in **Figure 15** is extended to a 4f-(correlation) configuration (**Figure 14**). In this case the diffractive spatial filter also automatically represents a spectral filter capable of selecting specific wavelengths. Since the temporal profile of laser pulses is directly related to their spectral distribution, such a configuration can be applied to temporal pulse shaping (Froehly *et al.* 1983, Nuss *et al.* 1994, Weiner *et al.* 1992). Similarly, time to space conversion in such setups can also be applied to ultrafast signal transmission (Konishi and Ichioka 1999).

3.02.1.8 Diffractive Lenses

It is possible to design diffraction gratings in such a way that they focus an incoming collimated beam of light into a focus spot, i.e., they work as lenses. To this end, it is necessary to design the diffracting microstructure such that the diffracted elementary waves interfere constructively in the focus plane. This is the case in the focus spot if the OPDs between these elementary waves are equal to multiples of the wavelength λ (**Figure 16**).

These straightforward requirements lead to the fundamental design rules for diffractive lenses:

$$r_j^2 + f^2 = (f + j\lambda)^2 \Rightarrow r_j^2 = 2j\lambda f + (j\lambda)^2$$

$$\xrightarrow{\text{paraxial: } f >> j\lambda} r_j^2 \approx 2j\lambda f \Rightarrow f = \frac{r_1^2}{2\lambda} \qquad [26]$$

Here, r_j is the radius of the diffracting rings, f is the focal length, and λ is the design wavelength (Miyamoto 1961, Nishihara and Suhara 1987). Very often, the paraxial approximation of this relation is used for the design of diffractive lenses (Sinzinger and Jahns 2003). As expressed by eqn [26] the structure of the diffractive lens is periodic in r^2 with period r_1 (Jahns and Walker 1990). Thus, the distance between the rings decreases with increasing radius. This can also be understood from eqn [12] where we discussed the phase profile of a thin paraxial (quadratic) lens. Conceptually it is easy to understand that beams hitting the lens at the edges need to be deflected to a larger diffraction angle, which can be achieved only with smaller grating periods.

As in linear diffraction gratings, the form factor (i.e., the shape of the individual periods of the diffractive grating) determines the distribution of the diffraction amplitudes, i.e., the diffraction efficiency. For a conventional binary amplitude diffractive lens a large portion of the energy is deflected into higher diffraction orders. It is important to note that these higher diffraction orders form additional foci f_m as indicated in **Figure 16**. Consequently, such a diffractive lens

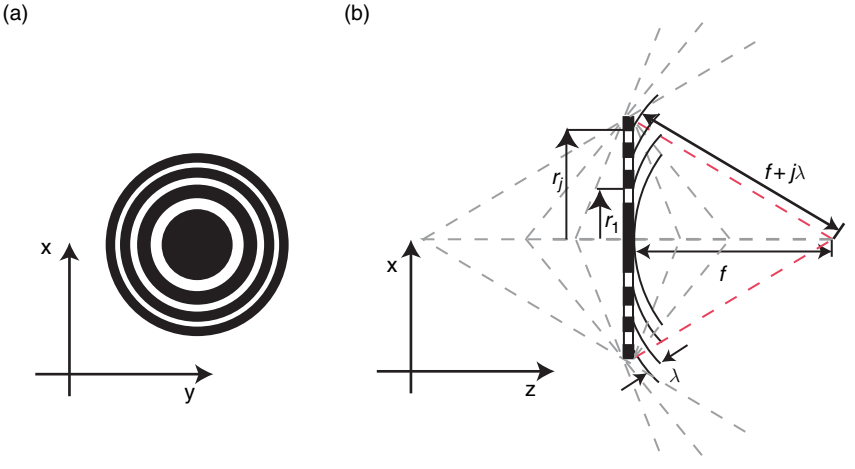


Figure 16 Microstructure of a diffractive lens. (a) View from the top and (b) side view with indication of various diffraction orders. Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

is a multifocus lens that additionally acts both as a positive and as a negative lens:

$$f_m = \frac{r_1^2}{m2\lambda},\ m \text{ is an integer} \qquad [27]$$

Besides the low diffraction efficiency, these higher foci lead to a significant contrast reduction due to stray and background light. In order to increase the diffraction efficiency and contrast of the diffractive lens, the same approach is feasible as for linear gratings. Multilevel phase steps within one period of the circularly symmetric diffractive lens lead to a largely increased efficiency and significantly less background noise. As seen in eqn [15] only eight phase levels per 2π are sufficient to reach a theoretical diffraction efficiency of 95% in the scalar approximation.

The limitations of diffractive lenses result from the fabrication processes. As we have seen, the smallest features in diffractive lenses occur at the edges of the lenses. Thus, for a given minimum feature size $w_{\min}$, the maximum aperture angle (or NA) is limited to (N is the number of phase levels):

$$\mathrm{NA} = \frac{\lambda}{N w_{\min}} \qquad [28]$$

The so-called photon sieve (**Figure 17**) has been suggested to overcome this limitation with the aim of building diffractive lenses for X-rays. Due to their extremely short wavelength, the features of conventional diffractive microlenses for this wavelength regime are extremely small and often beyond what can be fabricated. Instead of the ring structure, Kipp *et al.* (2001) suggested a structure consisting of small holes in a nontransparent screen. Again in order to achieve focusing properties, it is necessary that the waves diffracted at the holes interfere constructively in the focal plane. This can be achieved by arranging the holes of the sieve on the transparent rings of the Fresnel ring structure of conventional diffractive lenses. However, if the holes are larger than the ideal ring, diameter portions of the diffracted light interfere constructively in the focal plane. This results in a decrease in the focus intensity. Nevertheless, even larger holes can have a net contribution to the focus intensity (Cao and Jahns 2002). The potential of the photon sieve thus lies in the possibility to trade focusing power (NA) for lens efficiency.

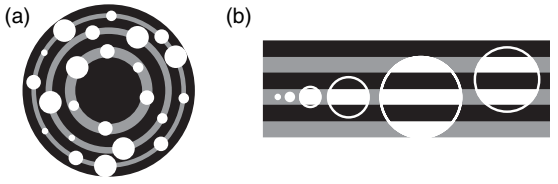


Figure 17 The concept of the photon sieve. (a) Focusing with a distribution of holes in a screen and (b) effect of the size of the holes with respect to the corresponding Fresnel rings. Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

Two features of diffractive lenses can be pointed out here as their most outstanding advantages. First, the diffractive implementation is not limited to the quadratic or spherical shape that we have been discussing above. It is, for example, equally straightforward to implement aspherical lenses in the diffraction concept. Optimized astigmatic diffractive lenses, for example, have been used to optimize folded optical systems fabricated using the concept of planar optics (Jahns and Acklin 1993, Shiono and Ogawa 1992). Second, just like linear diffraction gratings, diffractive lenses have a strong dispersion with an opposite sign rather than the normal dispersion of refractive elements. The focal length of a diffractive lens depends on the illumination wavelength according to the following equation:

$$\frac{\partial f}{\partial \lambda} = -\frac{f}{\lambda} \qquad [29]$$

This chromatic behavior can be exploited, for example, in optical system design for chromatic error correction. To this end, however, diffractive lenses are necessary, which exhibit a very high efficiency over a very broad spectral band. This requirement was thought to be unsolvable for a long time until multilayer diffractive elements had been suggested by Nakai in the late 20th century (Nakai and Ogawa 2002, Nakai and Ogawa). Such multilayer DOEs consist of a stack of two or more blazed diffractive DOEs fabricated from different materials. With a suitable choice of materials it is possible to achromatize the functionality of the individual zones of the stacked DOE. Thus, the multilayer DOE has an efficiency close to 100% over the whole visible spectrum. This development triggered an enormous development leading to the successful application of DOEs in photographic objective lenses as well as in high-quality inspection microscope objective lenses (Dobschal *et al.*). The compactness, the lightweight, the complex optical functionality, and good possibilities for integration are the most important advantages.

3.02.2 Refractive Microlenses and Microlens Arrays

Microlenses and microlens arrays turn out to be the most crucial components for microsystems, which incorporate optical functionality. As in macroscopic optical systems, the most important functionality in optical microsystems is the imaging, focusing, or collimation of light (Daly 2001, Herzig 1997, Sinzinger and Jahns 2003). Various new innovative fabrication technologies allow fabrication of microlenses with good optical quality in large quantities at a relatively low cost per element (Borrelli 1999). With increasing availability, numerous applications of optical microlenses and microsystems have been developed within the past decades. These applications range from sensing, special purpose imaging, optical security devices to optical interconnects.

In this section, we summarize the basics of microlenses as well as microlens arrays and introduce the latest trends in microlens fabrication and applications.

3.02.2.1 Performance of Lenses at Small Dimensions

The properties of microlens arrays compared to macro-optical elements are dominated by two specifics, i.e., the small physical dimensions of the individual microlenses and the array character (periodicity) if a number of microlenses are illuminated by a single expanded beam of light. In this case, the microlens array acts like a diffraction grating where the shape of the individual grating period has the phase profile of the lens.

First, we want to consider the effect of miniaturization on the performance of the individual lens elements. To this end, it only makes sense to compare lenses with equal optical performance but different physical dimensions (diameters) (**Figure 18(a)**). This means that we compare lenses that have, for example, the same (diffraction limited) optical resolution (Lohmann 1989). The optical resolution is defined as the size of the smallest resolvable features δx in an image with

$$\delta x \approx \frac{\lambda}{\mathrm{NA}}; \quad \mathrm{NA} = n \sin u \qquad [30]$$

where λ is the wavelength and NA is called the numerical aperture, with n being the refractive index of the surrounding medium and $2u$ the aperture angle of the lens. Thus, lenses with equal NA have the same diffraction-limited point spread function (PSF). On the other hand, if we consider, for example, the amount of wavefront aberrations to be constant, their effect in the image plane is reduced for lenses with a smaller diameter due to their smaller focal lengths (at constant NA). In summary, this results in the observation that the performance of microlenses is limited by the diffraction limit rather than by the aberrations, which generally limit the performance of macroscopic lenses. Thus, it becomes easier to fabricate lenses with diffraction-limited performance if their diameters are smaller. On the other hand, if we consider the space-bandwidth product (SBP), i.e., the number of image points resolvable with the corresponding lenses, which is an important quality criterion for optical systems and elements, their performance improves significantly with increasing diameters (at constant NA). This is a result of the larger image field size, which increases quadratically with the lens diameter. These circumstances are shown in **Figure 18(b)**, which illustrates the fact that macrolenses perform better if aberrations can be controlled (black curve).

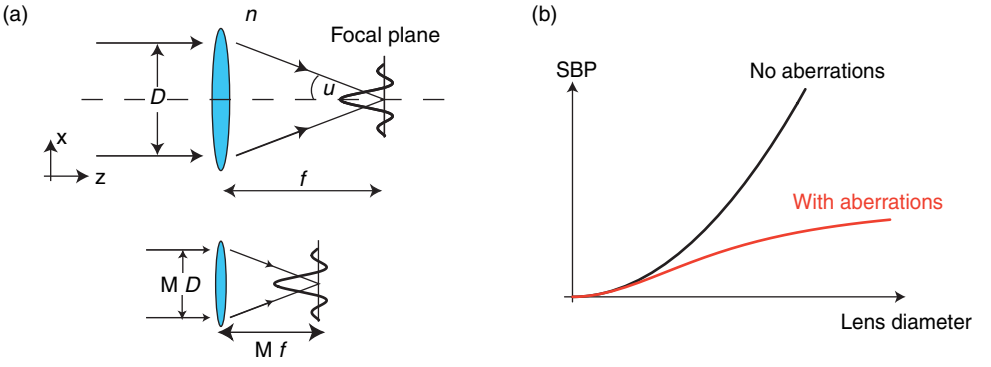


Figure 18 Scaling behavior of lenses. (a) Scaling (factor M) with constant numerical aperture; (b) the curves show the number of resolvable image points (SBP, space bandwidth product) of (micro-)lenses as a function of the lens diameter. Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

The Fresnel number N_F (eqn [6]), which we have already discussed in the context of diffraction theory, is another important parameter, which is useful for judging the quality of (micro-) lenses. As we mentioned above, the Fresnel number is a measure of the relative impact of diffraction on the propagation of a wavefront. As the Fresnel number decreases, diffraction gains more and more importance, which has significant effect on the focusing and imaging behavior of wavefronts, such as focus shifts and space invariance of the imaging behavior (Li 1981, Li and Platzer 1983, Sheppard 1986, Wang *et al.* 1993). The Fresnel number can be interpreted as the ratio of the NA and the diffraction angle (eqn [8]). Consequently, at constant NAs, the Fresnel number decreases linearly with decreasing lens diameters. In other words, diffraction gains more and more importance on the wavefronts propagating behind microlenses, even if the NA and with it the diffraction limit are kept constant. Thus, in order to calculate the best focus of microlenses with small diameters, it is significant to take into account the diffraction at the aperture stop of the lens. It is straightforward to see that diffraction at the aperture stop, which by itself has a certain focusing effect, limits the maximum focal length of lenses with the corresponding diameters (Ruffieux *et al.* 2006). For diffraction gratings, a rule of thumb states that at least 10 periods of a diffraction grating should be illuminated in order to achieve a clear-cut diffractive behavior. Similarly, one can state that the behavior of microlenses with $F_N > 10$ can be estimated to a decent precision (errors $< 10\%$) according to the laws of geometrical optics.

Besides the smaller sensitivity to aberrations, the use of microlenses and microlens arrays is interesting because they enable very compact imaging systems. As a rule of thumb for conventional optical elements, it can be assumed that the lens diameter needs to be at least as large as the size of the image field. For imaging systems with constant NA, this automatically means that large field sizes come along with large distances between the object and the image (**Figure 19**). Especially for large image fields, macroscopic lens systems thus tend to become large, also in axial direction, i.e., in the z-direction. By microchannel imaging where microlens arrays are applied for such an imaging task it is, however, possible to separate image size from the length of the imaging system. Since each individual microlens system images only a small section of the object, the distance between the object and the image can remain rather small. Thus, extremely compact imaging systems for large object sizes are possible using microlens arrays. This is the fundamental concept of microlens lithography, which has gained significant commercial importance during the last decade (Völkel 1996).

3.02.2.2 Microlens Arrays

With the introduction of microlenses into optical systems, very often a reduction in the pupil size of the individual microchannel system is introduced. However, in many applications the overall pupil remains macroscopic in its dimensions while a microlens array splits the aperture into a large number of subapertures (**Figure 19**). Consequently, it is also generally necessary to consider the effect of the array properties on the performance of the complete system. In this section, we first discuss some of the important array parameters of periodic micro-optical components. Then we outline the analytical treatment of those array parameters as well as their effect on the performance of the array. Finally, we address

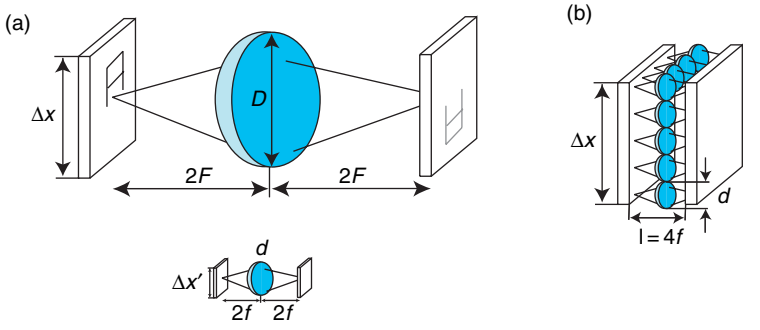


Figure 19 Scaling of the image field size. (a) Conventional imaging system at different scales and (b) microchannel imaging system. Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

the new possibilities that arise when the regularity of these arrays is broken.

3.02.2.2.1 Array parameters

In addition to the performance parameters of the individual elements, it is necessary to consider the so-called array parameters when using arrays of micro-optical components. The most important array parameter is the uniformity of the elements throughout the array. Specifically for the fabrication of large arrays of micro-optical components (e.g., microlens arrays), it is quite a challenging task to achieve uniform properties for all the lenses in the array. On the other hand, it is similarly difficult to characterize the uniformity of the optical performance if the number of individual elements (microlenses) is very large. Shack–Hartmann sensors (Beyerlein *et al.* 2005, Pfund *et al.* 2001), Moirè sensors as well as arrays of microinterferometers (e.g., of the Smartt type) have been discussed for this so-called array testing (Schwider and Sickinger 1997; Sinzinger and Jahns 2003, Chap. 4).

The second important array parameter is the fill factor of the array. It corresponds to the packaging density of the individual components and determines whether there are gaps between the individual microlenses in the array. Obviously, also the shape of the array (array grid, e.g., hexagonal or Cartesian grid) as well as the shape of the individual microlenses influence the packaging density. Microlenses with circular apertures or profile shapes, for example, cannot completely fill the pupil aperture without dead spaces in between the microlenses. Microlenses and micro-optical elements with square apertures can be packed to extremely high fill factors. The diffractive implementation is well suited for the fabrication of such elements with square apertures. On the other hand, microlens arrays with high fill factors have also been achieved for refractive microlens arrays fabricated through reflow and optimized ion exchange in glass.

3.02.2.2.2 Regular microlens arrays

Generally, microlens arrays are considered to be packed as densely as possible on a Cartesian grid. In this case, if a significant number of microlenses are illuminated within the pupil of the optical system, the microlens array can be treated like a diffraction grating where the form of the individual microlens corresponds to the form of one period of the diffraction grating. This situation is illustrated in **Figure 6** and analyzed in Section 3.02.2.3. For microlenses, the form factor corresponds to the Fourier transform of the microlens profile and thus provides a uniformly illuminated area in the Fraunhofer diffraction plane. The extension of this homogeneously illuminated area depends on the NA of the individual microlenses. This area is sampled by an array of sharp peaks, which have the origin in the periodicity of the microlens array. It depends on the specific application if this periodicity has a positive or negative effect on the desired output pattern (**Figure 20**).

Microlens arrays, for example, have been suggested for multiple beam splitting or array generation (Streibl *et al.* 1991). In this application, the goal is to generate an array of uniformly illuminated diffraction orders in the Fraunhofer diffraction plane. To this end, microlens arrays can be very useful. However, instead of observing the distribution in the focus plane of the microlenses, the distribution in the Fraunhofer diffraction plane is considered. In combination with an array of an additional phase correcting field lenses, this approach enables the efficient generation of a large number of orders with very good uniformity in the intensity distribution. To this end, obviously the periodicity of the microlens array is useful and necessary for the optical performance.

Microlens arrays, on the other hand, are also suggested for beam homogenizing in the so-called fly's eye condenser configuration. Considering the effect of one single microlens in the Fraunhofer diffraction plane of the condenser lens, we again find a uniformly illuminated area. Since the same area is illuminated by every single microlens, conceptually, a very uniform illumination results in this area. However, problems arise for coherent sources. In this case, the interference of the individual beams destroys the uniformity in the illuminated area. Again, the sharp peaks of the array factor show up as expected theoretically. Thus, for

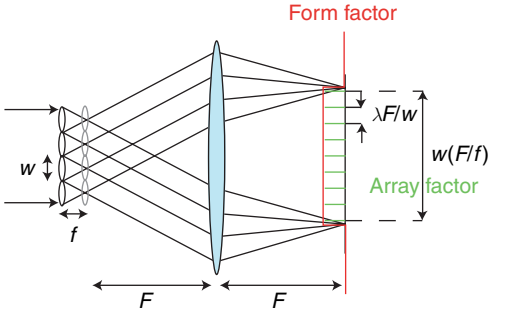


Figure 20 Form factor and array factor of a microlens array used, for example, as fly's eye condenser for beam homogenization. Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

beam homogenizing, the regularity of the microlens arrays results in serious problems.

An additional effect to be observed in periodic microlens arrays is the Talbot effect. This results in the observation of a large number of focal planes at periodic distances. It is thus quite challenging, for example, to experimentally find the first and the best focal plane of microlens arrays.

3.02.2.2.3 Nonuniform microlens arrays

For microchannel imaging of large arrays of data channels it has been pointed out that arrays of microlenses with varying parameters offer additional degrees of freedom for the optimization (Sinzinger and Jahns 1997). For example, the focal length of each microlens can vary depending on the position of the lens within the array. Similarly, additional deflection angles can be introduced and can vary for each individual element of the array. These additional design parameters allow one to adjust the properties of the microlens array to the very specific imaging task (Wippermann *et al.* 2006a). For an optimized microchannel imaging systems, for example, this effectively resulted in the combination and integration of the microlens array with an additional macroscopic lens element, which represents the variation of the array parameters across the array (Sinzinger and Jahns 2003, Chap. 8). The fabrication of such a microlens array with spatially varying microlens properties is even more challenging than the fabrication of a regular microlens array. A diffractive implementation is possible without additional effort, while refractive implementations, for example, can be realized using direct writing schemes such as direct electron or laser beam writing (Wippermann *et al.* 2006b).

Recently nonuniform arrays of microlenses for which the term chirped arrays has been coined have been demonstrated for various applications. The additional design freedom of the individual microlenses can be applied, for example, for the optimization of specific imaging situations like microchannel systems for ultraflat camera objective lenses (Duparré *et al.* 2005a,b). Inhomogeneous microlens arrays can also be advantageous as beam homogenizers (Wippermann *et al.* 2007). Thus, for a completely random distribution of the microlens arrays within the array, a very uniform illumination profile without peaks due to the periodicity results. In this case, the array factor is a constant, which depends on the number of microlenses. However, there is a certain peak observed in the zeroth diffraction order (Lohmann 2006, p. 241).

3.02.2.3 Fabrication of Microlenses

In this section, we outline some of the most widely applied principles for the fabrication of microlens or a microlens array. In Section 3.02.2, the structure of diffractive microlenses by identifying them as specific forms of diffraction gratings has been discussed. The fabrication of diffractive microlenses is possible through standard lithographic approaches using suitable optical materials. In this section, the focus is on the fabrication of refractive microlenses. The most widely used techniques such as photoresist reflow, ultraprecision micromachining for micro-optics fabrication, as well as the ion exchange technique for the fabrication of gradient index (GRIN) microlenses have been briefly described.

3.02.2.3.1 Functional principles of optical components

As described above, the TEA optical elements are identified as phase profiles modifying the incident wavefront. For refractive optical components in general and thus also in microlens arrays this phase profile is generated by a variation of the optical path length (OPL) of optical rays incident onto the component at different lateral locations. The OPL is a product of the physical path length l and the refractive index $n(\lambda)$ of the propagation medium at the corresponding wavelength:

$$\mathrm{OPL} = ln(\lambda) \qquad [31]$$

From this equation two fundamental possibilities to generate a local variation of the OPL (or a phase profile) can be identified (**Figure 21(a)**) (Sinzinger and Jahns 2003). If the light, for example, propagates along different physical path lengths through a dielectric medium with homogenous refractive index, the concept of conventional refractive lenses results. Their functionality can also be described by refraction of the rays at the surface, which defines the element. Alternatively, the physical path length

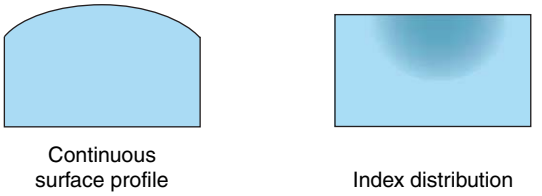


Figure 21 The functionality of refractive microlenses generated by (a) a surface profile and (b) a refractive index distribution.

inside the element is constant for all locations but the refractive index of the dielectric medium varies locally (**Figure 21(b)**). In this case, at least if we have a smoothly varying refractive index profile we speak of GRIN optical media. Propagation in such a medium can also be approximated in geometrical optics by assuming a large number of transitions between media with slightly different refractive indices (Iga *et al.* 1984).

Based on these two basic functional principles of refractive microlenses, various different fabrication technologies can be applied for their fabrication.

3.02.2.3.2 Reflow microlenses

The conceptually simplest and quickest techniques for microlens fabrication exploit the natural effect of surface tension. This leads to the formation of spherical islands, for example, once liquids are dropped onto surfaces or once cylindrical islands on substrates are melted. These spherical islands or caps can be exploited as microlenses. The most widely applied technique has been suggested originally by Popovic who fabricated cylindrical islands of photoresist through lithographic techniques (Daly 2001, Hutley 1990, Popovic *et al.* 1988). Upon heating these photoresist islands above the melting temperature of the resist material, spherical microlenses are formed. In spite of this straightforward fabrication concept the technology proves to be rather challenging. Specifically, for the reproducible fabrication of microlenses with specific optical parameters, it is necessary to control all processing parameters to very high precision. The significance of the different parameters can be understood by considering various parameters influencing the fabrication result.

The focal length of the resulting plano-convex microlenses can be determined from the refractive index of the lens material n and the radius of curvature r_c of the spherical surface:

$$f_{\text{reflow}} = \frac{r_c}{n-1} \qquad [32]$$

In order to determine r_c, we compare the volume of the cylindrical islands of photoresist after lithographic fabrication V_{cyl} and the spherical photoresist caps V_{cap} after the reflow process:

$$V_{\text{cyl}} = \pi\left(\frac{D}{2}\right)^2 t; \quad V_{\text{cap}} = \pi h^2 \frac{3r_c - h}{3} \qquad [33]$$

where D is the diameter of the photoresist cylinders, which is also considered to be the diameter of the resulting microlenses; t is the thickness of the photoresist layer; and h is the height of the spherical photoresist caps resulting after melting (sometimes also called the sag of the microlens) (**Figure 22**). For simplifying the following derivation, we assume that the photoresist volume and the diameter of the contact area between the resist and the substrate surface do not change during the reflow process. Even though these are rather crude assumptions, the results show the relations that we are interested in. For more precise calculations, the volume shrinkage as well as a detailed analysis of the surface energies need to be performed.

Under the above assumptions, we can derive equations for the radius of curvature as well as the relation between the resist thickness and the sag and diameter of the formed microlenses:

$$r_c = \frac{h^2 + \frac{D^2}{4}}{2h}; \quad t = \frac{h}{6}\left(3 + 4\frac{h^2}{D^2}\right) \qquad [34]$$

The impact of these equations is illustrated by the curves in **Figure 23**. They show the dependence of the focal lengths on the diameter and on the thickness of the photoresist layer. It is obvious that in order to work in an appropriate parameter region it is necessary to be able to fabricate rather thick photoresist layers with very good uniformity. This proves to be one of the most serious challenges in the fabrication of

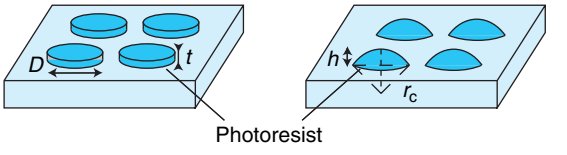


Figure 22 Photoresist reflow for shaping spherical microlenses from lithographically fabricated photoresist cylinders.

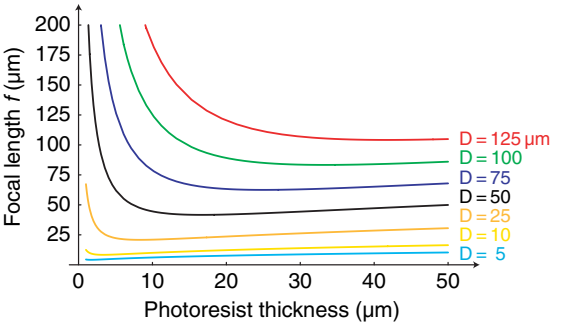


Figure 23 Illustration of the relations in eqn [34]; focal lengths resulting from the reflow process as a function of the photoresist thickness for various lens diameters.

microlenses with diameters larger than several hundred microns. Thanks to the progress in the development of highly viscous photoresists, very uniform layers with thicknesses of several tens of microns can be fabricated nowadays. The next step after the coating of sufficiently large and uniform photoresist layer thickness is the melting process, which can become quite challenging if the cylinder volume is quite large. The ideal lens shape is achieved only if the whole resist volume reaches temperatures above the glass temperature simultaneously. Once the volume is too large, the outer regions reach high temperatures too quickly and can be cured before the center portions reach the melting point. These and other additional problems result in the fact that very good reflow microlenses can be fabricated only in a rather limited parameter range up to diameters of about 500 μm. At the same time, surface tension always results in rather small radii of curvature, i.e., large NAs. The reflow process is thus applicable for microlenses with diameters <2 mm and NA > 0.1 (Sinzinger and Jahns 2003, Chap. 4). Efforts to extend this parameter range include preshaping of the photoresist as well as the application of contact coatings in order to influence the contact angles between the resist and the substrate and as such also the radius of curvature. In spite of the technological challenges outlined above, this fabrication approach has been commercialized and successfully applied to the fabrication of very large arrays of microlenses. Obviously to this end, the additional challenges of uniform processing conditions over the whole wafer area need to be solved (Suss 2006).

In order to transfer the photoresist microlenses into materials with advantageous optical and physical parameters, such as chemical, mechanical, or thermal stability, generally a reactive ion etching process is performed. This allows the transfer of the profile into the glass substrates. Through a precise control of the composition of the processing gas, the relative etching rates between the photoresist and the glass can be controlled (so-called differential etching) in such a way that the profile depth inside the substrate can be changed (Eisner and Schwider 1996, Stern and Jay 1994). Even the profile itself can be optimized by a variation of the conditions during the etching process. Alternatively, galvanic forming and replication techniques can be applied to the transformation and replication of the reflow photoresist microlenses (Gale 1997).

A large variety of approaches to microlens fabrication have been suggested, which are all based on the principles of surface tension. The conceptually simplest technique uses small droplets of curable materials such as polymers, which are applied to a planar substrate. Due to surface tension, these drops assume spherical profiles, which actually provide microlenses of a reasonable quality. Alternative ways to exploit surface tension in a similar way are techniques that exploit local volume growth or shrinkage for microlens formation. Even in compound semiconductor materials, surface tension is helpful to achieve polishing of multilevel phase profiles to achieve analog refractive micro-optical elements (Herzig 1997; Sinzinger and Jahns 2003, Chap. 4).

3.02.2.3.3 Ultraprecision micromachining for microlens fabrication

Within the last two decades, the mechanical precision of micromachining tools has been refined to a very high precision. Thus, the fabrication of high-quality microlenses and microlens arrays is nowadays feasible with high precision micromachining such as single point diamond turning or ultraprecision milling. Diamond turning provides the very good surface quality for spherical surface profiles. Ultraprecision milling, on the other hand, provides more flexibility with respect to aspherical or freeform profiles. However, the surface quality achieved with these milling processes is limited and additional polishing steps are necessary for high-quality elements. In spite of the challenges related to the fabrication of micro-optical elements using ultraprecision micromachining, the approach looks very promising and is the only fabrication approach providing microlenses with diameters in the range of a few millimeters, which is not accessible for alternative methods. This is even more important since these sizes can turn out to be the most important for application in complex microsystems (Lee and Cheung 2002, Zhang *et al.* 2005).

3.02.2.3.4 Planar GRIN microlenses fabricated by ion exchange – GRIN lenses

GRIN elements are a different type of refractive optical elements. In GRIN components, rather than being refracted at a surface profile, the light is deflected by a refractive index gradient within the substrate. In this section, the focus is on the fabrication of spherical index gradients and their application as microlenses. Historically, the field of GRIN optics gained significance with the introduction of GRIN optical fibers. The guiding mechanism in these fibers is based on a radial refractive index profile in the core

of the GRIN fibers. This index profile leads to the bending of the ray trajectory in such a way that for suitable shapes of the index profiles the light is constantly refocused during propagation through the fiber (**Figure 24**). In comparison to conventional step index fibers, this has the tremendous advantage that GRIN fibers can be designed to have zero waveguide mode dispersion. Mode dispersion is the relative difference between the propagation constants for different waveguide modes (which can be identified with propagation directions) and normally limits the transmission bandwidth of multimode waveguides. Thus, even if a large number of modes are guided in the GRIN fiber with suitable index profile all OPLs of the light through the fiber are equal. In other words, the GRIN fiber acts like a series of microlenses distributed along the fiber length such that the optical beam is constantly refocused when propagating through the index gradient. Thus, it has been realized soon that a short piece of such a GRIN fiber effectively acts as a microlens if the index profile is designed properly (Rawson *et al.* 1970, Tomlinson 1980). In the first approximation, the parabolic index profile is described by the following equation

$$n(x) = n_0\left(1 - \frac{1}{2}\alpha^2 x^2\right) \qquad [35]$$

where α is the design parameter defining the shape of the profile. In this case, the ray trajectory can be approximated as in **Figure 24** (Saleh and Teich 1991):

$$x(z) = x_0 \cos(\alpha z) + \frac{\Theta_0}{\alpha}\sin(\alpha z) \qquad [36]$$

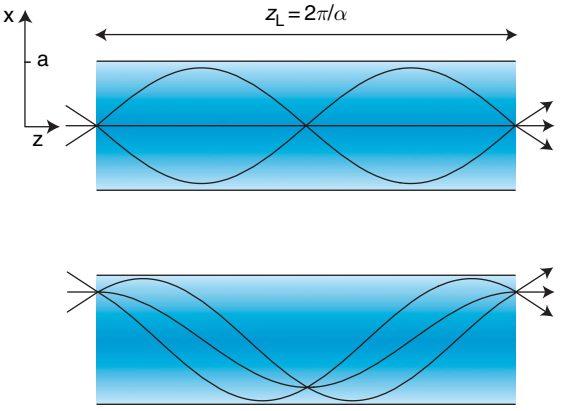


Figure 24 Light propagation in GRIN rod lenses with radial index profiles. Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

Due to their lengthy shape, those lenses consisting of a short piece of GRIN fiber are called GRIN rod lenses and became well known under the trade name SELFOC™ (micro-) lenses also (Iga *et al* 1984). Large numbers of those rod lenses stacked to a 1D array have been used in scanners, copying machines, and laser printers for scanning and imaging the document onto the photosensitive drum. As in microlens imaging in general, the use of microlenses in this application enables the imaging of widely extended objects with very compact systems. For this application, GRIN rod lenses have the additional advantage that they can be designed to form upright images with a magnification of 1. With conventional microlenses this is achieved through stacking of at least two microlens arrays, which leads to additional efforts with respect to the assembly technology.

On the other hand, due to the complex fabrication process of fiber pulling, GRIN rod lenses cannot be easily fabricated in batch-like processes. Nor is it possible to fabricate 1D or 2D arrays of such lenses for integration in complex microsystems directly. Therefore, in the early 1980s the focus in GRIN micro-optics shifted to the fabrication of planar GRIN microlenses. Different from GRIN rod lenses for planar lens functionality, index profiles with specific point symmetries are necessary. As a suitable fabrication technology an ion exchange process in glass has been suggested by Oikawa and later on refined by numerous other authors (Oikawa 1994, Oikawa and Iga 1982).

In order to generate the index profile with point symmetry, which is necessary to achieve planar microlens functionality, the substrate (generally a glass substrate with well-controlled properties and ion composition) is coated with a mask providing local protection from the subsequent ion exchange process. After etching the mask structures (e.g., small circular openings can be used for microlens fabrication), the substrate is immersed into a melt containing ions suitable for replacing some of the ions from the substrate (**Figure 25**). The ion exchange takes place at temperatures of several hundred degrees Celsius in a diffusion-like process. At these temperatures, the ion mobility is large enough so that some of the ions (generally the alkali ions, such as Na^+) from the glass are replaced by ions from the melt (e.g., Ag^+ or Tl^+). Due to different polarization characteristics of the ions, this exchange influences the dielectric properties of the glass matrix and effectively changes the refractive index of the substrate. For a given material and ion combination,

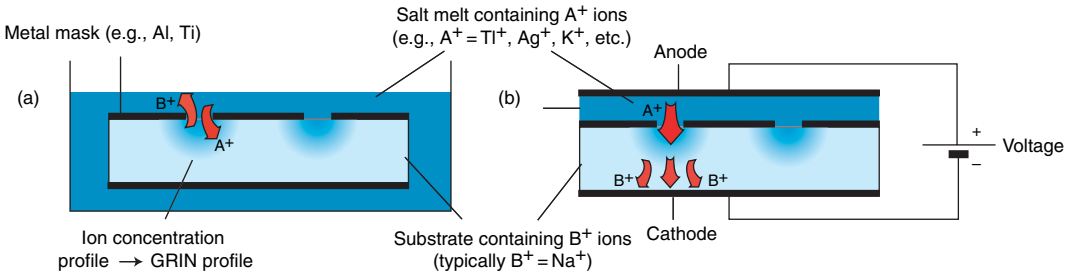


Figure 25 The ion exchange process for the fabrication of planar GRIN. (a) Thermal ion exchange and (b) field-assisted ion exchange. Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

the amount of index change is proportional to the percentage of exchanged ions. Since the ion exchange is an isotropic process, it does not stop at the edges of the mask. The ratio of exchanged ions drops gradually from a maximum at the surface of the substrate areas with no mask layer to zero at the maximum penetration depth of the ions inside the substrate (typically on the order of several hundred micrometers) or at similar distances in a lateral direction under the mask. An ion concentration profile with point symmetry is generated, which causes a refractive index distribution of a corresponding shape. The resulting refractive index profile is characteristic of the substrate and the exchanging ion pairs. If the exchange process is sufficiently nonlinear, the radial index distribution acts as a microlens. In order to accelerate and optimize the thermal ion exchange process described above, an electric field can be applied during the diffusion process. In the field-assisted ion exchange, also called electromigration, the electric field is used to increase the ion mobility across the interface between the substrate and the salt melt. Due to this electric field, the movement of the ions from the melt into the substrate is accelerated and the penetration depth of the ions is increased significantly. Additionally the shape of the electric field directly influences the resulting index profile.

The ion pairs used in the exchange process and the type of glass substrate are the most important parameters in the ion exchange process. A variety of different ion pairs have been investigated for the fabrication of planar GRIN microlenses. The most promising results were achieved with sodium–thallium and sodium–silver ion exchange. Both ion combinations yield comparable maximum index changes of about $\delta n = 0.1$ at the substrate surface in commercially available glass substrates (Findakly 1985, Oikawa 1982). Extremely large index changes (e.g., $\delta n > 0.2$) have been achieved in glasses with high sodium content (~25 vol.%) (Coutaz and Jaussand 1982, Misawa *et al.*1984). These substrates tend to suffer from high internal tension, which builds up because of the different diameters of the exchanged ions. Therefore, most of the microlenses have been fabricated with lower index changes being achieved in glasses with lower sodium content.

Generally, it is the maximum achievable index modulation that determines the NAs of the resulting planar GRIN microlenses. Consequently, the variety of NAs that has been demonstrated are rather limited and restricted to fairly small values of $NA < 0.1$ at typical diameters below 1 mm. Large NAs up to $NA = 0.6$ have been demonstrated by Oikawa; however, in this case a significant portion of the focusing effect stems from the surface profile effect, which was caused by surface swelling during the ion exchange process. This was caused by the fact that the thallium ions diffusing into the glass substrate have a larger diameter than the sodium ions, which are replaced in the glass matrix.

Bähr *et al.* demonstrated that complex mask shaping in combination with field-assisted diffusion and postdiffusion are useful tools in order to significantly extend the parameter region of ion-exchanged GRIN microlenses. Most importantly, they managed to fabricate microlens arrays with very high fill factors (Bähr and Brenner 2001).

3.02.2.4 Implementation of Active Microlenses

For various applications, it is desirable to be able to tune the parameters of an optical system within a certain range of parameters. This is a specific challenge in compact microsystems where the manual tuning or the integration of actuators for automatic tuning by positioning of the individual components is challenging. In this section, some of the techniques

that have been suggested for the fabrication of tunable microlenses are discussed briefly.

3.02.2.4.1 Microlenses based on liquid crystal elements

3.02.2.4.1.(i) Fundamentals of liquid crystals

Liquid crystal materials are the most widespread materials used in order to achieve dynamic optical functionality. This is due to the tremendous development in the field of flat panel displays, which has been triggered by the consumer market. In order to understand the functionality and the tuning behavior of liquid crystal-based optical elements, it is necessary to discuss the microscopic structure of LC materials. Without going into details, we try to focus on the basics necessary to understand the working principles.

After the discovery of the LC effect on cholesteryl benzoate by the Austrian biologist F. Reinitzer in the late 19th century, the challenge to be overcome before the effect could become of practical use was to find LC materials that were sufficiently stable at room temperature. Nowadays, a large variety of different types of LC materials are available and are used in an even larger variety of applications (Chigrinov 1999, Collings and Hird 1997, Lueder 2001).

The microscopic structure of LCs can be considered to consist of long chain-like molecules (other shapes of the molecules such as disc-like shapes can also be observed but can be ignored here for simplicity). In contrast to regular liquids, these organic or polymer molecules are subject to significant intermolecular forces. In the LC phase, these intermolecular forces result in the arrangement of the molecules in regular orientations as it is common in solid (crystalline) materials. Like crystalline materials, LCs can also change their intrinsic order in a phase transition, i.e., a transition between different states of order. The anisotropic structure of the molecules in combination with the intrinsic order of the LC results in interesting anisotropic optical functionality. A further consequence of the anisotropic susceptibility of the molecules is that, for example, external electric fields can be used for the dynamic manipulation of the states of order in the crystals. Energy considerations lead to the observation that the molecules align along the field lines of the external electric field.

The optical functionality of the LC material depends on the relative orientation of the molecules with respect to the polarization of the incident wavefront. In the liquid phase with the molecules floating around statistically, every orientation has the same probability. Thus, no anisotropic optical functionality is observed. However, the intrinsic order in the LC phase leads to optical anisotropy, i.e., birefringence. Different states of polarization transmitted through the LC thus experience different refractive indices. If the material thickness is optimized this anisotropy can be exploited to achieve a rotation of the polarization direction for the transmitted light. If the orientation of the LC molecules is changed by applying an electric field of a specific orientation, the anisotropy of the LC is changed and so is the polarization state of the transmitted light. If stacked between two polarizers with corresponding orientation such as a switchable LC cell, it can thus be used for amplitude modulation. This is the very basic functional principle of LC displays.

For efficient (micro-)optical elements the implementation of phase only modulators is desirable. To this end, the orientation of the LC molecules and the polarization of the incident wavefront need to be synchronized. Phase-only modulation requires that the plane of polarization is always oriented parallel or perpendicular to the mean orientation of the LC molecules (called the director). In order to avoid a change in the polarization state upon propagation through the cell, this needs to be true for all switching states. Such a configuration is possible if the molecules are aligned in the plane parallel to the cell surface. If the external electric field then triggers a rotation around an axis parallel to the surface but perpendicular to the director, light polarized parallel to the director experiences a variation in the refractive index but no influence on the state of polarization. In this case, the plane of polarization remains parallel to the main axes of the LC. Thus, the anisotropy does not affect the incident direction of polarization.

In spite of the identical fundamental functional principle, a large variety of alternate LC structures have been developed mainly for display purposes. During the fabrication of the cell, it is possible to influence the orientation of the LC molecules in the relaxed state, i.e., in absence of an electric field. The structure of the substrate surfaces, for example, has a significant influence on the alignment of the LC molecules. The addressing concept of the LC layer is of crucial importance for the quality of the optical functionality. For display purposes, the active matrix thin film transistor (TFT) structure provides best results with respect to contrast, interpixel crosstalk, and gray scale representation.

3.02.2.4.1.(ii) Liquid crystal (micro-)lenses

3.02.2.4.1.(ii).(a) Dynamic microlenses on pixelated LC devices

Throughout the last decade, the quality of LC TFT displays has improved significantly. Enhanced microstructuring techniques allow fabrication of pixelated displays with pixel sizes significantly smaller than 10 μm and very high fill factors in the area of 95%, i.e., very little space between the pixels. Once the polarizers that are necessary to achieve gray scale display functionality are removed such high-quality displays enable the implementation of phase diffractive lens functionality. Generally, refractive lenses cannot be implemented in LC cells for displays since the limited thickness of the LC layers enables only a phase modulation depth of about 4π at visible wavelengths. This modulation depth is sufficient for standard diffractive lenses with decent performance. Limitations of such diffractive elements can be easily derived from the relations about diffractive microoptics presented above. The pixel size, for example, which results in a coarse spatial quantization, limits the NA as well as the quality of diffractive microlenses (Carcolé *et al.* 1994, Davis *et al.* 1989, Tam *et al.* 1992).

3.02.2.4.1.(ii).(b) Modal liquid crystal lenses

Liquid crystal lenses with analog refractive phase profiles can be realized using relatively thick layers of LC in order to enable a large maximum phase shift when applying an external electric field (Naumov *et al.* 1998). **Figure 26** shows the basic structure of a modal LC lens. The necessary analog phase profile is generated through a suitable electrode structure generating the electric field inside the LC layer. The LC response to high-frequency voltage can be exploited in order to generate the suitable electric field distribution. Once a high-frequency voltage is applied to the annular electrode, the cell can be considered as a series of low-pass filters (**Figure 27**). The LC acts like a capacitor in this context. The longer the effective high resistance electrode section, the larger is the voltage drop in this layer. This low-pass characteristic is described by the following equation:

$$V_o = V_i \frac{|X_C|}{\sqrt{X_C^2 + X_R^2}}, \quad \text{with } X_C = \frac{1}{j\omega C} \text{ and } C \sim \frac{1}{d} \qquad [37]$$

where V_i and V_o are the input and output voltages to the low-pass filter, respectively; X_C and X_R are the complex resistances of the capacitors and the resistor, respectively, and C is the capacitance of the LC layer.

The remaining voltage drop occurs across the LC leading to the stable electric field distribution inside the crystal. Thus, from the characteristic of low-pass filters, we find that for a suitable relation between the high-resistance electrode and the capacity of the LC layer a parabolic electric field distribution results. Since the capacitance of the layer depends on the

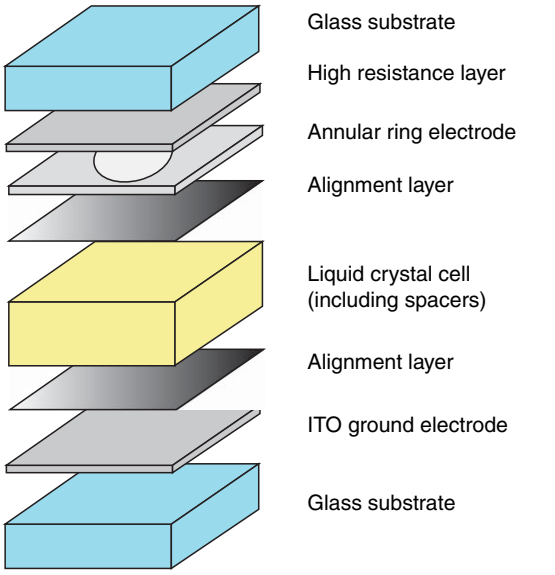


Figure 26 Schematic structure of a liquid crystal modal lens.

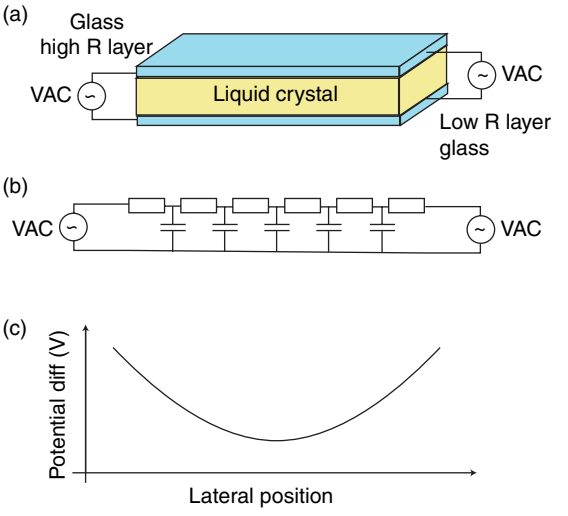


Figure 27 Electric equivalent of a modal liquid crystal lens. (a) Contacting of the cell; (b) electric equivalent as a series of low pass filters; and (c) potential profile inside the liquid crystal.

layer thickness, the resistance of the electrode layer needs to be adjusted to the corresponding layer thickness in order for the modal effect to occur at reasonable frequencies.

For modal lenses, nematic LCs are used and aligned homogeneously. This is achieved, for example, through a slight polishing step, where the substrate surface gets the necessary preferential orientation along which the LC molecules tend to align. In the electric field, the LC molecules align about the ordinary axis and change the effective refractive index of light polarized properly. Portions of the incident beam polarized perpendicular to the director will not be affected by the change of the cell due to the electric field. By stacking two of these cells with perpendicular orientation, polarization-independent modal lenses can be realized (Love 1993). The theoretical phase modulation depth, which can be achieved with modal lenses, can be calculated from the birefringence of the LC and the thickness of the LC cell. Typical values for birefringence are $\Delta n = 0.225$, which results in a phase depth of about seven wavelengths in a cell with a thickness of 20 μm. This also illustrates the limitations of this technological approach to fabricate dynamically tunable microlenses. Due to the relatively small relative refractive index change, a large thickness of the layer is necessary in order to achieve decent microlens performance. On the other hand, increasing the thickness results in slow tuning speeds as well as increasing stray light due to scattering in the LC volume. Modal LC microlenses thus have rather low NAs in the range of 10^{-2}.

One of the advantages of this approach for the fabrication of tunable microlenses can be found in the planar cell structure. Since the LC lenses exploit index variations of the LC material rather than surface profiles, the surface of the cell is not changed during the tuning process. Thus, these LC lenses are extremely well suited for micro-optical systems integration, e.g., in planar optical systems. A planar integrated micro-optical system that incorporates dynamic functionality based on modal LC microlenses has recently been demonstrated (Amberg *et al.* 2006).

3.02.2.4.2 *Liquid tuneable microlenses*

The difference between the index modulation generated in tunable LC microlenses, is that in liquid microlenses, the alternative approach is to tune the shape of the elements (**Figures 28** and **29**). As in fabrication of reflow microlenses, the natural formation of spherically shaped droplets through surface tension is exploited in these cases. It has been discussed above that the spherical droplets themselves provide decent lens functionality. This approach has the advantage, that effectively, the large index difference between the liquid lens medium and the environment is available for the optical functionality and the tuning. This results in a significantly wider tuning range, which is available for liquid lenses compared to tunable optical elements based on index modulation, e.g., in LCs. Here, the concept of two different approaches to the generation of liquid tunable lenses is discussed.

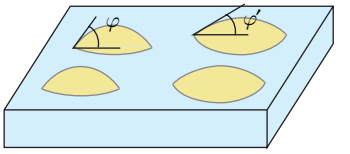


Figure 28 Concept of switching in liquid lenses. The wetting angle of the liquid lens can be changed dynamically.

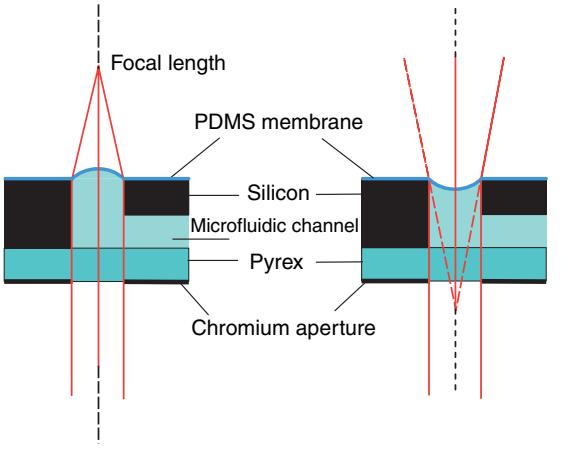


Figure 29 Basic structure of a tunable liquid lens. (Source: Werber and Zappe 2005.)

3.02.2.4.2.(i) *Microfluidic liquid microlenses*

The implementation of tunable microlenses with the help of complex fluidic systems has been demonstrated by several authors. In the general approach, microcavities filled with liquid are used, which are covered by a flexible membrane (e.g., PDMS of a thickness of 50 μm) (Chen *et al.* 2004, Krupenkin *et al.* 2000, Sugiura and Morita 1993, werber and zappe 2005). By varying the pressure inside the cavity, this membrane is bent to form a plano-concave microlens of a variable radius of curvature. Arrays of microlenses with comparable properties can be

realized by connecting a single reservoir to various membrane-covered microcavities through a microfluidic channel system. Analog tuning of the pressure leads to analog tuning of the focal length. With the application of negative pressure, the implementation of negative, that is, plano-concave, microlenses is also possible. This yields a large tuning range, which so far has not been achieved by any other technological approach. Due to the fact that these liquid microlenses are working on the basis of refraction at the interface between the liquid and the surrounding atmosphere, relatively large NAs can be achieved (Xie and Lu 2005). Werber *et al.* demonstrated that high optical quality of the lens performance can be achieved over the large tuning range (Werber and Zappe 2005). Instead of mechanically controlling the pressure Dong suggested the use of hydrogel materials, which can be activated, for example, by chemical (pH)-, thermal-, electromagnetic-, or electric field-induced stimuli. By fabricating a hydrogel ring structure this stimulus can be exploited for pressurizing the microfluidic channel and thus tune the microlens (Dong *et al.* 2006).

3.02.2.4.2.(ii) Electrowetting for the fabrication of tunable microlenses A further technological approach to activate microlens fabrication is based on a phenomenon called electrowetting (Bergé and Peseaux 2000, Vallet *et al.* 1999). Here, the formation and tuning of the shape of liquid droplets is realized through the application of electric fields. By tuning a voltage applied between an electrode immersed into a droplet of conductible or polar liquid and an electrode below the dielectric substrate carrying the droplet, the shape of the droplet can be tuned within a broad range. This effect can be best understood by performing an energy balance considering the surface energy between the liquid and the substrate γ_{sl} and the electrostatic energy $(1/2)\varepsilon\varepsilon_0(U^2/d)$ in the capacitor between the two electrodes:

$$\gamma_{sl}(U) = \gamma_{sl} - \frac{1}{2}\varepsilon\varepsilon_0\frac{U^2}{d} \qquad [38]$$

Here, U is the applied voltage and d is the thickness of the dielectric layer (dielectric constant ε). This surface energy, which depends on the applied voltage, influences the wetting angle between the substrate and the liquid and thus the extension of the droplet.

For practical purposes, the electrode in contact with the droplet is disturbing the optical performance of these liquid droplet microlenses. In order to overcome this problem, Mönch *et al.* (2005) developed a microsystem where the electrowetting liquid is immersed into a microcavity together with a nonpolar liquid, which effectively forms the lens (**Figure 30**). By varying the wetting angle between the polar liquid and the sidewalls of this microcavity the shape of the lens liquid inside the cavity can be varied. For this effect, the name competitive electrowetting has been coined. Compared to the conventional electrowetting effect, where high voltages result in strong positive lens functionality, the effect is reverse in competitive wetting since the polar liquid is not identical with the lens liquid. Competitive electrowetting has significant advantages with respect to lens quality and stability (vibration damping). However, the total achievable optical functionality and tuning range are reduced

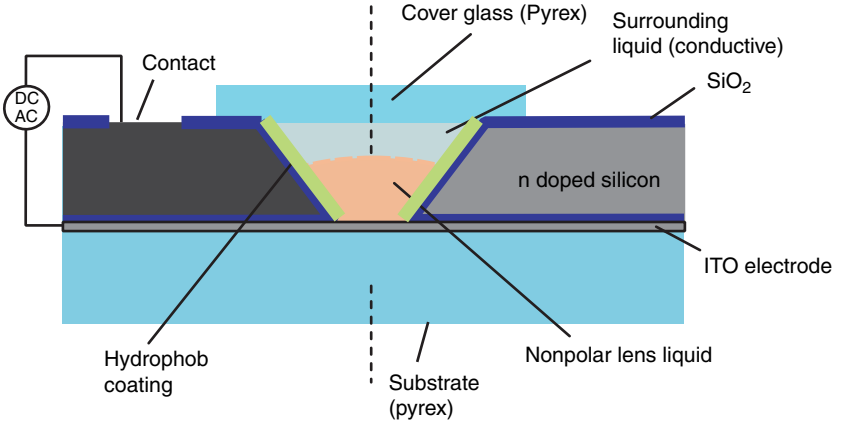


Figure 30 Schematic setup of tunable liquid microlenses generated by competitive electrowetting. (Source: Mönch W, Krogmann F, Zappe H 2005.)

due to the embedding of the tunable lens into the polar liquid, which causes a significant immersion effect. Consequently, microlenses demonstrated with this technology so far showed rather low NAs. A linear tuning range over a relatively broad range of focal lengths was demonstrated by Mönch *et al.* (2005).

Miniature- or microlenses are currently being commercialized for applications in miniature camera objectives for mobile phones or personal digital organizers (pdas).

3.02.2.5 Applications of Microlens Arrays

3.02.2.5.1 Wavefront sensing – Shack–Hartmann sensor

One of the most widespread application of microlens arrays is in wavefront sensing. To this end, the microlens array is used for wavefront division. Each single section of the wavefront is focused onto a position sensitive detector by the individual microlenses. Thus, the position is a measure for the inclination (gradient) of the specific section of the wavefront (Malacara 1992). The local gradients of the wavefront expressed by the inclination angles α_x and β_y with respect to the x- and y-axes, respectively, can be calculated from the following equation (**Figure 31**):

$$\sin\alpha_x = \frac{\Delta x}{\sqrt{\Delta x^2 + \Delta y^2 + f^2_{\text{microlens}}}}; \qquad \sin\beta_y = \frac{\Delta y}{\sqrt{\Delta x^2 + \Delta y^2 + f^2_{\text{microlens}}}} \qquad [39]$$

From the measurement of these local gradients the shape of the wavefront can be calculated, e.g., by polynomial fitting. One challenge connected to Shack–Hartmann wavefront measurements is interchannel crosstalk due to difficulties when assigning the focus spots to the correct microchannels, which is specifically important for the measurement of wavefronts with large local gradients (i.e., curvatures). This problem has been overcome with optimized evaluation algorithms resulting in a large achievable dynamic range. The smallest radius of curvature that can be measured with Shack–Hartmann sensors is approximately given by twice the wavelength of the individual microlenses (Pfund and Beyerlein 2002). The measurement precision depends on the spatial resolution of the position sensing system. Subpixel resolution can be achieved with suitable algorithms when determining the center of gravity of the focus spots (Pfund *et al.* 1998).

The conceptually very simple principle of Shack–Hartmann sensors allows the implementation of high-quality wavefront sensors with very simple optical setups. In recent decades, the development of arrays of microlenses with small submillimeter apertures and very good uniformity combined with high-resolution CCD or CMOS cameras and optimized evaluation software have triggered a wide range of new application areas for Shack–Hartmann sensors. Alternative methods for high-quality wavefront sensing require, for example, complex interferometric setups, which are potentially very sensitive to instabilities of the configuration, and often require multiple measurements for unambiguous results (Koch *et al.* 2000).

As a specific advantage, Shack–Hartmann wavefront sensors yield very quick results, which is specifically important, for example, for real-time applications in adaptive optics. The evaluation of the wavefront sensor is directly used for an optimization of the optical system in order to achieve optimum optical performance. Adaptive systems and thus Shack–Hartmann sensors are of great importance, e.g., in medical imaging or astronomy where the effects of atmospheric distortions on the optical imaging quality are to be corrected. Further fields of application are laser beam characterization and consumer optics testing (Hardy 1978, Tyson 1998).

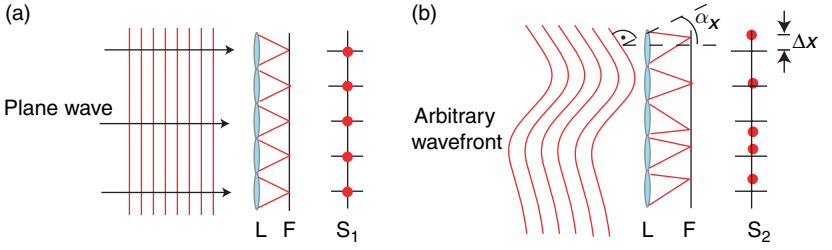


Figure 31 Schematic illustrating the concept of the Shack–Hartmann wavefront sensor. (a) Plane wave incident onto a microlens array L and top view of the focal plane S1 (distance F) and (b) arbitrary wavefront incident onto the microlens array L (again with top view of the focal plane S_2). Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

Finally, Shack–Hartmann sensing is an impressive application of large microlens arrays. Due to the potentially very fast measurements, Shack–Hartmann sensors have also been suggested for array testing of the properties of the array of micro-optical elements, specifically microlens arrays. Due to the very large number of elements present in these arrays, it is a serious challenge to test, e.g., arrays uniformity. With precise Shack–Hartmann measurements, which allow a thorough characterization of an individual element within a few seconds, it can become affordable to even perform 100% testing of arrays of thousands of micro-optical elements (Beyerlein *et al.* 2005).

3.02.2.5.2 Microlens array imaging

For various situations in everyday life it is desirable to develop ultracompact optical systems for imaging or image sensing. Compact digital cameras, for example, are nowadays integrated into mobile phones or handheld organizers. Very often, the size of the objective lenses really determines the limits to the integration of the whole system. Even though very small microlenses can be fabricated they have strict fundamental limitations with respect to the space bandwidth product (i.e., the number of resolvable pixels) as has been pointed out earlier. Thus, for conventional single pupil imaging systems there is a direct conflict between the size of the imaging system and the achievable image quality. Microlens array imaging to a certain extent allows overcoming this trade-off. There is a direct relation between the field size, the lens diameter, and the length of the imaging system. In order to be able to reduce the overall extension of the whole imaging system it is necessary to reduce the diameter of the individual imaging elements. In this case, a large number of those individual pupils are necessary in order to provide the necessary field size and overall pupil diameter. This is the concept of microchannel imaging (**Figure 19**) (Andersen 1979).

Besides the potentially very compact longitudinal extension of the microchannel imaging system, an additional advantage results from the discussion of the scaling behavior of optical lens elements. Since for large lens diameters it is becoming increasingly difficult to control wavefront aberrations, the quality of the individual microlenses used in microlens array imaging can be significantly lower. Also, because microlens arrays can be fabricated in large arrays using lithographic fabrication and replication processes, such systems can be significantly cheaper than the systems with large lenses. At the same time the systems are compact since the geometrical length l is now determined by the diameter d of the microlenses rather than by the diameter of the total image field.

However, for the generation of high-quality images with this setup several aspects have to be considered in designing the system. First, if a single microlens array imaging system is to be applied for the imaging task, each of the microlenses generates an inverted image of a portion of the object plane. Unless the object has a very specific symmetry, this local inversion distorts the image. This can be avoided by using two microlens arrays forming an array of microtelescopes (Sinzinger and Jahns 2003, Chap. 14).

Since the field of view (FOV) of each microlens is not limited to the diameter of the lens, multiple images of each part of the object field are formed in the image plane. The origin for these alternate imaging channels, for example, can be due to diffraction at the multiple aperture grid of the microlens array or due to scattering effects. Geometrical considerations illustrate that these multiple images coincide with a strictly symmetrical configuration and form a so-called composite image. The relation to be fulfilled for the formation of the composite image is given by the following equation:

$$z_C = z_O \qquad [40]$$

However, as illustrated in **Figure 32**, this does not necessarily occur in the same plane I, where sharp images are formed by the individual microtelescopes. The distance z_I can be found by applying the basic imaging equation for each of the two imaging steps. If $z_I \neq z_C$, in the image plane, multiple images of the object occur as indicated in **Figure 32**. High-quality imaging is possible in symmetrical setups, where the

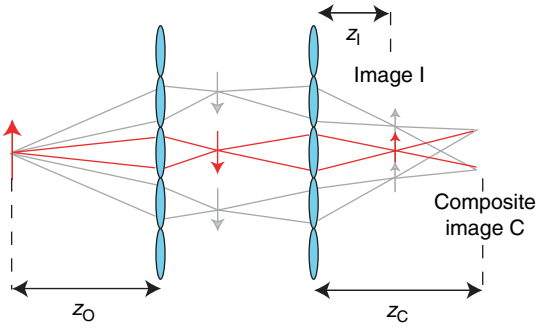


Figure 32 Microlens array imaging: geometrical considerations (detailed explanations see text). Adapted from Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim, Germany, with permission.

composite image plane coincides with the image plane of the microlenses. This symmetrical configuration leads to imaging systems that are free from antisymmetrical aberrations such as coma or distortion.

Second, in completely symmetrical microsystems where the image plane I and the composite image plane C coincide, the only source of background noise stems from light transmitted through noncorresponding microlenses. Field lenses as well as high aspect ratio walls separating the optical channels help to reduce this background noise due to crosstalk between the imaging channels (Duparré *et al.* 2004, Völkel *et al.* 1996).

Furthermore, for the design of a microlens array imaging system, it is necessary to consider the image field size of the microlenses as well as the fill factor of the arrays. For covering the whole image plane uniformly, it is required that the lenses are capable of imaging field sizes larger than the lens area. Depending on the lens quality, this can be achieved at the expense of lateral resolution if larger object and image distances are chosen.

Array imaging systems as described above have already been used for years in fax machines, photocopiers, and similar applications. Frequently, 1D arrays of GRIN SELFOC lenses are used for imaging. In such systems, it is important to achieve good uniformity in the intensity over the image plane. Aperture stops in the entrance and exit pupils as well as in the field lenses can be applied for increasing the uniformity. Combinations of refractive and diffractive microlens arrays with interesting imaging behavior have been demonstrated (Andersen 1979). The improvement of lithographic fabrication technology for microlens arrays with good imaging quality and high packaging density allows constructing array imaging systems for high-resolution imaging. One of the high-end applications of such systems is microlens lithography, which is being commercialized, for example, by Suss Microtec, Inc., Neuchatel (Suss 2006). In this application, microlens arrays are used for the imaging of a lithographic mask onto the wafer. Due to the small aperture, even diffraction-limited microlens imaging systems do not allow submicron resolution. Thus, with regard to the lateral resolution they cannot compete with sophisticated lens systems designed, for example, for high-resolution lithography. Microlens array imaging systems with a resolution of 2–5 μm have been demonstrated. The potential of microlens lithography lies in the possibility of using arrays with a large number of microlenses in imaging systems, which can, nevertheless, be compact and cheap. They can be applied for projection lithography of large field sizes without the need of step and repeat or scanning systems, which cause subsequent stitching errors. In microchannel or microlens lithography, compared to lensless proximity printing, which can also conceptually work on very large field sizes, large depths of focus are achieved. This is of specific importance for applications in microsystems fabrication, where often resist depths of several tens to hundreds of micrometers are to be achieved. Due to the parallel arrangement of a large number of microimaging systems, the overall SBP is very large. Of course, the fabrication of huge numbers of microlenses with uniform optical properties over the whole array turns out to be a severe technological challenge.

Significant efforts have been undertaken to optimize ultracompact imaging systems based on microlens arrays. Duparré *et al.* (2004) fabricated flat camera objective lenses, for example, for mobile phone systems. The challenge in this context is to image a large environmental scene onto a small camera detector, e.g., a CCD or a CMOS array. Again, the microchannel approach is necessary in order to avoid the unfavorable longitudinal size of single pupil imaging systems. Duparré *et al.* suggested an approach where each pixel is illuminated through an individual microlens channel. This configuration corresponds to the so-called apposition eye, which is well known from nature since the eyes of small invertebrates are built this way (Horridge 1977, Snyder 1977). In order to achieve the large viewing angle, in this configuration, the microlens array is combined with an array of detectors, eventually combined with pinholes, with a pitch significantly smaller than the microlens array. From geometrical considerations, it is easy to understand the functional principle (**Figure 33**). Each microlens array generates an individual image of the object. The pinholes in front of the detector array are actually sampling each of these images at specific positions. Thus, the detector signals combine to a compound image of the complete object. The maximum FOV of such an apposition imaging system is thus determined by the NAs of the individual microlenses:

$$\mathrm{FOV} = \arctan\left(\frac{a}{f}\right) < \arctan\left(\frac{p}{f}\right) = 2\mathrm{NA} \qquad [41]$$

Here, a is the size of an image formed by the individual microlenses. Consequently, a is smaller than the diameter D of the microlenses, which

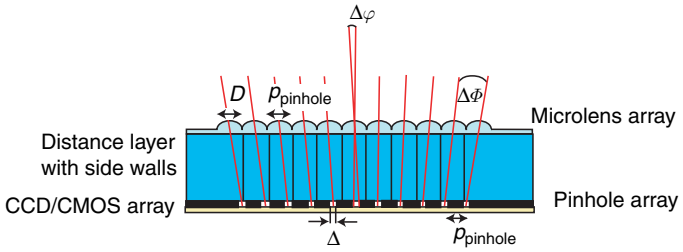


Figure 33 Compound eye imaging sensor; due to the mismatch in the period of the microlens array and the pinhole array, a large field of view can be achieved.

corresponds to the optimum fill factor of the pitch $p_{\text{microlens}}$ of the microlens array.

Other important parameters of the system are the viewing angle difference $\Delta\Phi$ between the individual microchannels, the so-called interommatidial angle, and the acceptance angle of the individual microchannel $\Delta\varphi$ where

$$\begin{aligned}\Delta\Phi &= \arctan\left(\frac{\Delta p}{f}\right);\\ \Delta\varphi &= \arctan\left[\frac{\text{FWHM}(\Delta \otimes \text{PSF}_{\text{microlens}})}{f}\right]\\ &\approx \sqrt{\left(\frac{\Delta}{f}\right)^2 + \left(\frac{\lambda}{D}\right)^2}\end{aligned} \qquad [42]$$

and $\Delta p = p_{\text{microlens}} - p_{\text{pinhole}}$ is the pitch difference between the microlens and the pinhole arrays.

FWHM denotes the full width half maximum diameter, $\text{PSF}_{\text{microlens}}$ denotes the point spread function of the individual microlens (diameter, D), Δ is the size of the pinhole, and $\otimes$ denotes the convolution. From the relations in eqn [42] the achievable angular resolution can be estimated (Duparré *et al.* 2005a). In the compound eye configuration, as every microlens is used for the generation of one pixel of the image, the overall resolution is quite limited. Thus, it can be more appropriate to apply this approach for the implementation of imaging sensors as has already been suggested and demonstrated by Sanders and Halford (1995).

Again in such a multichannel imaging concept, interchannel crosstalk is the most severe challenge. This can be avoided by stacking the microlens and the pinhole array with an array of absorbing walls separating the channels optically. These walls can be fabricated by ultraprecision micromachining or deep etching techniques (**Figure 33**).

For an optimization of the imaging quality, i.e., for the correction of the large field curvature, the array of microlenses needs to be fabricated with spatially variable lens properties (Duparré *et al.* 2005b, Sinzinger and Jahns 1997). The potentially large viewing angles of the microlenses at the edges of the described apposition eye imaging systems result in a significant amount of spherical aberrations. This can be reduced by adjusting the properties of the individual imaging channels. To this end, the imaging properties of each of the microlenses are adjusted to the specific imaging situation of the individual microchannels. In this case, the regularity of the microlens array is destroyed. For microlens arrays with varying properties of the individual microlenses the term chirped microlenses has been coined (Duparré *et al.* 2005b).

The most critical feature of microlens arrays for imaging systems is the uniformity of the microlenses. The availability of cheap high-quality microlens arrays will further increase the interest in array imaging techniques such as integral photography or Gabor's superlens. Here, the array imaging approach is generalized to the use of microlens arrays that differ in pitch or focal lengths. Significant potential can also be found in combination with DOEs to form more general array of optical elements (Davies *et al.* 1994, Hembd-Solner *et al.* 1999, Stevens and Harvey 2002).

3.02.2.5.3 Laser diode beam shaping and beam homogenization

One of the commercially most interesting applications of microlenses is laser diode beam shaping and beam homogenization (Leger 1997, Turunen and Wyrowski 1997). The challenge here is to cope with the different divergence angles as well as the different waist positions of the laser radiation emitted, for example, from edge-emitting laser diodes. Microlenses with specific astigmatism that can be integrated directly within the semiconductor substrate can potentially perform this task. An approach for their realization as GRIN microlenses

using ion exchange in glass substrates has been suggested. Such approaches in combination with optimized microtechnological approaches for the assembly and integration of the corresponding elements directly with the laser diodes offer unique possibilities for compact laser sources with preshaped beam profiles.

For many tasks, for example, in material processing or lithography, very homogeneous beam profiles are required. To this end, it is necessary to convert a Gaussian beam profile into a rectangular profile. Various different approaches to beam homogenization mostly based on DOEs have been suggested (Sinzinger *et al.* 1995). One of the most widespread techniques applies microlens arrays in the configuration of a fly's eye condenser. However, in this case, as described above, the periodicity of the array (array factor) results in destruction of the periodicity due to strong peaks in the diffraction pattern. Recently, promising approaches that focus on the use of irregular microlens arrays have been suggested (Wippermann *et al.* 2007).

3.02.3 Summary and Conclusions

Diffractive optics and microlenses are of crucial importance in microsystems, which incorporate optical functionality. This chapter outlines the fundamental theoretical and technological aspects of these fields. The main focus is to present an overview, describe current trends as well as most popular current applications, and provide crosslinks to the literature for a more thorough treatment of the various topics. It is certainly safe to forecast that microoptical elements and systems based on diffractive optics and microlens technology will even gain in significance with the progress of functional integration in complex microsystems in the near future.

Acknowledgment

The author would like to acknowledge the active collaboration with an enthusiastic team of colleagues and students with whom it is a lot of fun to work. Specific thanks also goes to agencies and state departments who provide funds for our research, which helps us to keep learning, i.e., the Thüringer Minsisterium for Wirtschaft, Technologie und Arbeit, the Thüringer Kultusministerium, The German and Europeas Science foundations (DFG and ESF). Specific thanks for proof reading go to Sebastian Stoebenau who did his job once more with amazing diligence.

Most importantly I would like to thank my wife Sabine and daughter Ronja for everything.

References

Amberg M, Oeder A, Hands P J W, Love G, Sinzinger S 2006 Tunable planar integrated optical systems. *OSA Annual Meeting*. Rochester, NY, USA

Andersen R H 1979 Close-up imaging of documents and displays with microlens arrays. *Appl. Opt.* **18**, 477–84

Bähr J, Brenner K-H 2001 Realization of refractive continuous phase profile elements with high design freedom by mask structured ion exchange. *Prof. SPIE* **4437**, 50–60

Bartelt H 1984 Computer-generated holographic component with optimum light efficiency. *Appl. Opt.* **23**, 1499–502

Bergé B, Peseaux J 2000 Variable focus lens controlled by an external voltage: An application to electrowetting. *Eur. Phys. J. E* **3**, 159–63

Beyerlein M, Pfund J, Dorn R 2005 Automatisierte 100%-Charakterisierung mikrooptischer Bauteile. *Photonik* **4/2005**, 64–6

Born M, Wolf E 1980 *Principles of Optics*, 6th edn. Pergamon Press, London

Borrelli N F 1999 *Microoptics Technology: Fabrication and Applications of Lens Arrays and Devices*. Marcel Dekker, New York, NY

Bryngdahl O, Wyrowski F 1990 Digital holography – Computer generated holograms. In: Wolf E (ed.) *Progress in Optics XXVIII*. North Holland, New York, NY

Cao Q, Jahns J 2002 Focusing analysis of the pinhole photon sieve: Individual far field model. *JOSA A* **19**, 2381–93

Carcolé E, Campos J, Bosch S 1994 Diffraction theory of Fresnel lenses in low resolution devices. *Appl. Opt.* **33**, 162–74

Cassasent D (ed.) 1978 *Optical Data Processing*. Springer, New York, NY

Caulfield J 1979 *Handbook of Optical Holography*. Academic Press, San Diego, CA

Chen J, Wang W, Fang J, Varahramyan K 2004 Variable focusing microlens with microfluidic chip. *Micromech. Microeng* **14**, 675–80

Chigrinov V G 1999 *Liquid Crystal Devices: Physics and Applications*. Artech House, Boston, MA

Collings P J, Hird M 1997 *Introduction to Liquid Crystals*. Taylor and Francis, London

Coutaz J L, Jaussand P C 1982 High index gradient in glass by ion exchange. *Appl. Opt.* **21**, 1063–5

Daly D 2001 *Microlens Arrays*. Taylor and Francis, London

Davies N, McCormick M, Brewin M 1994 Design and analysis of an image transfer system using microlens arrays. *Opt. Eng.* **33**, 3624–33

Davis J A, Cottrel D M, Lilly R A, Connely S W 1989 Multiplexed phase encoded lenses written on spatial light modulators. *Opt. Lett.* **14**, 420–2

Dobschal H J, Steiner R, Rudolf K, Hage K, Brunner R 2001 *German Pat*. 101 30 212 6

Dong L, Agarwal A K, Beebe D J, Jiang H 2006 Adaptive liquid microlens activated by stimuli-responsive hydrogels. *Nature* **442**, 551–3

Dorsch R, Lohmann A W, Sinzinger S 1994 Fresnel ping-pong algorithm for two-plane computer-generated display holograms. *Appl. Opt.* **33**, 869–75

Duparré J, Dannberg P, Schreiber P, Bräuer A, Tünnermann A 2004 Artificial apposition compound eye fabricated by micro-optics technology. *Appl. Opt.* **43**, 4303–10

Duparré J, Dannberg P, Schreiber P, Bräuer A, Tünnermann A 2005a Thin compound-eye camera. *Appl. Opt.* **44**, 2949–56

Duparré J, Wippermann F, Dannberg P, Reimann A 2005b Chirped arrays of refractive ellipsoidal microlenses for aberration correction under oblique incidence. *Opt. Express* **13**, 10539–51

Eisner M, Schwider J 1996 Transferring resist lenses into silicon by RIE. *Opt. Eng.* **35**, 2979–82

Fienup J R 1982 Phase retrieval algorithms – A comparison. *Appl. Opt.* **21**, 2758–69

Findakly T 1985 Glass waveguides by ion-exchange. *Opt. Eng.* **24**, 244–50

Froehly C, Colombeau B, Vampouille 1983 Shaping and analysis of picosecond pulses. In: Wolf E (ed.) *Progress in Optics XX.* North Holland, New York, NY

Gale M 1997 Replication. In: Herzig H P (ed.) *Micro-optics: Elements, Systems and Applications.* Taylor and Francis, London

Gerchberg R W, Saxton W O 1972 A practical algorithm for the determination of phase from image and diffraction plane figures. *Optik* **35**, 237–246

Goodman J 1996 *Introduction to Fourier Optics*, 2nd edn. McGraw-Hill, Boston, MA

Gross H 2005 *Fundamentals of Technical Optics*. In: Gross H (ed.) *Handbook of Optical Systems.* Wiley-VCH, Weinheim, Germany, Vol. 1

Gruber M 2001 Diffractive optical elements as raster image generators. *Appl. Opt.* **40**, 5830–9

Hardy J W 1978 Active optics: A new technology for the control of light. *Proc. IEEE* **66**, 651–97

Hembd-Solner C, Stevens R F, Hutley, M C 1999 Imaging properties of the Gabor superlens. *J. Opt. A Pure Appl. Opt.* **1**, 94–102

Herzig H P (ed.) 1997 *Micro-optics: Elements, Systems and Applications.* Taylor and Francis, London

Horner J L (*ed*.) 1987 *Optical Signal Processing.* Springer, New York, NY

Horridge G A 1977 The compound eyes of insects. *Sci. Am.* **237**, 108–20

Hutley M C 1990 Optical techniques for the generation of microlens arrays. *J. Mod. Opt.* **37**, 253–65

Iga K, Kokubun Y, Oikawa M 1984 *Fundamentals of Microoptics*. Academic Press, Tokyo

Jahns J 1994 Diffractive optical elements for optical computing. In: Jahns J and Lee S H (eds.) *Optical Computing Hardware*, Academic Press, San Diego, CA, pp. 137–67

Jahns J, Acklin B 1993 Integrated planar optical imaging system with high interconnection density. *Opt. Lett.* **18**, 1594–6

Jahns J, Lohmann A W 2004 Temporal filtering by double diffraction. *Appl. Opt.* **43**, 3306–9

Jahns J, Walker S J 1990 Two-dimensional array of microlenses fabricated by thin film deposition. *Appl. Opt.* **29**, 931–6

Jahns J, Lohmann A W, Bohling M 2006 Talbot bands and temporal processing of optical signals. *JEOS* **1**, 06001

Javidi B 2006 *Optical Imaging Sensors and Systems for Homeland Security Applications*. Springer, New York, NY

Javidi B, Horner J L (eds.) 1994 *Real Time Optical Information Processing.* Academic Press, Boston, MA

Joannopoulos J D, Meade R D 1995 *Photonic Crystals.* Princeton University Press, Princeton, NJ

Kipp L, Skibowski M, Johnson R L, Berndt R, Adelung R, Harm S, Seemann R 2001 Sharper images by focussing soft X-rays with photon sieves. *Nature* **4**, 184–8

Knop K 1978 Rigorous diffraction theory for transmission gratings with deep rectangular profiles. *J. Opt. Soc. Am.* **68**, 1206–10

Koch J A, Presta R W, Sacks R A, Zacharias R A, Bliss E S, Dailey M J, Feldmann M, Grey A A, Holdener F R, Salmon J T, Seppala L G, Toeppen J S, Van Atta L, Van Wonterghem, Whistler W T, Winters S E, Woods B W 2000 Experimental comparison of a Shack–Hartmann sensor and a phase-shifting interferometer for large-optics metrology applications. *Appl. Opt.* **39**, 4540–6

Konishi T, Ichioka Y 1999 Ultrafast image transmission by optical time-to-two-dimensional-space-to-time-to-two-dimensional-space conversion. *JOSA A* **16**, 1076–88

Kraus Th F, De La Rue R M 1999 Photonic crystals in the optical regime – Past, present and future. *Prog. Quant. Electr.* **23**, 51–96

Krupenkin T, Yang S, Mach P 2000 Tunable liquid micro-lens. *Appl. Phys. Lett.* **82**, 316–18

Lee W B, Cheung B C F 2002 *Surface Generation in Ultra-precision Diamond Turning: Modelling and Practices.* John Wiley and Sons, New York, NY

Leger J R 1997 Laser beam shaping. In: Herzig H P (ed.) *Micro-optics: Elements, Systems and Applications*, Taylor and Francis, London

Li Y, Platzer H 1983 An experimental investigation of diffraction patterns in low Fresnel number focussing systems. *Opt. Acta* **30**, 1621–43

Li Y, Wolf E 1981 Focal shifts in diffracted converging spherical waves. *Opt. Commun.* **39**, 211–15

Lohmann A W 1971 An interferometer based on the Talbot effect. *Opt. Commun.* **2**, 413–15

Lohmann A W 1988 An array illuminator based on the Talbot effect. *Optik* **49**, 41–5

Lohmann 1989 Scaling laws for lenses. *Appl. Opt.* **28**, 4996–8

Lohmann A W 2006 *Optical Information Processing*. Universitätsverlag Ilmenau, Germany

Love G 1993 Liquid crystal modulators for unpolarized light. *Appl. Opt.* **32**, 2222–3

Lueder E 2001 *Liquid Crystal Devices*. John Wiley and Sons, Chichester, NY

Malacara D 1992 *Optical Shop Testing.* Wiley, New York

Misawa S, Oikawa M, Iga K 1984 Maximum and effective numerical aperture of a planar microlens. *Appl. Opt.* **23**, 1784–6

Miyamoto K 1961 The phase Fresnel lens. *JOSA* **17**, 17–21

Mönch W, Krogmann F, Zappe H 2005 Variable Brennweite durch flüssige Mikrolinsen. *Photonik* **4/2004**, 44–6

Nakai T, Ogawa H 2002 Research on multi-layer diffractive optical elements and their applications to photographic lenses. *Tech. Digest. 3rd Int. Conf. Optics-photonics Design and Fabrication*. Tokyo, Japan

Nakai T, Ogawa T 2000 *European Pat*. 107 29 06

Naumov A F, Koktev M Y, Guralnik I R, Vdovin G V 1998 Liquid crystal adaptive lenses with modal control. *Opt. Lett.* **23**, 992–8

Naumov A F, Love G D, Loktev M Yu, Vladimirov F L 1999 Control optimization of spherical modal liquid crystal lenses. *Opt. Express* **4**, 344–52

Nishihara H, Suhara T 1987 Micro Fresnel lenses. In: Wolf E (ed.) *Progress in Optics XXIV.* North Holland, New York, NY

Nuss M C, Li M, Chiu T H, Weiner M C 1994 Time to space mapping of femtosecond pulses. *Opt. Lett.* **19**, 664–6

Oikawa M 1994 Physics of planar microlenses. In: Jahns J, Lee S H (eds.) *Optical Computing Hardware.* Academic Press, San Diego, CA

Oikawa M, Iga K 1982 Distributed index planar microlens. *Appl. Opt.* **21**, 1052–6

Petit R (ed.) 1980 *Electromagnetic Theory of Gratings.* Springer, Berlin

Pfund J, Beyerlein M 2002 Shack–Hartmann-Sensoren für die Qualitätskontrolle in klassischer und Laser Optik. *Photonik* **4/2002**, 6–8

Pfund J, Lindlein N, Schwider J, Burow R, Blümel Th, Elssner K-E 1998 Absolute sphericity measurement: A comparative study on the use of interferometry and a Shack–Hartmann sensor. *Opt. Lett.* **23**, 742–4

Pfund J, Lindlein N, Schwider J 2001 Nonnull testing of rotationally symmetric aspheres: A systematic error assessment. *Appl. Opt.* **40**, 439–46

Popovic Z D, Sprague R A, Connel G A N 1988 Technique for the monolithic fabrication of microlens arrays. *Appl. Opt.* **27**, 1281–4

Rawson E G, Herriot D R, McKenna J 1970 Analysis of refractive index distributions in cylindrical graded-index glass rods (GRIN rods) used as image relays. *Appl. Opt.* **9**, 753–9

Ruffieux P, Scharf T, Herzig H P, Völkel R, Weible K J 2006 On the chromatic aberrations of microlenses. *Opt. Express* **14**, 4687–94

Saleh B, Teich M 1991 *Fundamentals of Photonics*. Wiley Interscience, London

Sanders J S, Halford C E 1995 Design and analysis of apposition compound eye optical sensors. *Opt. Eng.* **34**, 222–35

Schwider J, Sickinger H 1997 Arraytests for microlenses. *Optik* **107**, 26–34

Sheppard C J R 1986 Imaging in optical systems of finite Fresnel number. *JOSA A* **3**, 1428–32

Shiono T, Ogawa H 1992 Reflection aspherical microlens for planar optics fabricated by electron beam lithography. *Opt. Lett.* **17**, 565–7

Sinzinger S 2002 Microoptically integrated optical correlators for security applications. *Opt. Commun.* **209**, 69–74

Sinzinger S, Jahns J 1997 Integrated microoptical imaging system with high interconnection density fabricated in planar optics. *Appl. Opt.* **36**, 4729–35

Sinzinger S, Jahns J 2003 *Microoptics*, 2nd edn. Wiley-VCH, Weinheim, Germany

Sinzinger S, Brenner, K- H, Moisel J, Spick T, Testorf M 1995 Astigmatic gradient-index elements for laser diode collimation and beam shaping. *Appl. Opt.* **34**, 6626–32

Snyder A W 1977 Acuity of compound eyes: Physical limitations and design. *J. Comp. Physiol. A* **116**, 161–82

Soifer V, Kotlyar V, Doskolovich L 1997 *Iterative Methods for Diffractive Optical Elements Computation.* Taylor and Francis, London

Solymar L, Cooke D J 1981 *Volume Holography and Volume Gratings*. Academic Press, London

Stern M B, Jay T R 1994 Dry etching of coherent refractive microlens arrays. *Opt. Eng.* **33**, 3547–51

Stevens R F, Harvey T G 2002 Lens arrays for three-dimensional imaging systems. *J. Opt. A Pure Appl. Opt.* **4**, 17–21

Streibl N, Nölscher U, Jahns J, Walker S 1991 Array generation with Lenslet arrays. *Appl. Opt.* **30**, 2739–42

Sugiura N, Morita S 1993 Variable-focus liquid-filled optical lens. *Appl. Opt.* **32**, 4181–6

Suss 2006; http.//www.suss-microoptics.com

Tam E C, Zhou S, Feldman M 1992 Spatial-light-modulator-based electro-optical imaging system. *Appl. Opt.* **31**, 578–80

Tomlinson W J 1980 Aberrations of GRIN rod lenses in multimode optical fibers. *Appl. Opt.* **19**, 1117–26

Turunen J, Wyrowski F 1997 Diffractive optics for industrial and commercial applications. Akademie Verlag, Berlin

Tyson R K 1998 *Principles of Adaptive Optics.* Academic Press, Boston, MA

Vallet M, Vallade M, Bergé B 1999 Limiting phenomena for the spreading of water on polymer films by electrowetting. *Eur. Phys. J. B* **11**, 583–91

Vander Lugt A 1992 *Optical Signal Processing.* Academic Press, Boston, MA

Völkel R, Herzig H P, Nussbaum P, Dändliker R, Hugle W B 1996 Microlens array imaging system for microlens lithography. *Opt. Eng.* **35**, 3323–30

Wang W, Friberg A T, Wolf E 1993 Structure of focused fields in systems with large Fresnel numbers. *JOSA A* **12**, 1947–53

Weaver C S, Goodman J W 1966 A technique for optically convolving two functions. *Appl. Opt.* **5**, 1248–9

Weiner A M, Leard D E, Reitze D H, Paek E G 1992 Femtosecond spectral holography. *IEEE J. Quant. Electron.* **28**, 2251–61

Werber A, Zappe H 2005 Tunable microfluidic microlenses. *Appl. Opt.* **44**, 3238–45

Wippermann F, Radtke D, Amberg M, Sinzinger S 2006a Integrated free-space optical interconnect fabricated in planar optics using chirped microlens arrays. *Opt. Express* **14**, 10765–78

Wippermann F, Radtke D, Zeitner U, Duparré J W, Tünnermann A, Amberg M, Sinzinger S, Reinhardt C, Ovsianikov A, Chichkov B N 2006b Fabrication technologies for chirped refractive micro-lens arrays. *SPIE Proc. Vol. 6288 Current Developments in Lens Design and Optical Engineering VII*, USA, San Diego, CA, pp. 14–15

Wippermann F, Dannberg P, Bräuer A, Sinzinger S 2007 Improved homogenization of fly's eye condenser setups under coherent illumination using chirped microlens arrays. *Photonics West 2007*, San Jose, CA, USA

Xie Y, Lu Z 2005 Tunable liquid microlens with three-dimensional adjustment of the position of the focal plane. *J. Opt. Tech.* **72**, 445–7

Zhang C, Rentsch R, Brinksmeier E 2005 Advances in Ultrasonic assisted lapping of microstructures in hard-brittle materials: A brief review and outlook. *Int. J. Machine Tools Manufact.* **45**, 881–90

Biography

Stefan Sinzinger received his Dipl.-Phys. and Dr. degrees from the Friedrich-Alexander Universität (FAU) Erlangen-Nürnberg (Germany), Institute for Applied Optics (Prof. Dr. A. W. Lohmann), in 1989 and 1993, respectively, where he was involved in the design, fabrication, and characterization of diffractive and refractive micro-optical elements. Between 1990 and 1991 he spent 9 months at the NEC Research Center performing research on computer holography. In 1994, he joined Prof. Dr. J. Jahns at the Institute for Information Technology (Optische Nachrichtentechnik) at the Fernuniversität Hagen where he was performing research on planar micro-optical systems integration. In 1998 he received the Bennigsen-Foerder Award for young scientists in Nordrhein-Westfalen for the project Mikrooptischer Lesekopf für Kopierschutzhologramme (Micro-optical Readout Device for Security Holograms). After finishing his Habilitation on Planar-optical Systems Technology in 2002 he became Professor of Optical Engineering (Technische Optik) at the Technische Universität Ilmenau, Fakultät für Maschinenbau (Mechanical Engineering Department). He (co-)authored more than 30 publications in peer reviewed journals as well as more than 70 conference proceedings including various invited presentations. He is coauthor of the textbook Microoptics published in the second edition in 2003 by Wiley-VCH and editor of the textbook Optical Information Processing (author A.W. Lohmann) published in 2006 by the Universitätsverlag Ilmenau. He is performing as a reviewer for various international optical journals and is Associate Editor of the SPIE Journal on Microlithography, Microfabrication, and Microsystems. His current research is focused on the design, fabrication, and integration of micro-optical elements and hybrid optofluidic microsystems, the manipulation of micro- and nanoparticles through electromagnetic fields, as well as ultra-precision machining for optical microsystems.

3.03 Photonic Crystals

Susumu Noda[1], Department of Electronic Science and Engineering, Kyoto University, Kyoto, Japan

Nomenclature

2D	Two-dimensional
3D	Three-dimensional
PC	Photonic constant
a	Lattice period
f_i	Resonance frequency
Q	Quality factor
Q_i	Quality factor in-plane
Q_v	Quality factor vertically
SOI	Silicon on insulator
λ	Wavelength
T	Transmittance
R	Reflectance
D	'Drop' efficiency, for emission vertically from a defect
V	Modal volume

3.03.1 Introduction

The photonic crystal represents one of the most exciting recent developments in optics. Through periodic structuring of optically transparent materials with features whose spacing is on the order of the wavelength of light, a photonic band structure is generated in the optical transmission spectrum (Noda and Baba 2003, Yablonovitch 2000). Thereby, in the same manner that the electronic band gap, which results from the periodic atomic lattice structure in crystalline materials, prohibits the transport of electrons in a certain energy range, the photonic band gap prohibits the transport of photons in a certain energy range.

As shown in **Figure 1**, both 2D and 3D photonic crystals may be conceived (Ho *et al.* 1990, Noda *et al.* 1996, 2000), in which the uniform periodic structure prohibits transmission of light for wavelengths in the photonic band gap. Such photonic band gap structures become even more interesting through the introduction of defects, as shown in **Figure 2**, allowing light to be controlled in a number of ways. For example, by introducing a line-shaped defect, it is possible to form an ultra-small waveguide that permits the transmission of light only along the defect. Alternatively, light can be trapped at certain points by introducing point defects, thereby forming a photonic nanocavity.

By combining these line and point defects, it is possible to design ultra-small photonic circuits (or chips) with various functions. Furthermore, the band gap itself makes it possible to suppress spontaneous emission, which is a fundamental factor limiting the performance of various photonic devices. Finally, the recent introduction of the photonic heterostructure concept has further expanded the spectrum of applications for photonic crystals.

Photonic crystals are thus artificially structured materials whose optical behavior may be finely tuned by varying their periodic features. Two-dimensional structures, more easily fabricated than their 3D counterparts, form the basis for a wide

[1] Introduction by Hans Zappe, Department of Microsystems Engineering, University of Freiburg, Freiburg, Germany

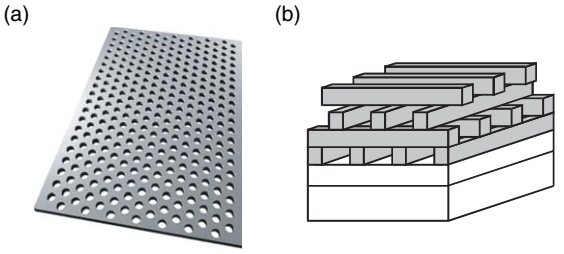


Figure 1 Schematic representations of (a) a 2D photonic crystal and (b) a 3D photonic crystal.

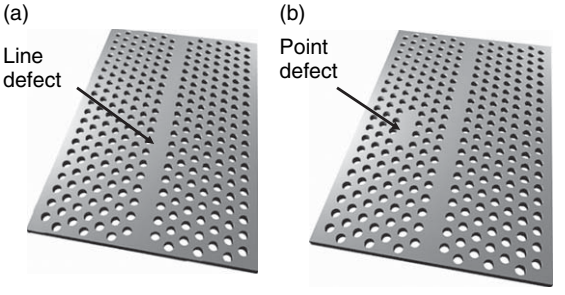


Figure 2 Schematic representations of (a) a line-defect waveguide introduced into a 2D photonic crystal slab and (b) a combined point-defect resonator and line-defect waveguide system.

variety of waveguide configurations and have enjoyed considerable popularity in a range of microsystems applications. For this reason, we focus on the design, composition, and application of 2D photonic crystals in this chapter.

3.03.1.1 Photonic Crystals in Microsystems

Optical microsystems are typically assembled from a variety of micro-optical components, including microlenses (Chapter 3.02) and micro-mirrors (Chapter 3.01). Optical systems based on integrated optics (Chapter 3.05) rely on the guiding of light in waveguides; these play a central role in the application of photonic microsystems to optical communications (Chapter 3.07). As photonic crystals move from the laboratory to the real world, they will allow the development of new types of waveguide-based optical components and will ultimately prove to be a driving force in reducing the size while enhancing the functionality of integrated optical systems.

Numerous applications have emerged for photonic crystals (Benisty *et al.* 2006, François *et al.* 2002, Prather *et al.* 2006), with relevance to microsystems. Premier among these are 2D waveguide structures with ultra-compact couplers, splitters and bends, as we will see in greater detail below. Using fabrication techniques borrowed from microsystems engineering, where soft lithography is of particular interest (Pisignano *et al.* 2004), photonic crystal waveguide structures have been used as compact waveguide circuits and circuit boards (Lee *et al.* 2006) and their dispersive properties have allowed the conceptualization of wavelength monitors (Viasnoff-Schwoob *et al.* 2005) as well as multichannel filters (Kee *et al.* 2006). Further uses of photonic crystal structures in passive optical systems include resonators for ultimate use as optical force sensors (Stomeo *et al.* 2006), polarizing beam splitters (Zhen and Li 2005), and large-area biosensors (Block *et al.* 2007).

Photonic crystals have also proven useful when combined with active optical elements, particularly light emitters. For example, nanoimprint lithography has been employed for the fabrication of photonic crystal structures on green LEDs (Kim *et al.* 2007), thereby strongly enhancing the luminescence intensity. Alternatively, using a focused-ion beam to generate the periodic structure on the surface of a vertical cavity laser, single-mode operation with high side-mode suppression has been demonstrated in this class of semiconductor laser (Leisher *et al.* 2006). As photonic crystals continue to find more applications in new and different areas of micro-optics, optical microsystems will benefit from the concomitant improvements in functionality.

3.03.1.2 Microsystems in Photonic Crystals

Just as microsystems may take advantage of advances in photonic crystals, these in turn profit from the capabilities offered by mechanical, electrical, and fluidic microsystems. The primary areas of interest include the use of microsystems to tune photonic crystals or to aid in their assembly.

The etch and assembly technologies used in microsystems engineering have proven beneficial for the flexible manufacture of optical devices based on photonic crystals. A general-purpose platform may be modified to act as a waveguide, a modulator, or a switch, based on details in the fabrication process (Zhou *et al.* 2005, 2006). Sophisticated MEMS assembly techniques have been employed to allow manufacture of 3D photonic crystal devices with considerable flexibility in the number and composition of the vertical layers (Aoki *et al.* 2003).

Finally, prospects for tuning the transmission characteristics of photonic crystals have recently improved due to the availability of various microsystems-based

actuation mechanisms. Precisely applied deformation of suspended membranes into which photonic band-gap structures have been etched has been shown to be effective in tuning the optical behavior (Rajic *et al.* 2003). Nano/micromechanical actuators have been employed to distend a flexible photonic crystal substrate (Park and Lee 2004), essentially stretching the periodic structure and thereby changing its period and thus optical transmission. Tunability has also been achieved using micro-fluidic structures (Chapter 2.13), whereby the refractive index contrast of individual holes of an etched photonic crystal may be addressed and modified using injection of fluids (Erickson *et al.* 2006); the entire microsystem consists of photonic, fluidic, and control layers, assembled into a single opto-fluidic microsystem.

These examples demonstrate the utility of photonic crystals for advanced optical microsystems as well as the importance of MEMS technologies for further advances in the photonics field. As the photonic crystal field matures, we expect that the spectrum of applications will continue to grow.

3.03.2 Two-Dimensional Photonic Crystal Slabs

Two-dimensional photonic crystal research initially targeted structures with periodicity in two dimensions and a third dimension assumed to be of infinite length (Joannopoulos *et al.* 1997), but recently the focus of research has been on slab structures (Noda and Baba 2003, Noda *et al.* 2000a, Painter *et al.* 1999), such as the one shown in **Figure 1(a)**, that have a thickness of the order of the wavelength of the light. The photonic crystal in this figure has a triangular lattice structure. The slab material is assumed to be a high refractive index medium, such as Si and III–V semiconductors, and the lattice points are assumed to be composed of a low refractive index medium, such as air. In 2D photonic crystal slabs, the confinement of light occurs in the in-plane direction due to a photonic band gap effect, and light is confined in the perpendicular direction by total internal reflection due to a difference in refractive indices. Pseudo-3D light control becomes possible as a result. It is, of course, important to optimize the structure so that the leakage of light in the direction perpendicular to the slab is minimized. Recently, as described in Section 3.03.2.1, 2D photonic crystal slabs have been subjected to various types of band gap/defect engineering, and research with the aim of realizing ultrasmall photonic circuits has been advancing steadily.

3.03.2.1 Control of Light by Combined Line- and Point-Defect Systems

As shown in **Figure 2(a)**, when a line defect is introduced into a 2D photonic crystal slab, it will act as an ultrasmall optical waveguide. In order to operate the waveguide efficiently, it is necessary to devise a way in which the leakage of light in the direction perpendicular to the slab can be nullified. In 1999, an optical waveguide experiment was reported (Baba *et al.* 1999), but the perpendicular leakage of light was not taken into consideration and the propagation loss was hence extremely large, exceeding 70 dB mm^{-1}. A year later in 2000, detailed results of an examination of the effects of the slab thickness and the ratio of the low refractive index (the medium (air) that forms the lattice points) to the high refractive index (the slab material) were reported, and a theoretical zero-loss waveguide structure was proposed for the first time (Chutinan and Noda 2000). In 2001, based on these design guidelines, a new waveguide experiment was conducted in which a low propagation loss of 7 dB mm^{-1} was obtained (Notomi *et al.* 2001). At the same time, discussions were taking place regarding the effect of modulation of the waveguide width on various waveguide characteristics. It was also pointed out that an increase in propagation loss was caused by vertical asymmetry in the slab (Tanaka *et al.* 2003). Regarding this, the importance of using crystals with a 2D in-plane full band gap was pointed out (Takayama *et al.* 2005). Recently, an extremely low waveguide loss less than 0.7 dB mm^{-1}, one-tenth of the previous best value, has been achieved (Kuramochi *et al.* 2005, McNab *et al.* 2003, Sugimoto *et al.* 2004b), and it should be possible to fabricate waveguides with propagation losses of less than 1 dB cm^{-1} in the near future. In view of the merits of the miniaturization made possible by the use of photonic crystals, these recent developments show that the minimization of waveguide loss in 2D crystal slabs has advanced to the point where it can almost be ignored.

In conjunction with studies of straight-line waveguides, detailed investigations of waveguide bends have also been carried out. It was initially pointed out that the bending of a waveguide was possible in purely 2D crystals, in regions of very broad wavelength transmission (Joannopoulos *et al.* 1997, Mekis *et al.* 1996), but that, in slab structures, the effect of light scattering to the outside of the slab meant that the low-loss band region in the bend part of the waveguide could not be enlarged to a great extent (Chutinan and

Noda 2000). Thereafter, the importance of controlling lattice point shapes positioned in the curved part was highlighted (Chutinan *et al.* 2002), and the enlargement and the shifting of the band area in the curved region were then reported (Sugimoto *et al.* 2004a, Takano *et al.* 2004). In 2004, it was also shown that low-loss curvature of a waveguide at a very broadband area might be possible even for a bend of 120° (Borel *et al.* 2004). In addition, there have been some studies of low-loss connections between line-defect waveguides and external optical systems (Gomyo *et al.* 2004, Miyai and Noda 2004, Miyai *et al.* 2002, Notomi *et al.* 2004, Shoji *et al.* 2002). Empirical studies have demonstrated that such waveguides can be connected with optical fibers with losses of less than a few decibels (Notomi *et al.* 2004, Shoji *et al.* 2002).

When a point defect is introduced into a photonic crystal, it becomes possible for this to act as a photonic nanocavity, in other words, as a trap of photons (Noda *et al.* 2000a, Painter *et al.* 1999). In photonic crystals, a point defect and a line defect can integrate quite spontaneously; thus, point defects have been extensively discussed with respect to their combination with line defects. High-Q nanocavities and fusion with nonlinear and/or active media will be discussed in Sections 3.03.2.3 and 3.03.2.4. **Figure 2(b)** shows such a combined system of defects (Akahane *et al.* 2003a, Noda *et al.* 2000a). Here, the point defect is formed by filling three lattice points with a high refractive index medium. If the resonant frequency of the point defect is defined as f_i and light with various wavelengths is introduced from the waveguide, light that has a frequency of f_i is trapped by the point defect. The trapped light resonates in the point-defect cavity and is emitted perpendicular to the slab surface due to a breakdown in the total internal reflection conditions. Such optical behavior in a combined line- and point-defect system could be used, for example, as a surface emitting-type ultracompact channel add/drop functional device. It would also be possible to apply it as an ultracompact device for sensing and/or trapping of nanomaterials.

In combined line- and point-defect systems, a maximum of 50% of the light introduced into the waveguide is trapped by the point defect and is emitted to free space (Chutinan *et al.* 2001), which can be deduced by coupled mode theory. This maximum efficiency is obtained when $Q_{in} = Q_v$, where Q_{in} and Q_v are defined as the quality factors of point-defect cavity for in-plane and vertical directions, respectively. More concretely, Q_{in} is determined by the optical coupling between the point-defect cavity and line-defect waveguide, and Q_v is determined by the total internal reflection conditions of the point defect in the direction perpendicular to the slab. The overall Q factor is expressed by $1/Q = 1/Q_{in} + 1/Q_v$ and it determines the wavelength resolution. **Figure 3** shows an

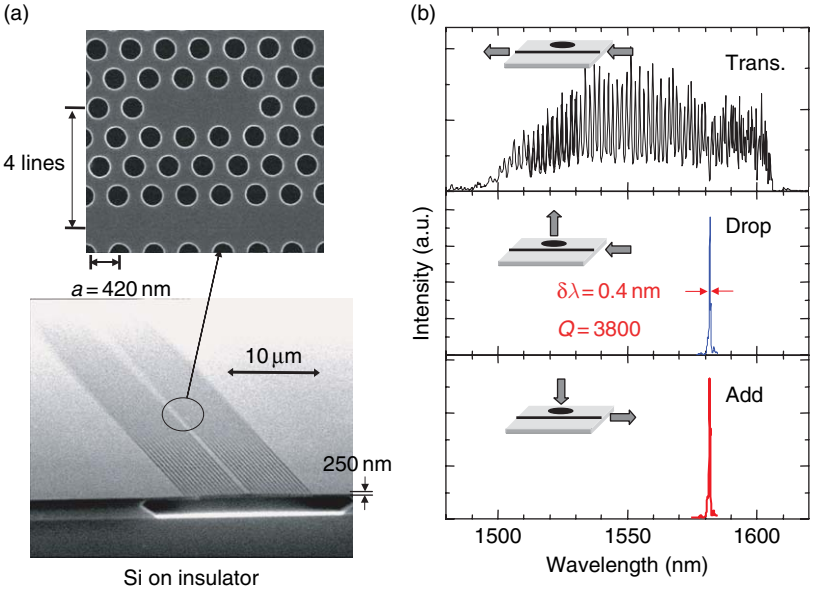


Figure 3 Device based on the photonic crystal slab design in **Figure 2(b)** and its characteristics: (a) electron microscopic image and (b) transmission, drop, and add spectra.

electron microscopic image of a sample composed of a combined point and line defect, and its transmission, drop, and add spectra (Akahane *et al.* 2003a). The device was created using $Si/SiO_2/Si$; that is, an SOI system. A photonic crystal is formed in Si in the top layer. The SiO_2 layer beneath was removed. In the transmission spectrum shown in **Figure 3**, the many sharp resonance peaks observed are attributed to Fabry–Perot resonance at both the edges of the waveguide. This phenomenon can be eliminated by applying an antireflection coating at the waveguide surfaces. An important point to note is the sharp decrease in transmission seen in the vicinity of 1.58 μm, which was attributed to the emission of light through the point defect in the perpendicular direction. **Figure 3** also shows the add spectrum obtained when light was introduced from free space into the point defect. It can be clearly seen that light is effectively entering and exiting through an extremely small point-defect cavity. The resolution determined from these drop and add spectra was estimated to be 0.4 nm, and a relatively large Q value of 3800 was obtained. By comparison with theoretical calculations, a high optical drop or add efficiency of up to 50% was obtained.

As outlined earlier, the use of 2D photonic crystal slabs has allowed the investigations of the basic behavior of line and point defects, individually and in combination, and a basis has been established for the realization of advanced functional photonic circuits.

3.03.2.2 The Effect of Introducing Heterostructures

Along with the advances in band gap/defect engineering described earlier, new concepts for controlling light with a greater degree of freedom are also emerging. One of these is the fabrication of photonic crystal heterostructures (Kawakami *et al.* 2003, Song *et al.* 2003). In the field of semiconductor electronics, the fundamental technology involving heterostructures is concerned with optical/electronic devices such as laser diodes and transistors, the importance of which does not need elaboration here. However, in addition to their functions in electronic systems, heterostructures are also playing important, albeit different, roles in photonic crystal research. The purpose of developing heterostructures in photonic crystals is explained below using an in-plane heterostructure (Song *et al.* 2003) in a 2D crystal slab.

Figure 4 shows an example of an in-plane heterostructure. This photonic crystal has the same basic structure as that shown in **Figure 3**, except that the lattice constant has been slightly varied to give seven different values. **Figure 4** also shows an electron microscopic image of each part of the sample actually created. The schematic drawing emphasizes the variation in lattice constant, but in the fabricated sample the lattice constant in adjacent areas differed by only 1.25 nm, too small to be discerned from the electron microscopic image. One of the promising properties of this type of in-plane heterostructure is the possibility of obtaining an optimized multiwavelength behavior

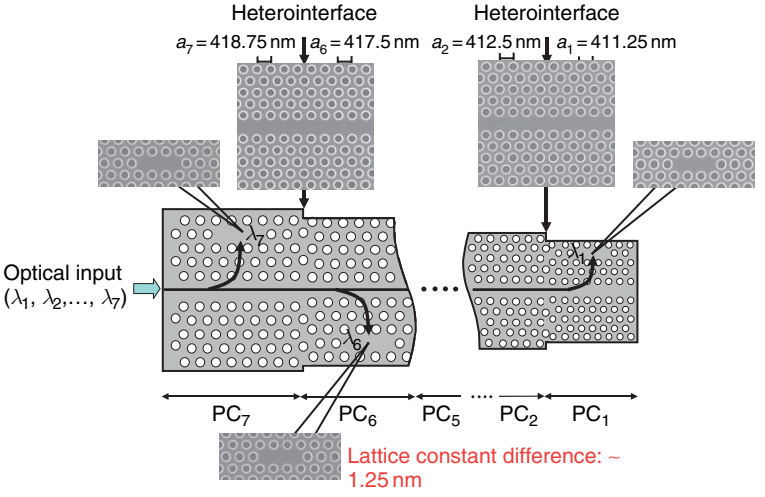


Figure 4 Example of an in-plane heterostructure: the insets show electron microscopic images of a fabricated sample. The schematic drawing highlights the differences in lattice constant, which is too small (1.25 nm) to be seen in the microscope images.

while maintaining the high values of Q (3800) and high drop efficiency (50%) for the configuration shown in **Figure 3**. As described in the previous section, the overall point-defect Q value is determined by a combination of the in-plane value Q_{in} and the vertical value Q_v, and the radiation efficiency is highest when $Q_{in} = Q_v$. If the size of only one point defect is changed, in turn changing the operational wavelength, the effective refractive index of the defect will also be altered, resulting in a change in Q_v. On the contrary, Q_{in} will be relatively unaffected: therefore, the overall Q value will be altered and the drop efficiency will change together with the wavelength. The difference in the overall Q value will also cause a change in the wavelength resolution. When a heterostructure such as that in **Figure 4** is used to vary the structure of the overall crystal proportionally, although the operating wavelength will change with the lattice constant, the Q value and drop efficiency, which are dimensionless quantities, will be maintained at constant values. In other words, multiwavelength operations will be possible as long as the wavelength resolution and drop efficiency are kept constant. **Figure 5** shows the results of such an experiment. It shows the light emitted from the point defect in each region as the irradiated wavelength is varied. **Figure 6** shows the corresponding drop spectra. The spectrum for λ_7 has been omitted due to light scattering effects at the waveguide edge. The relationship between the transmission and drop spectra is also shown. It can be seen that the drop wavelength interval was almost constant at approximately 5 nm, as per the design, and the Q value was maintained at 3800 in all wavelength regions. It is also demonstrated that the drop efficiency was essentially constant for λ_1 to λ_4.

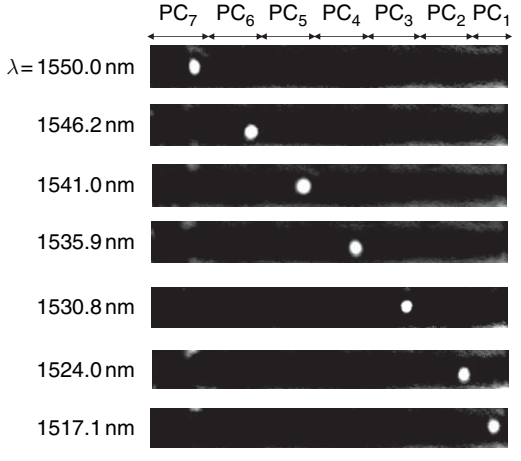


Figure 5 Multichannel drop operation of a device with an in-plane heterostructure.

Detailed calculations showed that in these wavelength areas a high drop efficiency of approximately 40% was obtained. In contrast, although the Q values for λ_5 and λ_6 were similar to those for λ_1 to λ_4, it is clear that far greater drop efficiencies than the theoretical maximum of 50% were obtained. The reason for this unexpected observation is concealed in the relationship between the transmission spectrum and the drop spectrum in the inset to **Figure 6**. It can be seen that λ_5 and λ_6 are outside the transmission area, that is, the cutoff area. **Figure 7(a)** illustrates this mechanism, where a heterostructure composed of two photonic crystals, PC_m and PC_n, that have different lattice constants is shown: a part of the light introduced into the waveguide in PC_m is transmitted (T) without being trapped by the point defect. A different part of the light is trapped by the point defect, but it is reflected (R) to the entrance side of the crystal without being emitted in the upward direction. The rest of the light is emitted upward from the point defect; that is, it drops (D). As described previously, based on coupled mode theory, it can be deduced that light D will have a maximum radiation efficiency of 50% and that $T = R = 25\%$. When there is a heterointerface and the emitted light cannot enter PC_n, or in other words, a cutoff condition is reached, the transmitted light T will be reflected by the interface. When the reflected light T interferes with R in an

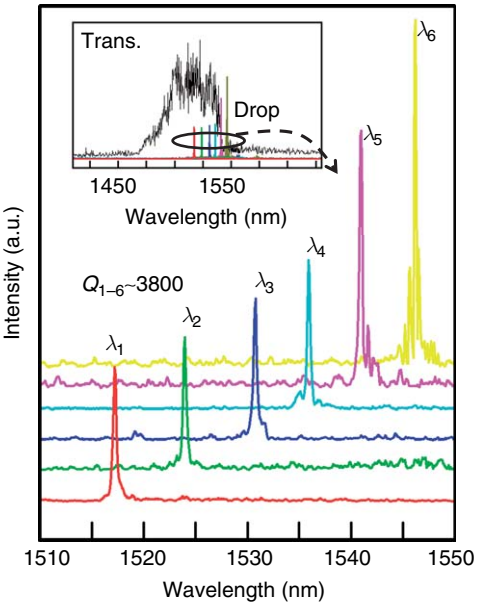


Figure 6 Drop spectra of a device with an in-plane heterostructure: the inset shows the relationship between the transmission spectrum and the drop spectrum. (Source: Song B S, Noda S, Asano T 2003 Photonic devices based on in-plane hetero photonic crystals. *Science* **300**, 1537; **Figure 1**).

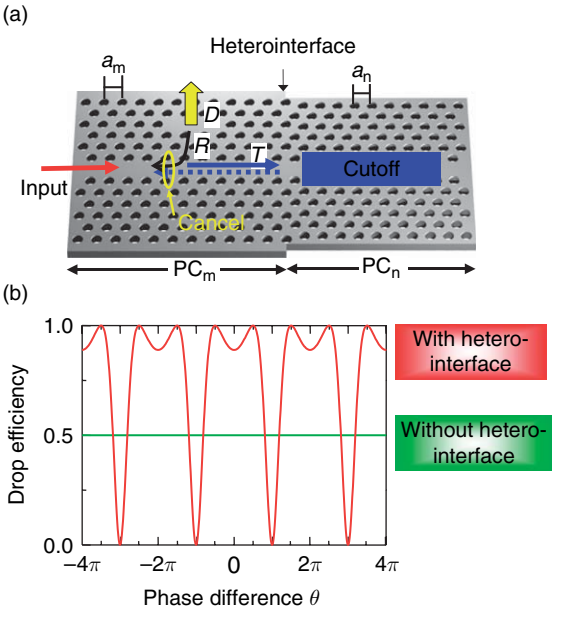


Figure 7 Schematic drawing showing (a) the effects of reflection at a heterointerface (*T*, *R*, and *D* represent the transmitted light (transmission rate), reflected light (reflection rate), and drop light (drop efficiency), respectively) and (b) the calculated properties of these effects.

out-of-phase fashion, T and R will cancel each other out and, as a result, the efficiency of light D becomes 100%. **Figure 7(b)** shows the calculated radiation efficiency when reflection at the interface is incorporated. The phase difference (θ) between light T reflected by the heterointerface and light R reflected by the defect is shown on the horizontal axis, and a high D efficiency can be obtained due to the broad phase conditions. If θ is carefully selected, it is possible to obtain a light radiation efficiency of 100%. The experimental results for λ_5 and λ_6 in **Figure 6** clearly show this effect.

As described above, important effects can be expected from the use of heterostructures, such as dramatically improving the degrees of freedom available in the design of devices and improving the operation efficiency. As described later, reflection at a heterointerface also plays an important role in the realization of ultrahigh-Q nanocavities. Furthermore, chirped photonic crystal waveguides could be made feasible through heterostructure deformations, and it might also be possible to use such heterostructures in a broad sense for controlling light propagation (Mori and Baba 2004). In the future, it is expected that heterostructures will continue to gain in importance as a technology for the realization of photonic crystal devices.

3.03.2.3 High-Q Nanocavities

The devices shown in **Figures 3** and **4** utilize photons that have been emitted to free space from a point-defect cavity. However, if the Q factor of the point defect could be increased sufficiently, the emission of photons outside the slab plane would be suppressed, even if photons were trapped by the point defect. An in-plane light circuit can then be composed (Takano *et al.* 2004, 2006).

It is thus expected that the development of high-Q nanocavities with very small modal volumes (V) will advance to enable the development of in-plane photonic circuits and to play an extremely important role in a broad range of other applications. Examples include the creation of strong coupling system between photon and electron systems for quantum computing, single-photon source for quantum communication, future photonic memory, and supersensitive sensing. The applications envisaged will be based on the extreme strengthening of the interaction between light and matter by strongly trapping light in a very small area over an extended timescale. In the following, discussions on recent advances in high-Q nanocavities are described.

The Q factor of the point-defect cavity shown in **Figure 3** was 3800, but recently, as shown in **Figure 8**, a phenomenon has been discovered whereby Q can be increased by at least a factor of 10 by simply tuning the lattice points (air holes) at both edges of the point-defect cavity (Akahane *et al.* 2003b). Before this phenomenon was found, it was revealed that a major cause

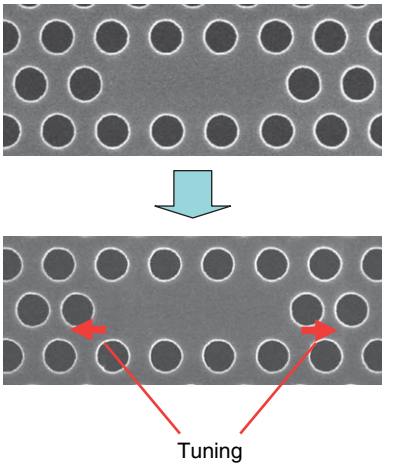


Figure 8 Method of enlarging the Q value of a donor-type defect in which three lattice points are filled (**Figure 3**). (Source: Akahane Y, Asano T, Song B S, Noda S 2003b High-Q photonic nanocavity in a two-dimensional photonic crystal. *Nature* **425**, 944–7; **Figure 1**).

of the leakage was light within the cavity being emitted in the direction perpendicular to the slab, due to abrupt reflection at the ends of the cavity along the long axis. In order to prevent such leakage, it is important that steep changes in the electric field gradient at the edge of the defect be avoided: the electric field gradient should have a gradual (ideally Gaussian) distribution (Akahane *et al.* 2003b). The tuning of the lattice points, as shown in **Figure 8**, corresponds to a change in the Bragg reflection condition at the cavity edge. The Bragg reflection is determined by a summation of partial reflections at a series of lattice points near the cavity. When lattice points near the cavity edges are tuned, the Bragg reflection condition should be modified. Because the phases of partial reflections at the tuned rods are changed, the resultant phase mismatch weakens the magnitude of Bragg reflection. To compensate for the reduction in the reflection, light is considered to penetrate more inside the Bragg reflection mirror and be reflected perfectly. It means that the electric field profile at the cavity edge becomes gentler. Based on the above ideas, nanocavities with various lattice point shifts were created, as shown in **Figure 9**, and the corresponding Q factors were measured. The highest value of Q (45 000) was obtained when the lattice point was shifted by approximately $0.15a$ (60 nm), where a is lattice constant. When the modal volume was standardized, this value corresponded to a record light-confinement effect.

Thereafter, tuning was performed not only for lattice points directly adjacent to the nanocavity but also for those that were two and three points away. A Q value of 100 000 was attained as a result (Akahane *et al.* 2005). Furthermore, very recently it was discovered that a photonic double heterostructure (**Figure 10**) that utilizes reflection based on mode gap effects at in-plane heterointerfaces, discussed previously, is able to achieve even higher Q values of ~1 000 000 (Asano *et al.* 2006, Song *et al.* 2005).

It can be concluded from the above results that, in accordance with the guidelines derived from Gaussian confinement, the use of even simple structures such as a 2D slab, together with lattice point tuning and the concept of a photonic double heterostructure, means that the realization of a cavity that does not leak is no longer just a dream.

3.03.2.4 Introduction of Nonlinear and/or Active Functionality

As described earlier, various advances have been made in 2D photonic crystal slab structures. All of these have occurred based on the use of material with a constant refractive index that forms the photonic crystals, the

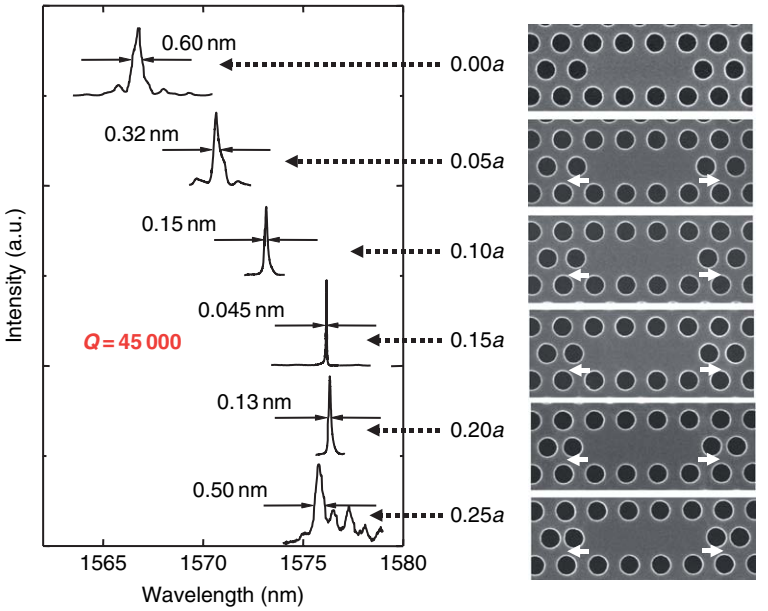


Figure 9 Electron microscopic images of a nanocavity with various lattice point shifts and their corresponding resonance spectra. (Source: Akahane Y, Asano T, Song B S, Noda S 2003b High-Q photonic nanocavity in a two-dimensional photonic crystal. *Nature* **425**, 944–7; **Figure 4**).

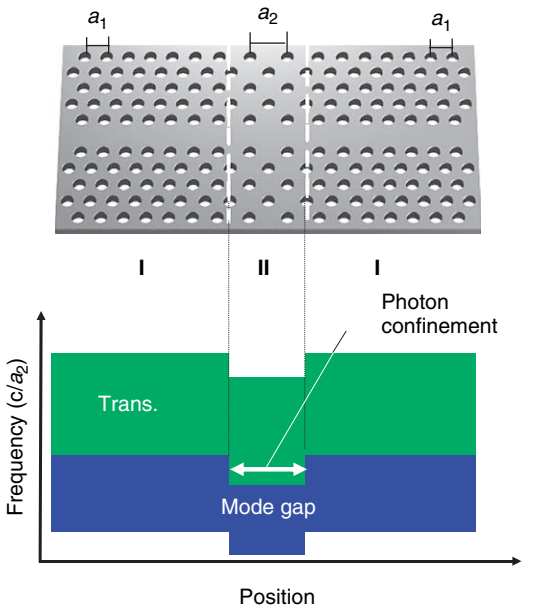


Figure 10 Photonic double heterostructure and its photon-confinement mechanism (c = light velocity; a_1, a_2 = lattice constants of individual photonic crystal units). (Source: Song B S, Noda S, Asano T 2005 Ultra-high-Q photonic double-heterostructure nanocavity. *Nat. Mater.* **4**, 207–10; **Figure 1**).

material used being transparent. As ever, more advanced functionality is demanded, dynamic control of the photonic crystals becomes important. Great effort is currently being undertaken to realize such control. For example, in **Figure 3** the resonance (drop) wavelength is fixed, but if the refractive index of the defect is dynamically varied, it becomes possible to tune the resonance wavelength. Changes in the refractive index can be induced by thermal effects and/or carrier plasma effects (Asano *et al.* 2005, Baba *et al.* 2003). **Figure 11** shows an example of the outcome of such tuning. By introducing light into the point defect and thermally changing the refractive index of the defect, tuning of the resonance wavelength by up to 5 nm is achieved. This device is extremely compact, so that even the use of a thermal effect allows a dynamic speed to be obtained in a matter of microseconds. Implementation of a high-speed, nonlinear effect has recently been tried. There have also been recent reports of the fabrication of a bistable switch that is operated at a low-input intensity (Notomi *et al.* 2005).

Advances have also been made in the realization of extremely small active devices by the introduction of active media into a 2D photonic crystal slab. Two-dimensional photonic crystal point-defect lasers were first reported in 1999 (Painter *et al.* 1999). Thereafter, various improvements to the structure of the resonator were made, improving the performance. In 2004, a current injection operation into a single point-defect device was carried out (Park *et al.* 2004). Research into line-defect lasers was continuing at the same time, and, in 2002, the first oscillations from a line-defect laser were reported (Sugitatsu and Noda 2003). This showed that substantial wavelength tuning was possible by varying the width of the waveguide (Sugitatsu *et al.* 2004). Very recently, an attempt was made to introduce quantum dots (Arakawa and Sakaki 1982, Nabetani *et al.* 1994, Tabuchi *et al.* 1992) into the high-Q nanocavity shown in **Figure 8** (Yoshie *et al.* 2004). Further research and development related to single-photon sources and optical quantum information devices is expected using methods that introduce an active medium.

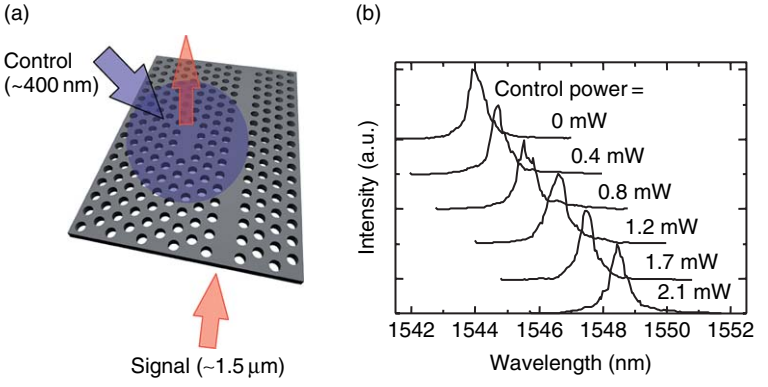


Figure 11 Tuning of resonance spectra: (a) schematic representation of light incident on a point defect and (b) changes in resonance spectra for various incident control light intensities. (Source: Asano T, Kunishi W, Nakamura M, Song B S, Noda S 2005 Dynamic wavelength tuning of channel-drop device in two-dimensional photonic crystal slab. *Electron. Lett.* **41**, 37–8; **Figure 2**).

3.03.3 Summary and Future Prospects

Recent advances in photonic crystals have been discussed in terms of band–gap/defect engineering. Although many of the explanations given may have been insufficient due to space limitations, it is hoped that the reader gets a sense of the steady advances made in the last few years in fundamental technology for operating and controlling light in extremely small volumes, which was the initial expectation of photonic crystals.

It is expected that, over the next 10 years, nanoprocessing technology will dramatically advance and that more reliable and precise devices will continue to be developed. In the case of 2D photonic crystal slabs, there is promise of remarkable advances in Si-based systems, together with progress in integration with electronic circuits. Further advances can be expected in combined optical and electronic circuits equipped with features such as optical switching, tuning, and delay functionality. It is expected that the main components of such circuits will be optical, and optical/electronic chips will be developed. There is no doubt that the size and power consumption of such devices will be more than tens to hundreds of times smaller than they are now. Advances in a large number of applications can also be expected, including developments in next-generation miniaturized multiwavelength light sources based on the addition of active functionality, supersensitive sensors, optical memory functions that require a high Q value, and single-photon light sources.

Acknowledgments

This work was supported by the Core Research for Evolutional Science and Technology (CREST) program of the Japan Science and Technology Agency, by a Grant-in-Aid for Scientific Research and an IT project grant of the Ministry of Education, Culture, Sports, Science, and Technology of Japan, by the 21 COE program, and by the Venture Business Laboratory of Kyoto University.

References

Akahane Y, Asano T, Song B S, Noda S 2003a Investigation of high-*Q* channel drop filters using donor-type defects in two-dimensional photonic crystal slabs. *Appl. Phys. Lett.* **83**, 1512–14

Akahane Y, Asano T, Song B S, Noda S 2003b High-*Q* photonic nanocavity in a two-dimensional photonic crystal. *Nature* **425**, 944–7

Akahane Y, Asano T, Song B S, Noda S 2005 Fine-tuned high-*Q* photonic-crystal nanocavity. *Opt. Express* **13**, 1202–14

Aoki K, Miyazaki H T, Hirayama H, Inoshita K, Baba T, Sakoda K, Shinya N, Aoyagi Y 2003 Microassembly of semiconductor three-dimensional photonic crystals. *Nat. Mater.* **2**, 117–21

Arakawa Y, Sakaki H 1982 Multidimensional quantum well laser and temperature dependence of its threshold current. *Appl. Phys. Lett.* **40**, 939–41

Asano T, Kunishi W, Nakamura M, Song B S, Noda S 2005 Dynamic wavelength tuning of channel-drop device in two-dimensional photonic crystal slab. *Electron. Lett.* **41**, 37–8

Asano T, Song B S, Noda S 2006 Analysis of the experimental *Q* factors ($\sim$1 million) of photonic crystal nanocavities. *Opt. Express* **14**, 1996–2002

Baba T, Fukaya N, Yonekura J 1999 Observation of light propagation in photonic crystal optical waveguides with bends. *Electron. Lett.* **35**, 654–5

Baba T, Shiga M, Inoshita K, Koyama F 2003 Carrier plasma shift in GaInAsP photonic crystal point defect cavity. *Electron. Lett.* **39**, 1516–18

Benisty H, Lourtioz J-M, Chelnokov A, Combrié S, Checoury X 2006 Recent advances toward optical devices in semiconductor-based photonic crystals. *Proc. IEEE* **94(5)**, 997–1023

Block I D, Chan L L, Cunningham B T 2007 Large-area submicron replica molding of porous low-k dielectric filems and application to photonic crystal biosensor fabrication. *Microelectron. Eng.* **84(4)**, 603–8

Borel P I, Harpoth A, Frandsen L H, Kristensen M, Shi P, Jensen J S, Sigmund O 2004 Topology optimization and fabrication of photonic crystal structures. *Opt. Express* **12**, 1996–2001

Chutinan A, Noda S 2000 Waveguides and waveguide bends in two-dimensional photonic crystal slabs. *Phys. Rev. B* **62**, 4488–92

Chutinan A, Mochizuki M, Imada M, Noda S 2001 Surface-emitting channel drop filters using single defects in two-dimensional photonic crystal slabs. *Appl. Phys. Lett.* **79**, 2690–2

Chutinan A, Okano M, Noda S 2002 Wider bandwidth with high transmission through waveguide bends in two-dimensional photonic crystal slabs. *Appl. Phys. Lett.* **80**, 1698–700

Erickson D, Rockwood T, Emery T, Scherer A, Psaltis D 2006 Nanofluidic tuning of photonic crystal circuits. *Optics Lett.* **31(1)**, 59–61

François M, Danglot J, Grimbert B, Mounaix P, Muller M, Vanbésien O, Lippens D 2002 Photonic bandgap material for integrated photonic application: technological challenges. *Microelectron. Eng.* **61–62**, 537–44

Gomyo A, Ushida J, Shirane M, Tokushima M, Yamada H 2004 Low optical loss connection for photonic crystal slab waveguides. *IEICE Trans. Electron.* **E87C**, 328–35

Ho K M, Chan C T, Soukoulis C M 1990 Existence of a photonic gap in periodic dielectric structures. *Phys. Rev. Lett.* **65**, 3152–5

Joannopoulos J D, Villeneuve P R, Fan S 1997 Photonic crystals: putting a new twist on light. *Nature* **386**, 143–9

Kawakami S, Sato T, Miura K, Ohtera Y, Kawashima T, Ohkubo H 2003 3-D photonic-crystal heterostructures: fabrication and in-line resonator. *IEEE Photon. Technol. Lett.* **15**, 816–18

Kee C-S, Park I, Lim H 2006 Photonic crystal multi-channel drop filters based on microstrip lines. *J. Phys. D Appl. Phys.* **39**, 2932–4

Kim S H, Lee K-D, Kim J-Y, Kwon M-K, Park S-J 2007 Fabrication of photonic crystal structures on light emitting diodes by nanoimprint lithography. *Nanotechnol.* **18**, 055306

Kuramochi E, Notomi M, Hughes S, Shinya A, Watanabe T, Ramunno T 2005 Disorder-induced scattering loss of line-defect waveguides in photonic crystal slabs. *Phys. Rev. B* **72**, Art. No 161318

Lee E H, Lee S G, B H O, Park S G, Chung M Y, Kim K H, Song S H 2006 Fabrication and integration of VLSI micro/nano-photonic circuit board. *Microelectron. Eng.* **83(4–9)**, 1767–72

Leisher P O, Danner A J, Choquette K D 2006 Single-mode 1.3-μm photonic crystal vertical-cavity surface-emitting laser. *IEEE Photon. Technol. Lett.* **18(20)**, 2156–8

McNab S J, Moll N, Vlasov Y A 2003 Ultra-low loss photonic integrated circuit with membrane-type photonic crystal waveguides. *Opt. Express* **11**, 2927–39

Mekis A, Chen J C, Kurland I, Fan S H, Villeneuve P R, Joannopoulos J D 1996 High transmission through sharp bends in photonic crystal waveguides. *Phys. Rev. Lett.* **77**, 3787–90

Miyai E, Noda S 2004 Structured dependence of coupling between a two-dimensional photonic crystal waveguide and a wire waveguide. *J. Opt. Soc. Am. B* **21**, 67–72

Miyai E, Okano M, Mochizuki M, Noda S 2002 Analysis of coupling between two-dimensional photonic crystal waveguide and external waveguide. *Appl. Phys. Lett.* **81**, 3729–31

Mori D, Baba T 2004 Dispersion-controlled optical group delay device by chirped photonic crystal waveguides. *Appl. Phys. Lett.* **85**, 1101–3

Nabetani Y, Ishikawa T, Noda S, Sasaki A 1994 Initial growth stage and optical properties of a three-dimensional InAs structure on GaAs. *J. Appl. Phys.* **76**, 347–51

Noda S and Baba T (eds.) (2003) *Roadmap on Photonic Crystals,* Kluwer Academic Publishers, Dordrecht, the Netherlands

Noda S, Yamamoto N, Sasaki A 1996 New realization method for three-dimensional photonic crystal in optical wavelength region. *Jpn. J. Appl. Phys.* **35**, L909–12

Noda S, Chutinan A, Imada M 2000a Trapping and emission of photons by a single defect in a photonic bandgap structure. *Nature* **407**, 608–10

Noda S, Tomoda K, Yamamoto N, Chutinan A 2000b Full three-dimensional photonic bandgap crystals at near-infrared wavelengths. *Science* **289**, 604–6

Notomi M, Shinya A, Yamada K, Takahashi J, Takahashi C, Yokohama I 2001 Singlemode transmission within photonic bandgap of width-varied single-line-defect photonic crystal waveguides on SOI substrates. *Electron. Lett.* **37**, 293–5

Notomi M, Shinya A, Mitsugi S, Kuramochi E, Ryu H Y 2004 Waveguides, resonators and their coupled elements in photonic crystal slabs. *Opt. Express* **12**, 1551–61

Notomi M, Shinya A, Mitsugi S, Kira G, Kuramochi E, Tanabe T 2005 Optical bistable switching action of Si high-*Q* photonic-crystal nanocavities. *Opt. Express* **13**, 2678–87

Painter O, Lee R K, Scherer A, Yariv A, O'Brien J D, Dapkus P D, Kim I 1999 Two-dimensional photonic band-gap defect mode laser. *Science* **284**, 1819–21

Park H G, Kim S H, Kwon S H, Ju Y G, Yang J K, Baek J H, Kim S B, Lee Y H 2004 Low loss mode size converter from 0.3/ spl mu/m square Si wire waveguides to single mode fibers. *Science* **305**, 1444–7

Park W, Lee J-B 2004 Mechanically tunable photonic crystal structure. *Appl. Phys. Lett.* **85(21)**, 4845–7

Pisignano D, Persano L, Gigli G, Visconti P, Stomeo T, De Vittorio M, Barbarella G, Favaretto L, Cingolani R 2004 Planar organic photonic crystals fabricated by soft lithography. *Nanotechnology* **15**, 766–70

Prather D W, Shi S, Murakowski J, Schneider G J, Sharkawy A, Chen C, Miao B 2006 Photonic crystal structures and applications: perspective, overview and development. *IEEE J. Sel. Top. Quantum Electron.* **12(6)**, 1416–37

Rajic S, Corbeil J L, Datskos P G 2003 Feasibility of tunable MEMS photonic crystal devices. *Ultramicroscopy* **97**, 473–9

Shoji T, Tsuchizawa T, Watanabe T, Yamada K, Morita H 2002 Low loss mode size converter from 0.3 /spl mu/m square Si wire waveguides to singlemode fibres. *Electron. Lett.* **38**, 1669–70

Song B S, Noda S, Asano T 2003 Photonic devices based on in-plane hetero photonic crystals. *Science* **300**, 1537

Song B S, Noda S, Asano T 2005 Ultra-high-*Q* photonic double-heterostructure nanocavity. *Nat. Mater.* **4**, 207–10

Stomeo T, Errico V, Salhi A, Passaseo A, Cingolani R, D'Orazio A, De Sario M, Marrocco V, Petruzzelli V, Prudenzano F, De Vittorio M 2006 Design and fabrication of active and passive photonic crystal resonators. *Microelectron. Eng.* **83(4–9)**, 1823–5

Sugimoto Y, Tanaka Y, Ikeda N, Kanamoto K, Nakamura Y, Ohkouchi S, Nakamura H, Inoue K, Sakai H, Watanabe Y, Ishida K, Ishikawa H, Asakawa K 2004a Two-dimensional semiconductor-based photonic crystal slab waveguides for ultra-fast optical signal processing devices. *IEICE Trans. Electron.* **87C**, 316–27

Sugimoto Y, Tanaka Y, Ikeda N, Nakamura Y, Asakawa K, Inoue K 2004b Low propagation loss of 0.76 dB/mm in GaAs-based single-line-defect two-dimensional photonic crystal slab waveguides up to 1 cm in length. *Opt. Express* **12**, 1090–6

Sugitatsu A, Noda S 2003 Room temperature operation of a two dimensional photonic crystal slab defect-waveguide-laser with optical pump. *Electron. Lett.* **39**, 123–5

Sugitatsu A, Asano T, Noda S 2004 Characterization of line-defect-waveguide lasers in two-dimensional photonic-crystal slabs. *Appl. Phys. Lett.* **84**, 5395–7

Tabuchi M, Noda S, Sasaki A 1992 Science and Technology of Mesoscopic Structures edited by Namba S, Hanaguchi C, and Ando T, Springer, Tokyo, pp. 379–84

Takano H, Song B S, Asano T, Noda S 2006 Highly efficient multi-channel drop filter in a two-dimensional hetero photonic crystal. *Opt. Express* **14**, 3492–6

Takayama S, Kitagawa H, Asano T, Tanaka Y, Noda S 2005 Experimental demonstration of complete photonic band gap in two-dimensional photonic crystal slabs. *Appl. Phys. Lett.* **87**, Art. No 061107

Tanaka Y, Asano T, Akahane Y, Song B S, Noda S 2003 Theoretical investigation of a two-dimensional photonic crystal slab with truncated cone air holes. *Appl. Phys. Lett.* **82**, 1661–3

Viasnoff-Schwoob E, Weisbuch C, Benisty H, Cuisin C, Derouin E, Drisse O, Duan G-H, Legouézigou L, Legouézigou O, Pommereau F, Golka S, Heidrich H, Hensel H J, Janiak K 2005 Compact wavelength monitoring by lateral outcoupling in wedged photonic crystal multimode waveguides. *Appl. Phys. Lett.* **85**, 101107

Yablonovitch E 2000 How to be truly photonic. *Science* **289**, 557–9

Yoshie T, Scherer A, Hendrickson J, Khitrova G, Gibbs H M, Rupper G, Ell C, Shchekin O B, Deppe D G 2004 Vacuum Rabi splitting with a single quantum dot in a photonic crystal nanocavity. *Nature* **432**, 200–203

Zhen Y R, Li L-M 2005 A novel application of two-dimensional photonic crystals: polarization beam splitter. *J. Phys. D Appl. Phys.* **38**, 3391–4

Zhou W, Taysing-Lara M, Dang G, Mackie D M, Ervin M H, Newman P G 2005 Reconfigurable MEMS waveguide based on photonic crystal membrane. In: Abidi A, Lin S-Y, and Scherer A (eds.) Proc. SPIE, **5733**, 125–130

Zhou W, Mackie D M, Taysing-Lara M, Dang G, Newman P G, Svensson S 2006 Novel reconfigurable semiconductor photonic crystal-MEMS device. *Solid State Electron.* **50(6)**, 908–13

Biography

Susumu Noda received B.S., M.S., and Ph.D. degrees from Kyoto University, Kyoto, Japan, in 1982, 1984, and 1991, respectively, all in electronics. In 2006, he received an honorary doctor degree from Gent University, Gent, Belgium. From 1984 to 1988, he was with the Mitsubishi Electric Corporation, and was engaged in research on optoelectronic devices including AlGaAs/GaAs distributed feedback (DFB) lasers and multiple quantum well (MQW) DFB lasers. In 1988, he joined Kyoto University and is currently a Professor with the Department of Electronic Science and Engineering. His research interest covers physics and applications of photonic and quantum nanostructures including photonic crystals and quantum dots. He is an author of more than 200 scientific journal papers including *Nature* and *Science* on these research fields. Prof. Noda is a member of IEICE and the Japan Society of Applied Physics (JSAP). He received various awards including the IBM Science Award (2000), the JSAP Achievement Award on Quantum Electronics (2005), and Optical Society of America, Joseph Fraunhofer Award/Robert M. Burley Prize. From 2003 to 2005, he served as IEEE/LEOS Distinguished Lecturer.

3.04 IR Imaging

Masafumi Kimata, Department of Micro System Technology, College of Science and Engineering, Ritsumeikan University, Shiga, Japan

Glossary

Nomenclature *Definition*

Resonant optical cavity An infrared-absorbing structure having a metal reflector, a thin-metal film absorber, and a dielectric layer between them.

1/*f* noise The noise whose power is inversely proportional to the frequency.

Absorptivity The ratio of absorbing light energy to incident light energy.

Background fluctuation noise-limited *D** The D^* that is determined by the noise due to the fluctuation of radiation between the thermal infrared detector and the surroundings.

Background fluctuation noise-limited NETD The NETD that is determined by the noise due to the fluctuation of radiation between the thermal infrared detector and the surroundings.

Blackbody *D** The D^* specified for a blackbody light source having a specified temperature.

Curie temperature The temperature at which a phase transition due to heating occurs in a ferroelectric.

Cutoff wavelength The wavelength beyond which a photodetector has no useful response.

***D** (dee star, normalized detectivity)** A figure of merit that normalizes the detectivity in terms of the signal bandwidth and detector area.

Detectivity The inverse value of the noise equivalent power.

Dielectric bolometer A ferroelectric infrared detector that uses the temperature dependence of the dielectric constant near its Curie temperature.

Diode bolometer A thermometer or infrared detector using temperature current–voltage characteristics of a diode.

Field-enhanced pyroelectric coefficient The pyroelectric coefficient in which enhancement by the temperature dependence of the dielectric constant is taken into consideration.

Fill factor Ratio of the infrared-absorbing area to the pixel size.

Hybrid structure An IRFPA structure in which a detector chip is mated with a readout IC by bump bonding.

Infrared focal plane array (IRFPA) The infrared imaging device that has a number of photodetection pixels and readout electronic circuits.

Johnson noise A noise associated with the random motion of a carrier in a resistance.

Loss tangent A parameter that correlates the loss resistance in a ferroelectric with the impedance of its capacitive component.

Monolithic structure An IRFPA structure in which an array of pixels and readout circuits are integrated on a silicon chip.

Noise equivalent power The radiant power incident on an infrared detector that gives rise to a signal equal to the total rms noise.

Noise equivalent temperature difference (NETD) The temperature difference of a blackbody that produces a signal-to-noise ratio of one.

Pixel The unit cell of the imaging device.

Pyroelectric coefficient The gradient of the spontaneous polarization versus temperature curve.

Pyroelectric detector A ferroelectric infrared detector that uses the temperature dependence of spontaneous polarization.

Quantum infrared detector An infrared detector that detects the infrared rays by interaction of photon and electron/hole.

Resistance bolometer An infrared detector or thermometer that uses the temperature dependence of the resistance of a metal or semiconductor.

Responsivity The value that represents how sensitive a photodetector is, and is given by the output signal amplitude divided by the incident light power.

Scene temperature sensitivity The sensitivity defined as the output signal amplitude per Kelvin temperature change in the observed scene.

Seebeck coefficient The thermoelectric voltage per Kelvin temperature difference generated by the Seebeck effect.

Seebeck effect The effect in which an electromotive force is generated when two dissimilar conductors are connected and are given different temperatures at two junctions.

Spectral D^* The D^* specified for a monochromatic light source.

Shot noise A noise generated when carriers flow over some potential barrier.

SOI wafer A silicon wafer that consists of a single-crystal silicon substrate, a silicon dioxide layer, and a top single-crystal silicon layer.

Spontaneous polarization Polarization that appears in a pyroelectric material without an external applied electric field.

Temperature coefficient of resistance (TCR) The ratio of the resistance change due to 1 K temperature change to the resistance of the resistance bolometer.

Temperature fluctuation noise-limited D^* The NETD that is determined by the noise due to the fluctuation of heat conduction between the thermal infrared detector and the surroundings.

Temperature fluctuation noise-limited NETD The NETD that is determined by the noise due to the fluctuation of heat conduction between the thermal infrared detector and the surroundings.

Thermal conductance A parameter defined as the transported energy per Kelvin temperature difference.

Thermal infrared detector An infrared detector that detects the infrared rays by its change in temperature.

Thermal time constant A parameter that represents the thermal response speed.

Thermocouple A thermometer operated by the principle of the Seebeck effect.

Thermoelectric detector A thermal infrared detector that uses a thermometer based on the Seebeck effect.

Uncooled IRFPA An IFRPA that integrates thermal infrared detectors. Quantum IRFPAs operated at ambient temperature are sometimes categorized under uncooled IRFPA.

Abbreviations

Acronym	*Full spelling and explanation*
AFM	*Atomic Force microscope*. High-resolution microscope for observing material surfaces to an accuracy of atomic level.
BST	*Barium strontium titanate*. A material used for ferroelectric uncooled IRFPAs.
CCD	*Charge-coupled device*. A semiconductor device that has an array of MOS capacitors and has a function to transfer electronic charges.
CMOS	*Complementary metal oxide semiconductor*. MOS circuits or LSI having both n-channel and p-channel MOS transistors.
EDP	*Ethylenediamine pyrocatechol*. Name of an etchant for silicon anisotropic etching that contains ethylenediamine, pyrocatechol, and water.
FPA	*Focal plane array*. Imaging devices that integrate a great number of pixels and readout circuits.
FPN	*Fixed pattern noise*. A noise in FPAs that is a time-independent characteristic. The sources of this noise include pixel-by-pixel nonuniformity of responsivity and dark current, and nonuniformity in readout circuits.
IRFPA	*Infrared focal plane array*. FPAs having response in infrared spectral bands.
IVP	*Integrated vacuum package*. A configuration of vacuum package for uncooled IRFPAs that consist of an IRFPA chip and lid directly bonded to the IRFPA chip.
LSI	*Large-scale integration*. Silicon electronic devices that integrate a great number of transistors and other electronic components.
LWIR	*Long-wavelength infrared*. An infrared spectral band between 8 and 14 μm.
MOD	*Metal organic decomposition*. A thin-film deposition method that utilizes decomposition of metal organic precursors at elevated temperatures.
MOS	*Metal oxide semiconductor*. A transistor that has a source and drain diffusion region and a channel region that is formed in a metal oxide semiconductor structure.
MWIR	*Middle-wavelength infrared*. An infrared spectral band between 3 and 5 μm.
NEP	*Noise equivalent power*. A figure of merit for light detectors that gives the lowest detectable infrared radiation power for the detector and is defined as a ratio of total rms noise to responsivity.
NETD	*Noise equivalent temperature difference*. A figure of merit for infrared imaging system representing the target temperature difference that gives the output signal having the same magnitude as the rms noise.
PCT	*Lead calcium titanate*. A material used for ferroelectric uncooled IRFPAs.
PECVD	*Plasma-enhanced chemical vapor deposition*. CVD processes that utilize plasma to enhance chemical reaction rates of the precursors.
PLD	*Pulsed laser deposition*. A thin-film deposition method utilizing a technique called laser ablation.
PSG	Phosphosilicate glass.
PST	*Lead scandium tantalate*. A material used for ferroelectric uncooled IRFPAs.
PVDF	*Polyvinylidene fluoride*. A material used for ferroelectric uncooled IRFPAs.
PZT	*Lead zirconate titanate*. A material used for ferroelectric uncooled IRFPAs.
QDIP	*Quantum dot infrared photodetector*. A quantum infrared detector that has a quantum dot structure and operates based on the intraband transition.
QWIP	*Quantum well infrared photodetector*. A quantum infrared detector that consists of a multilayer of ultrathin semiconductor films and operates based on the intraband transition.
ROIC	*Readout integrated circuit*. Electronic circuits for readout operation of FPA.
SEM	*Scanning electron microscope*. A high-magnification microscope that uses a scanning electron beam.
SOI	*Silicon on insulator*. A silicon wafer having a thick single-crystal silicon substrate, a thin single-crystal silicon

	layer, and a silicon dioxide layer between them.
TCR	*Temperature coefficient of resistance*. A figure of merit for the resistance bolometer that gives a ratio of the change of resistance per Kelvin to the value of the resistance.
TFFE	*Thin-film ferroelectric*. A name expressing a monolithic ferroelectric uncooled IRFPA pixel.
TMAH	*Tetramethyl ammonium hydroxide*. An etchant for silicon anisotropic etching, the chemical formula of which is $(CH_3)_4NOH$.
Type II SLS	*Type II strained layer superlattice*. A quantum infrared detector that consists of a multilayer of ultrathin semiconductor films and operates based on the interband transition between adjacent semiconductor layers.

3.04.1 Introduction

Infrared (IR) imaging is a tool to see objects at ambient temperatures in complete darkness. It also offers temperature information. IR imaging is a passive imaging that does not require any illuminating light sources. Its usefulness has already been recognized in a number of application areas, including security, surveillance, search and rescue, fire fighting, traffic systems, law enforcement, process control, and predictive maintenance. The long-wavelength IR (LWIR; 8–14 μm) and middle-wavelength IR (MWIR; 3–5 μm) spectral bands are made available for the above-mentioned applications because of sufficient IR radiation energy and high atmospheric transmission. Both quantum and thermal IR detectors are used in IR imaging. Although the sensitivity and response speed of the quantum detectors is much higher than those of the thermal detectors, recent advances in IR focal plane arrays (IRFPAs) with thermal detectors based on microelectromechanical systems (MEMS) technology attract much attention particularly from the viewpoint of civil applications. The IRFPA with a thermal detector is called the uncooled IRFPA, because in contrast to the IRFPA with a quantum detector, it can be operated at ambient temperatures. In the thermal IR detector, the IR energy is converted to heat that changes the detector temperature, and the temperature change is measured by a thermometer in the detector. The responsivity of the thermal IR detector is inversely proportional to the thermal conductance between the detector and the substrate. Since MEMS technology makes it possible to fabricate structures with very low thermal conductance, remarkable performance improvement in uncooled IRFPAs has been accomplished by the introduction of MEMS technology. There are some books that include topics on thermal detectors and uncooled IRFPAs (Capper and Elliott 2001, Dereniak and Crowe 1984, Kruse 2001, Kruse and Skatrud 1997, Kruse *et al.* 1962).

According to Walsh and coworkers (Rai-Choudhury 2000), MEMS is the technology that enables the manufacture of products that would otherwise be impossible to produce, improves the manufacturing efficiencies by as much as an order of magnitude, improves the critical performance aspects of current products by as much as an order of magnitude, and/or improves the product quality by reducing the component size and number. Before the MEMS era, uncooled IRFPAs were manufactured in the form of a hybrid structure, in which a detector chip is mated with a Si readout integrated circuit (ROIC) using bump structures (Capper and Elliott 2001, Kruse 2001, Kruse and Skatrud 1997). The heat loss in this case was determined by the thermal conductance through the bump structure, and it was very difficult to reduce the thermal conductance because of limitation of the bump size and material. Uncooled IRFPAs based on MEMS technology have monolithic structures, in which freestanding thin-film thermometers are fabricated above or by the side of readout circuits. The freestanding structure is supported by thin and narrow legs that are made by the same technology as that for silicon large-scale integrations (LSIs), and their thermal conductance is much lower than that of the bump in the hybrid structure (Capper and Elliott 2001, Kruse 2001, Kruse and Skatrud 1997). Thus, MEMS technology has been a powerful method to achieve responsivity improvements of uncooled IRFPAs. Furthermore, the batch process by the MEMS technology promises low-cost, large-scale mass production, and it is expected to open huge markets for IR imaging systems. While the first-generation MEMS-based uncooled IRFPAs had single-level and double-level structures that were realized by the

conventional bulk and/or the surface micromachining technology, the second-generation devices having smaller pixel sizes have more complicated structures that require more sophisticated MEMS technology (Kosasayama 2004, Murphy 2002). Thus, the importance of MEMS technology is still growing for uncooled IRFPA development.

The uncooled IRFPA is one of the most successful examples of the so-called integrated MEMS device, so is the digital micromirror (DMD) device. It has a great number of tiny thermal IR detector pixels (typically from a few tens of thousands to a few hundreds of thousands) on a chip, along with readout circuits. To manufacture integrated MEMS devices, compatibility with Si-LSI processes is the most important issue to be addressed when designing the devices. The DMD is a digital device, which integrates small bistable movable mirrors and static random access memory cells to control them. On the other hand, the uncooled IRFPA has no moving components and has analog circuits that deal with minute analog signals from thermometers. Although the analog circuit technology is not a major theme of this chapter, it is important to note that the current performance of the uncooled IRFPAs would not be reached if low-noise analog circuits were not developed for them.

This chapter is devoted to uncooled IRFPA technology, including the basics and state-of-the-art technology. Before explaining individual IRFPA technologies, the classification and history of IR detectors and IRFPAs are reviewed in Section 3.04.2. Then, the basics of the thermal detectors and their fundamental performance limit are discussed along with some important formulae in Section 3.04.3. Following these introductory sections, individual uncooled IRFPA technologies, including the ferroelectric IRFPA, microbolometer IRFPA, thermoelectric IRFPA, diode IRFPA, and bimaterial and thermo-optical IRFPAs, are described in the continuing sections.

3.04.2 History and Classification

The history of IR detectors started with the discovery of IR radiation by Herschel in 1800 (Rogalski 1995). In his experiments, a thermal IR detector of a glass liquid thermometer was used. Since then, various IR detectors have been developed. The IR detectors are broadly categorized into two classes: thermal detectors and quantum detectors. **Figure 1** shows a classification of the IR detectors that are used for IRFPAs.

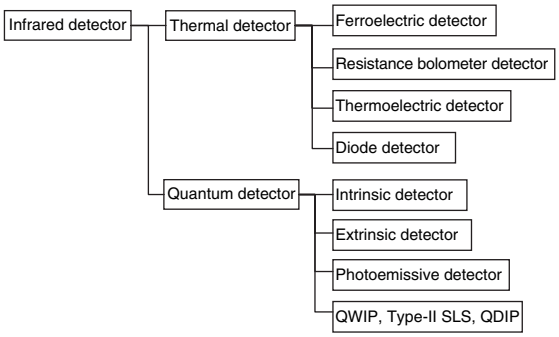


Figure 1 Classification of the infrared detectors used for infrared focal plane arrays.

The thermal detectors absorb IR radiation and change their temperature. The change in the temperature is measured by the thermometer fabricated on the detector. The thermal detectors are generally classified according to the type of thermometers, as shown in **Figure 1**. Although any physical mechanism having a temperature dependence can be applied to thermal detectors, practical thermal detectors currently used in IRFPAs are the pyroelectric detector, the dielectric bolometer, the resistance bolometer, the thermoelectric detector, and the diode bolometer.

The first thermal detector developed after Herschel's discovery was the thermoelectric detector. The thermoelectric detector utilizes the Seebeck effect that appears when two dissimilar metals and/or semiconductors are connected (Dereniak and Crowe 1984). The thermometer based on the thermoelectric effect is called the thermocouple and is used not only for IR detectors but also for contact-type thermometers. For IR detectors, serial-connected thermocouples called thermopiles are often used to enhance responsivity.

In 1881, Langley made the first resistance bolometer (Kruse and Skatrud 1997). His IR detector makes use of the temperature dependence of a Pt stripe resistance. Metals have positive temperature coefficients of resistances (TCRs), i.e., they increase their resistance as the temperature rises. The positive TCRs of metals are caused by the increased probability of scattering at higher temperatures. Both metals and semiconductors are used for resistance bolometers. In contrast to metals, semiconductors have negative TCRs. The characteristics of semiconductors reflect temperature dependences of the number of carriers and/or the mobility. Various sintered mixtures of oxides were developed for

semiconductor resistance bolometers. Typically, the sensitivity of the semiconductor bolometers is an order of magnitude larger than that of the metal bolometers.

Pyroelectric detectors and dielectric bolometers are made of ferroelectric materials. Pyroelectricity has long been known, but there were genuine efforts to understand this phenomenon only in the 19th century (Kruse and Skatrud 1997). The operation mechanism of the pyroelectric IR detector is based on the temperature dependence of the spontaneous polarization, and it is widely used as single-element IR sensors. The change in spontaneous polarization can be detected by measuring a transient current flowing through a capacitor having a pyroelectric material as an insulator. Hanel proposed the usage of the temperature-sensitive dielectric constant of ferroelectric materials in bolometers (Kruse and Skatrud 1997). Ferroelectric materials show large temperature dependence of the dielectric constant near the Curie temperature. We can apply this effect to thermometers for IR detectors. This class of ferroelectric IR detectors is called the dielectric bolometer.

The current–voltage characteristic of diodes is also temperature dependent and is used for contact-type thermometers and IR detectors. Application of the diode bolometer to the IR detector attracted attention with the progress in the development of 2D IRFPAs, because its most competitive feature when compared with other technologies is the compatibility with Si-LSI technology (Ishikawa *et al.* 1999).

Although thermal detectors have attractive features such as ambient-temperature operation and low-cost mass production, research and development of quantum detectors became the most important theme in IR technology from the mid-20th century because the first-generation IR imaging systems, which were mechanical scanning systems, required higher sensitivity and faster response for IR detectors. In the quantum detectors, the absorbed photons excite electrons in materials, changing their energy distribution. The change in the energy distribution is detected as a change in the resistance or in the current flow through an energy barrier. The operation mechanism of the photon detectors ensures a high-sensitivity and high-frequency response at the expense of the cooling requirement. While the thermal detectors for the LWIR spectral band can operate at ambient temperatures, the quantum detectors having spectral sensitivity in the MWIR and LWIR spectral bands generally have to be cooled down to cryogenic temperatures in order to reduce thermal excitation of the carriers.

Materials investigated in the early stage of quantum detector development were semiconductors of the lead salt family (PbS, PbSe, and PbTe) (Rogalski 1995). These materials were mainly used for single-element intrinsic photoconductive IR detectors. Both photoconductive and photovoltaic detectors are available with the intrinsic photodetection mechanism. The next important narrow gap semiconductor was single-crystal InSb (Capper and Elliott 2001). Extensive research on InSb during the 1950s and 1960s improved the crystal quality and purity, and thus contributed to IR detector technology. The bandgap energy of InSb is 0.23 eV at 77 K, which corresponds to a cutoff wavelength of 5.5 μm. The most important material in intrinsic IR detectors is HgCdTe, which was introduced by Lawson *et al.* in 1959 (Lawson *et al.* 1959). The bandgap energy of this alloy is adjustable, and as a result wavelength customization is possible in a very wide wavelength range between about 2 and 14 μm. In the late 1970s, and through the 1980s, HgCdTe technology was focused on developing the photovoltaic device because large-format IRFPAs needed IR detectors with low power dissipation and high impedance to interface to readout circuits. The photovoltaic HgCdTe detectors are widely used in the second-generation IR imaging systems, in which linear or 2D staring IRFPAs are incorporated. In contrast to mechanical scanning systems, every pixel in the staring IRFPA continuously stares the scene allocated for the pixel.

Si- and Ge-based extrinsic photoconductive detectors are being developed to extend the detection wavelength into the much longer wavelength in the IR region. Stressed Ge:Ga extrinsic photoconductors exhibit useful responses up to a wavelength of 200 μm (Rogalski 1995). Although extrinsic Si detectors are attractive for large-scale integration because of the compatibility with Si-LSI technology, low operation temperature restricted their applications to limited areas such as astronomy. In the early 1960s, the first real-time thermal imaging system operating in the LWIR spectral region was developed with a Hg-doped Ge extrinsic linear detector array (Rogalski 1995).

Noble proposed a metal oxide semiconductor (MOS) X–Y imaging array in 1968 (Noble 1968), and the charge-coupled device (CCD) was invented by Bole and Smith in 1970 (Boyle and Smith 1970). Such inventions triggered development of 2D staring

FPAs in IR spectral bands as well as in the visible range. Emergence of 2D staring IRFPAs changed the circumstances for IR detector development. Staring operation alleviates requirements for the sensitivity and response speed because of the long integration time in the 2D IRFPAs. In place of high sensitivity and fast response, capability of large-scale integration and high responsivity uniformity became important to realize high-performance IR imaging systems. It is important to note that nonuniformity of 2D IRFPAs sometimes limits the noise equivalent temperature difference (NETD) because IR imaging is high background imaging. This change in the circumstances sheds light on some photon detectors with low quantum efficiency, such as Si-based photoemissive detectors and GaAs-based quantum well IR photodetectors (QWIPs) (Rogalski 1995). It also revived the thermal detectors with relatively low responsivity and slow response speed.

Photoemissive detectors with Schottky barrier and heterojunction detect IR rays by using their potential barriers to measure the change in the energy distribution of carriers. Despite their relatively low quantum efficiency, PtSi Schottky-barrier IRFPAs were in the leading position in the array size race of IRFPAs for more than 10 years from the early 1980s due to their monolithic structure fabricated with a Si-LSI-compatible process and due to exceptionally high uniformity of responsivity. Nonuniformity of <1% is routinely obtained with PtSi Schottky-barrier IRFPAs. The first full-TV resolution IRFPA (Kimata *et al.* 1987) and the first megapixel IRFPA (Yutani *et al.* 1991) were developed with the PtSi Schottky-barrier technology in 1987 and 1991, respectively.

Quantum well or superlattice photodetectors are the newcomers in the quantum IRFPA field. This type of photodetector is divided into three groups:

(1) The QWIP (Rogalski 1995).
(2) The type II strained layer superlattice (SLS) (Rogalski 1995).
(3) The quantum dot IR photodetector (QDIP) (Lim *et al.* 2006).

The photodetection mechanism of QWIP is based on intersubband transition. Thanks to the mature material and process technology, GaAs-based QWIPs made significant progress in the 1990s. Compared to intrinsic detectors, QWIPs suffer from lower quantum efficiency and lower operation temperature. They, however, have found some application areas in which their advantages of large-scale and multiband integration are significant. The type II SLS uses the interband transition between the states in alternative layers. Since the type II SLS and QDIP IRFPAs offer higher performance and higher operating temperature when compared with QWIPs, research activities in this area are shifting from QWIP to them.

As explained in Section 3.04.1, the IRFPAs integrating thermal detectors are called uncooled IRFPAs. In contrast, the IRFPAs with quantum detectors are referred to as cooled IRFPAs. The term cooled IRFPA was not needed until the beginning of the 1990s because research activities and achievements in uncooled IRFPAs were limited and the term IRFPA generally meant the cooled IRFPA. In 1992, two outstanding results on the uncooled IRFPAs were disclosed. One was a hybrid IRFPA with barium strontium titanate (BST) ferroelectric detectors (Hanson *et al.* 1992) and the other was a monolithic IRFPA with VO_x resistance bolometer detectors (Wood *et al.* 1992). After these successful demonstrations, uncooled IRFPAs have made an amazing advance, and eventually, we have had to use the term, which discriminates IRFPAs with quantum detectors from those with thermal detectors.

The hybrid ferroelectric technology played an important role in the 1990s by leading commercialization of the uncooled IRFPAs. A lot of IR imaging systems, including enhanced vision systems for cars, were developed using hybrid ferroelectric IRFPAs. On the other hand, performance improvement has been proceeding in the resistance bolometer and diode bolometer, making best use of their monolithic structures manufactured with Si-LSI-like or Si-LSI-compatible processes. For resistance bolometer and diode bolometer IRFPAs, lower thermal conductance realized with advanced MEMS processes for these uncooled IRFPAs offered competitive sensitivities with the second-generation cooled IRFPAs, and monolithic structures enabled large-scale integration up to 640×480 elements (Murphy *et al.* 2005, Ueno *et al.* 2005).

Features of different kinds of IR detectors are summarized in **Table 1**.

3.04.3 Thermal Infrared Detector

This section outlines the operation and performance of thermal IR detectors and discusses the theoretical limit of thermal IR detectors and uncooled IRFPAs.

Table 1 Features of infrared detectors

Type		*Operation principle*	*Operation temperature*	*Sensitivity*	*Spectral response*	*Response time*	*FPA structure*	*Typical materials*
Thermal	Ferroelectric	Temperature dependence of spontaneous polarization (pyroelectric) Temperature dependence of dielectric constant (dielectric bolometer)	Room temperature	Low	Depending on absorber	Slow	Hybrid	BST, PST, PZT, PVDF
	Resistance bolometer	Temperature dependence of resistance						VO_x, a-Si
	Thermoelectric	Seebeck effect					Monolithic	Al/poly-Si n poly-Si/p poly-Si
	Diode	Temperature dependence of I–V characteristic						Si
Quantum	Intrinsic	Interband transition	Cryogenic temperature	High	Determined by bandgap energy	Fast		HgCdTe, InSb
	Extrinsic	Excitation from impurity level			Determined by impurity level		Hybrid	Si:In, Si:Ga, Ge:Cu, Ge:Hg
	Photoemissive	Emission over Schottky-barrier or hetero junction			Determined by barrier height			PtSi/p-Si, GeSi/p-Si
	QWIP type II SLS QDIP	Intraband or interband transition in superlattice structure		Medium	Tunable by design		Monolithic	GaAs- or InP-based compound semiconductors

Source: Kimata *et al*. (2006).

3.04.3.1 Operation of Infrared Detector

Figure 2 shows the general construction of the thermal IR detector in which a thermometer is thermally and electrically linked with a substrate with a support structure. This construction assumes an electrical readout. Although there are some examples that adopt optical readout methods and do not need electrical links, we limit our discussion to the electrical readout thermal IR detectors here. The IR radiation is absorbed in an IR absorber attached to the thermometer to change the temperature of the thermometer. The support structure provides the functions of mechanical support, thermal conducting paths to the substrate, and electrically conducting paths. If the thermal conductance of the support legs is sufficiently low, a detectable temperature change is generated at the thermometer, and a change in the output from the thermometer is read out through the support legs and the signal processing circuits on the substrate.

There are three issues to be addressed while designing the thermal IR detector. One is the type of thermometer. As reviewed in Section 3.04.2, we can utilize pyroelectric thermometers, dielectric bolometers, resistance bolometers, thermopiles, diode thermometers, bimaterial thermometers, and thermo-optical thermometers. The sensitivity and noise characteristics of the thermometer are the most important factors in selecting the thermometer. In addition to that, we have to give much importance to the material and manufacturing process for the thermometer, especially in developing the uncooled IRFPAs because monolithic integration of the thermometer on Si signal-processing circuits requires stringent rules for the material and process. The maximum process temperature after the metal interconnection process is often a serious barrier to deposit thin-film detector materials with good quality onto complementary metal oxide semiconductor (CMOS) readout circuits.

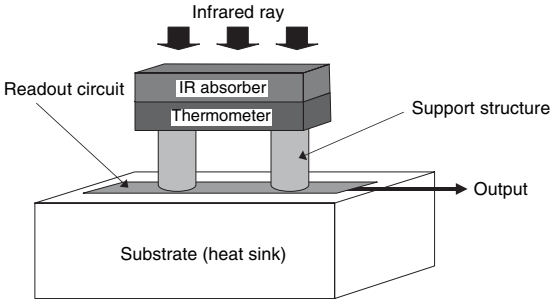


Figure 2 Schematic illustration of a thermal infrared detector, explaining the operation principle.

The second issue is related to the method of IR absorption. The design of the IR absorber influences the spectral absorptivity and heat capacity of the thermally isolated detector structure. Thermal IR detectors with higher sensitivity and faster response claim higher absorptivity and lower heat capacity. Deposition of metals such as Au at pressures of above 10^2 Pa makes flaky films called metal black with high absorptivity in a wide spectral band (Hirota and Morita 1998), and the metal black films are widely used in thermal IR detectors. Insulator films such as SiO_2 and SiN, which are used in Si-LSI processes, exhibit moderate absorption in the LWIR spectral band. Although the insulator films are not the best choice from the viewpoint of absorptivity, compatibility with Si-LSI processes is an attractive advantage of this IR absorber (Lenggenhanger *et al.* 1992). A thin metal film having a sheet resistance of 189 Ω per square can also be used for the IR absorber (Liddiard 1993). Single path through the metal film absorbs 50% of the incident IR radiation. More efficient IR absorption is possible with a 1/4 wavelength-resonant optical cavity (Capper and Elliott 2001). This resonant IR absorber consists of a metal reflector, a thin metal film having a sheet resistance of 377 Ω per square, and a dielectric film between them. The thickness of the dielectric film is tuned to be 1/4 of the wavelength of interest. Appropriate design of the resonant IR absorber achieves absorptivity more than 90% over the wavelength range of 8–14 μm. Any metal can be used for the thin metal film as resonant cavity absorber if its sheet resistance meets the requirement.

The last issue to be addressed is the structure of the thermal isolation between the detector structure and the substrate. The degree of thermal isolation is determined by the design of the support structure. Typical support structures in uncooled IRFPAs are a bump bond (Capper and Elliott 2001, Kruse 2001, Kruse and Skatrud 1997) and a microbridge (Capper and Elliott 2001, Kruse 2001, Kruse and Skatrud 1997), the details of which are discussed in Section 3.04.4 on ferroelectric IRFPAs and resistance bolometer IRFPAs. Higher thermal isolation is possible with the microbridge structure fabricated by MEMS technology (Wood 1993). Reduction of the thermal conductance has made possible recent remarkable improvements in the performance of uncooled IRFPAs.

The IRFPA is an imager that integrates a number of thermal IR detectors and has an electronic scanning capability. The pixel size of IRFPAs is from a few hundreds micrometer square to a few tens micrometer square, depending on the array size. The output from pixels is multiplexed by the vertical and horizontal scanning circuits, and is read out from a single or a few output ports. Since the output from the thermal IR detector is small, low-noise amplifiers are generally integrated on the chip. The simplest way to multiplexing is to activate only one selected pixel at a time, and scan the pixel in the X and Y directions. Application of this scanning scheme, however, is limited to relatively small-format IRFPAs with moderate sensitivity because it requires wide bandwidth for the pixel operation and has larger pixel noise when the array size becomes large. For high-performance uncooled IRFPAs, a line-addressed scheme is generally adopted. In the line-addressed IRFPA, a row having many pixels is selected simultaneously, and the output from the pixels is integrated by integrators located at the end of each column. This readout scheme makes it possible to reduce the bandwidth to a frequency corresponding to a horizontal scanning period. The CMOS technology makes it possible to integrate clock generators, bias generation/conditioning circuits, A/D converters, offset compensation circuits, on-chip device temperature monitor, and so on.

3.04.3.2 Performance of Thermal Detectors and IRFPAs

The responsivity R is defined as the output signal voltage V_S divided by the incident radiant power P_{IN} falling on the IR detector.

$$R = \frac{V_S}{P_{IN}} \qquad [1]$$

While the responsivity defined by eqn [1] is expressed in volts per watt, the unit for IR detectors that output current signals is amperes per watt. The responsivity generally depends on the wavelength, reflecting the wavelength-dependent absorptivity. As the thermal IR detector transduces the IR energy to a temperature change, and then to an electrical signal, the responsivity is broken up as given below:

$$R = \frac{\Delta T_D}{P_{IN}} \frac{V_S}{\Delta T_D} \qquad [2]$$

where ΔT_D is the temperature change at the detector due to the incident IR power P_{IN}. $\Delta T_D/P_{IN}$ in the right-hand side of eqn [2] is determined by the thermal balance consideration of the detector, which will be discussed later in this section, and $V_S/\Delta T_D$ is the sensitivity of the thermometer.

There are various noise sources in a thermal IR detector. The total noise gives the noise equivalent power (NEP) that is the lowest detectable IR radiant power for the detector. The NEP, P_N, is an important figure of merit of the photodetector, and it is given by (Hudson 1969):

$$P_N = \frac{V_{TN}}{R} \qquad [3]$$

where V_{TN} is the total root mean square (rms) noise voltage. If there are several uncorrelated noise sources, the total noise is the square root of the sum of the squares of the individual noise voltages. The NEP is the radiant power incident on an IR detector that gives rise to a signal equal to the total rms noise. The NEP is smaller for the more sensitive detector. It is, however, useful to use a figure of merit that has a large value when the detector is more sensitive, and the detectivity D is sometimes used in place of the NEP (Hudson 1969). The detectivity is simply the reciprocal of the NEP.

$$D = \frac{1}{P_N} \qquad [4]$$

The NEP and detectivity, which depend on the detector area A_D and the signal bandwidth B, are useful figures of merits in comparing specified detectors. However, further ingenuity is needed when we compare some detection mechanism and/or some material to others. Assuming that the noise is white (the noise amplitude is frequency independent), the noise voltage depends on the square root of the bandwidth. Extensive studies show that for most of the detection mechanisms, the signal-to-noise ratio depends on the square root of the detector area, and the detectivity varies inversely as the square root of the detector area. Then, a figure of merit that is independent of design parameters is defined as follows (Hudson 1969):

$$D^* = \frac{(A_D B)^{1/2} R}{V_{TN}} = \frac{(A_D B)^{1/2}}{P_N} \qquad [5]$$

where D^* (dee-star) is the normalized detectivity, the unit of which is cm $Hz^{1/2}$ W^{-1}. As the responsivity of IR detectors depends on the wavelength, we have to express D^* either in a spectral D^* (D^*_λ, where λ is the wavelength) or in a blackbody D^* ($D^*(T_B)$, where T_B is the blackbody temperature).

IR detectors and IRFPAs measure the temperature change or temperature distribution of objects. Therefore, it is worthwhile to know the sensitivity in terms of the temperature change of the observed objects. **Figure 3** shows a concept of IR imaging. The IR radiation from a scene segment, which has a one-to-one correspondence with a pixel on the IRFPA, passes through the atmosphere and is focused on the pixel by a lens. What we are interested in is the scene temperature sensitivity S, which is defined as follows:

$$S = \frac{V_S}{\Delta T_{SC}} \qquad [6]$$

where ΔT_{SC} is the scene temperature change. The scene temperature sensitivity is obtained by multiplying terms of a spectral integration of the thermal derivative of Planck's law in a spectral range of interest ($\Delta P_{SC}/\Delta T_{SC}$) and the energy transfer ratio between the scene and IFFPA ($\Delta P_{IN}/\Delta P_{SC}$) to eqn [2], namely:

$$S = \frac{\Delta P_{SC}}{\Delta T_{SC}} \frac{\Delta P_{IN}}{\Delta P_{SC}} \cdot \frac{\Delta T_D}{\Delta P_{IN}} \cdot \frac{\Delta V_S}{\Delta T_D} \qquad [7]$$

where ΔP_{IN} is the change in power falling on the detector due to the change in the object temperature. The scene temperature sensitivity means the output signal voltage due to 1 K temperature change at the observed object, and has the unit of volts per kelvin. As is evident in the above discussion, the scene temperature sensitivity is not defined for the detector itself, but for a detector module or imaging system, including optics. A useful figure of merit, the NETD is derived from consideration of the signal-to-noise ratio of the IR system (Kruse and Skatrud 1997):

$$\mathrm{NETD} = \frac{4F^2 V_{TN}}{\tau_O A_D R(\Delta P_{SC}/\Delta T_{SC})} = \frac{V_{TN}}{S} \qquad [8]$$

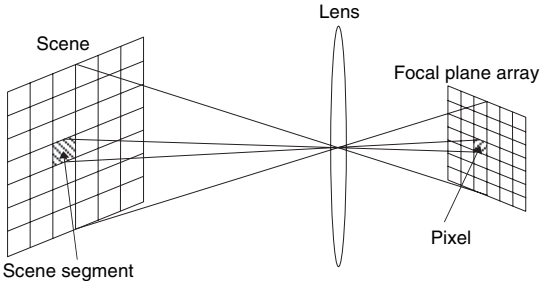


Figure 3 Concept of infrared imaging with focal plane array.

where F is F-number of optics and τ_O is the transmittance of optics. The NETD is the temperature difference of a blackbody that produces a signal-to-rms noise ratio of one.

3.04.3.3 Heat Balance in Thermal Detector

In order to understand the importance of thermal isolation, we will give a simple analysis for the heat balance in a thermal infrared detector here. The rate of energy accumulated at the thermally isolated detector structure equals the difference between the heat input and the heat loss. Although Joule heating is involved in some cases, the IR energy absorbed by the detector is the only source of heat input. An analysis including the Joule heating will be given in the section for resistance bolometer IRFPAs, in which the Joule heating restricts the operation condition.

There are three heat loss mechanisms: conduction, convection, and radiation. Conduction occurs in the way of heat flow from the thermally isolated detector structure to the substrate through the support structure. Heat conduction through the surrounding atmosphere is also an important heat loss mechanism if the IR detector is not mounted in an evacuated package. Assuming that the substrate temperature T_S equals the temperature T_A of the surroundings, the heat loss due to conduction E_{COND} is expressed as follows:

$$E_{COND} = (G_S + G_A)(T_D - T_S) \qquad [9]$$

where G_S is the thermal conductance of the support structure, G_A is the thermal conductance of the surrounding atmosphere, and T_D is the temperature of the thermally isolated detector structure.

The convection also takes place in nonevacuated packages. It is, however, a minor heat loss mechanism when compared with other heat loss mechanisms, and therefore it is neglected in the following analysis (Kruse and Skatrud 1997).

The thermally isolated detector structure radiates IR energy to its surroundings, and the surroundings radiate IR energy to the thermally isolated detector structure. If $T_D > T_S$, the radiation transfers energy from the thermally isolated detector structure to the surroundings. This radiation heat loss E_{RAD} is given by the following equation (Dereniak and Crowe 1984):

$$E_{RAD} = A_D \eta \sigma (T_D{}^4 - T_S{}^4) \qquad [10]$$

where A_D is the detector area, σ is the Stefan–Boltzmann constant, and η is the emissivity. As the temperature difference between the thermally isolated detector structure and the surroundings is very small in thermal IR detectors, we can treat the radiation mechanism like the conduction mechanisms with a thermal conductance G_{RAD} (Kruse 2002).

$$G_{RAD} = 4A_D\eta\sigma T_D^3 \quad [11]$$

Thus, the total heat loss E_{TOT} is given by

$$E_{TOT} = G_T(T_D - T_S) = G_T\Delta T_D \quad [12]$$

where

$$G_T = G_S + G_A + G_{RAD} \quad [13]$$

Eventually, we obtain the following heat balance equation.

$$C_H\frac{d(\Delta T_D)}{dt} = \eta P_{IN} - G_T \Delta T_D \quad [14]$$

where C_H is the heat capacity of the thermally isolated detector structure, and t is the time.

Assuming the incident IR energy to be a sinusoidal function,

$$P_{IN} = P_0 e^{j\omega t} = P_0 e^{j(2\pi f)t} \quad [15]$$

where P_0 is the amplitude of the sinusoidal incident IR energy, ω is the angular frequency, and f is the frequency. Then, the solution to the heat balance equation is as given below:

$$\Delta T_D = \frac{\eta P_0 e^{j\omega t}}{G_T + j\omega C_H} \quad [16]$$

The detector temperature also changes with time in a sinusoidal function, the amplitude of which is given by:

$$|\Delta T_D| = \frac{\eta P_0}{G_T(1+\omega^2\tau_T{}^2)^{1/2}} \quad [17]$$

where τ_T is the thermal time constant, defined as

$$\tau_T = \frac{C_H}{G_T} \quad [18]$$

At low frequencies for which $\omega t << 1$, ΔT_D is approximated as

$$|\Delta T_D| = \frac{\eta P_0}{G_T} \quad [19]$$

It is important to note that ΔT_D is inversely proportional to G_T. This is the reason why considerable efforts have been undertaken to fabricate a high thermal isolation structure in uncooled IRFPA development. G_{RAD} for a 50-μm-square pixel is $1.5 \times 10^{-8}\,W\,K^{-1}$ when the emissivity of the IR absorber and the fill factor is unity. The radiation sets the fundamental limitation of the heat loss. As the thermal conductivity of air is $2.5 \times 10^{-4}\,W\,cm^{-1}\,K^{-1}$, the thermal conductance through air is about $2.5 \times 10^{-5}\,W\,K^{-1}$ for a 50-μm-square pixel having an interference IR absorber suspended 2.5 μm above the substrate (Capper and Elliott 2001). The state-of-the-art technology has achieved a thermal conductance of the support structure of $1 \times 10^{-7}\,W^{-1}\,K^{-1}$ or less (Kruse and Skatrud 1997). Therefore, the dominant thermal loss mechanism in a nonevacuated package is the conduction through the surrounding air. The thermal loss through the surrounding air can be eliminated by encapsulating the IRFPA in a vacuum package having a pressure of <6.7 Pa (Capper and Elliott 2001). Then, the limiting factor for the heat loss is the conduction through the support structure in current vacuum packaged uncooled IRFPAs.

At high frequencies for which $\omega t >> 1$, ΔT_D falls off inversely with the frequency and is given by:

$$|\Delta T_D| = \frac{\eta P_0}{\omega C_H} \quad [20]$$

This means that the responsivity decreases as the modulation frequency of the incident IR power increases. At a frequency corresponding to the thermal time constant ($f_c = 1/2\pi\tau_T$), the responsivity becomes $\sqrt{1/2}$ of that in the low frequency region. Furthermore, for uncooled IRFPAs, image lag is observed in such a high-frequency region. IRFPAs operating at a TV frame rate (30 frames s^{-1}) typically have thermal time constants from 10 to 20 ms.

3.04.3.4 Fundamental Performance Limit

Here, we analyze the fundamental limit of the performance of uncooled IRFPAs according to Kruse (Kruse and Skatrud 1997). In addition to the electrical noise, heat exchange with the surroundings exhibits random fluctuation caused by the statistical nature of the mechanism. If the detector temperature is the same as the surroundings, the mean square temperature fluctuations $\overline{\Delta T_D{}^2}$ of the thermal IR detector over all frequencies is given by (Kruse and Skatrud 1997):

$$\overline{\Delta T_D{}^2} = \frac{kT_D{}^2}{C} \quad [21]$$

where k is the Boltzmann's constant and C is the harmonic average of the heat capacity, which is reduced to the heat capacity of the thermally isolated detector structure in the thermal IR detector. Assuming that fluctuations in the power interchange with the surroundings are independent of frequency, and the heat balance equation is valid for fluctuations in the detector temperature and interchanged power, we obtain the following (Kruse and Skatrud 1997):

$$\overline{P_F{}^2} = 4kT_D{}^2 G_T B \qquad [22]$$

where $\overline{P_F{}^2}$ is the mean square fluctuations in interchanged power and B is the bandwidth. $\overline{P_F^2}$ gives the fundamental limit for NEP P_{NFL} (Kruse and Skatrud 1997).

$$P_{NFL} = \left(\frac{\overline{P_F{}^2}}{\eta^2}\right)^{1/2} \qquad [23]$$

According to the definition of normalized detectivity, the temperature fluctuation noise-limited D^* becomes (Kruse and Skatrud 1997):

$$D^*{}_{TF} = \left(\frac{\eta^2 A_D}{4kT_D{}^2 G_T}\right)^{1/2} \qquad [24]$$

Using R/V_{TN} obtained from eqn [24], the temperature fluctuation noise-limited NETD is expressed as follows (Kruse and Skatrud 1997):

$$NETD_{TF} = \frac{8F^2 T_D (kBG_T)^{1/2}}{\eta \tau_O A_D (\Delta P_{SC}/\Delta T_{SC})} \qquad [25]$$

As discussed above, G_T is the sum of G_S, G_A, and G_{RAD}. G_S is the dominant component in the total thermal conductance in current uncooled IRFPAs fabricated by the MEMS technology that is mounted in a vacuum package. If we can make a support structure having a much lower thermal conductance than the radiation thermal conductance, the fluctuation in the radiation heat loss limits the performance. This situation gives the fundamental performance limit of the uncooled IRFPA, called the background fluctuation noise limit. Substituting the radiation thermal conductance into eqn [24], the background noise-limited D^* is given by (Kruse and Skatrud 1997):

$$D^*{}_{BF} = \left(\frac{\eta}{16k\sigma T_D{}^5}\right)^{1/2} \qquad [26]$$

Therefore, the background noise-limited NETD is expressed as (Kruse and Skatrud 1997):

$$NETD_{BF} = \frac{16F^2 (k\sigma B T_D{}^5)^{1/2}}{\tau_O (\eta A_D)^{1/2} (\Delta P_{SC}/\Delta T_{SC})} \qquad [27]$$

The detector temperature and the temperature of the surroundings are assumed to be the same in eqns [26] and [27]. When the detector temperature differs from that of surroundings, the background noise-limited D^* becomes (Kruse and Skatrud 1997):

$$D^*{}_{BF} = \left(\frac{\eta}{8k\sigma (T_D{}^5 + T_A{}^5)}\right)^{1/2} \qquad [28]$$

and the background noise-limited NETD is expressed as (Kruse and Skatrud 1997):

$$NETD_{BF} = \frac{8F^2 (2k\sigma B (T_D{}^5 + T_A{}^5))^{1/2}}{\tau_O (\eta A_D)^{1/2} (\Delta P_{SC}/\Delta T_{SC})} \qquad [29]$$

Figure 4 shows the temperature fluctuation noise-limited and background fluctuation noise-limited NETD for an uncooled IRFPA with a 50-μm-square pixel, calculated by eqns [25] and [27] (Kruse 1995). This graph is obtained assuming the parameters shown in the figure. The hatched area indicates the attainable NETD region. The temperature fluctuation noise-limited NETD decreases as $G_T^{1/2}$ by lowering the thermal conductance, and finally reaches the background fluctuation noise-limited NETD.

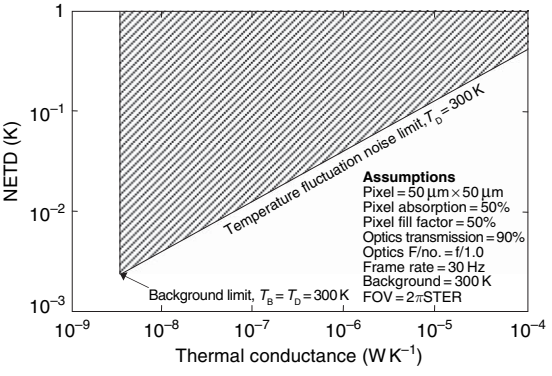


Figure 4 Temperature fluctuation noise limit and background fluctuation noise limit to noise equivalent temperature difference (NETD). NETDs in a hatched area is feasible. (Source: Kruse P W 1995 Uncooled IR focal plane array. *Proc. SPIE* **2552**, 556–63.)

3.04.4 Ferroelectric IRFPAs

3.04.4.1 Operation of Ferroelectric Infrared Detector

Figure 5 shows temperature dependences of the spontaneous polarization P_S, pyroelectric coefficient p, and dielectric constant ε of a ferroelectric (Capper and Elliott 2001). The spontaneous polarization becomes smaller with the increase in temperature. At the Curie temperature T_C, a phase transition occurs due to heating, and the ferroelectric is no longer polarized. The pyroelectric coefficient p is a measure of the rate of change of electric polarization with respect to temperature change, and is defined as follows:

$$p = \frac{\partial P_S}{\partial T} \qquad [30]$$

As depicted in **Figure 5**, the pyroelectric coefficient increases when the temperature becomes higher.

The pyroelectric IR detector uses the temperature dependence of spontaneous polarization with no applied bias voltage. A ferroelectric with higher pyroelectric coefficient has higher sensitivity for temperature. Although ferroelectrics have higher pyroelectric coefficients near T_C, most pyroelectric IR detectors are operated well below their T_C because the dielectric loss increases near T_C, degrading the performance. The dielectric bolometer uses both the temperature dependence of spontaneous polarization and the temperature dependence of dielectric constant near T_C. In this mode, an electric field is applied to the ferroelectrics, and the polarization and the pyroelectric effect are stabilized in the region of and above T_C (Capper and Elliott 2001). The electric field enhances the effective pyroelectric coefficient. The field-enhanced pyroelectric coefficient p_{FE} is given by (Hanson *et al.* 1992):

$$p_{FE} = p + \int_0^E \frac{\partial \varepsilon}{\partial T} dE \qquad [31]$$

where E is the applied electric field.

Consider a capacitor with a ferroelectric as an insulator. A change in temperature ΔT_D produces a transient current I_S to flow through an external circuit connected to the capacitor. The magnitude of I_S is given by:

$$I_S = pA_D \frac{d(\Delta T_D)}{dt} \qquad [32]$$

Using the heat balance equation, eqn [14], given in Section 3.04.3, we obtain the signal current as:

$$I_S = \frac{\eta p \omega A_D P_0}{G_T (1 + \omega^2 \tau_T{}^2)^{1/2}} \qquad [33]$$

For the pyroelectric IR detector and ferroelectric bolometer, the incident IR radiation is modulated, for example, with a light chopper, and the rms amplitude is given as:

$$I_S = \frac{\eta p \omega A_D P_0}{\sqrt{2} G_T (1 + \omega^2 \tau_T{}^2)^{1/2}} \qquad [34]$$

The capacitance has a loss resistance R_L that has a relationship with the pixel capacitor C_E, expressed as:

$$R_L = \frac{1}{\omega C_E \tan\delta} \qquad [35]$$

where $\tan\delta$ is the loss tangent, which means the angle between the resistance and the capacitive reactance. Then, the signal voltage V_S is given by the following equation (Kruse 2001):

$$\begin{aligned} V_S &= \frac{I_S R_L}{(1 + \omega^2 \tau_E{}^2)^{1/2}} \\ &= \frac{\eta p \omega A_D R_L P_0}{\sqrt{2}\, G_T (1 + \omega^2 \tau_E{}^2)^{1/2} (1 + \omega^2 \tau_T{}^2)^{1/2}} \end{aligned} \qquad [36]$$

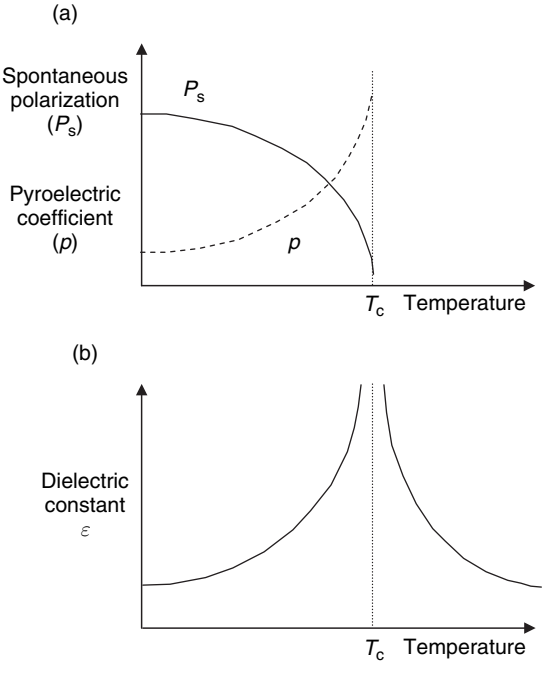


Figure 5 Temperature dependence of spontaneous polarization, pyroelectric coefficient, and dielectric constant of ferroelectric material, explaining the operation of ferroelectric infrared detector. (Source: Capper P, Elliott C T 2001 *Infrared Detectors and Emitters: Materials and Devices*. Kluwer Academic Publishers, Norwell, MA.)

where τ_E is the electrical response time defined as:

$$\tau_E = R_E C_E = \frac{1}{\omega \tan \delta} \quad [37]$$

Therefore, the responsivity is expressed as:

$$R = \frac{\eta p \omega A_D R_L}{\sqrt{2}\, G_T (1 + \omega^2 \tau_E^2)^{1/2} (1 + \omega^2 \tau_T^2)^{1/2}} \quad [38]$$

Assume that $\tau_E > \tau_T$, the responsivity rises with frequency at low modulation frequencies, reflecting the electrical response, then reaches a plateau, and finally decreases, being inversely proportional to the frequency at higher frequencies, influenced by the thermal response.

The electrical noise V_{EN} in the ferroelectric IR detector is associated with the loss resistance.

$$V_{EN} = \sqrt{\frac{kT_D}{\omega C_E \tan \delta}} \quad [39]$$

In the case of the ferroelectric IR detector, the noise is not white, but depends on the frequency according to the equivalent circuit consisting of the pixel capacitance and loss resistance.

3.04.4.2 Hybrid Ferroelectric IRFPAs

The development of ferroelectric uncooled IRFPAs started with hybrid structures in the mid-1970s. It took a long time for the ferroelectric technology to be recognized as a useful IRFPA technology. In 1992, Hanson and coworkers reported an impressive result with a 245 × 328-element hybrid ferroelectric uncooled IRFPA having a pixel structure as shown in **Figure 6** (Beratan *et al.* 1994, Hanson 1993, Hanson *et al.* 1992, Owen *et al.* 1992). The ferroelectric material used in their device is a BST ceramic. The first step of the ferroelectric wafer process is laser reticulation that makes grooves with a width of 10 μm between pixels in order to minimize the lateral thermal diffusion. The pixel pitch is 48.5 μm. After laser reticulation, a parylene backfill is deposited to replanarize the reticulated surface, and a common electrode and resonant cavity IR absorber is formed on the top of the BST wafer. Then, the BST wafer is thinned and polished to the final thickness of about 25 μm. The final step for the BST wafer is the formation of metal contacts to the Si ROIC and removal of the parylene backfill. On the Si ROIC, miniature organic mesas with over-the-edge metallization are fabricated for hybridization. This organic mesa structure reduces the thermal conductance from the BST detector to the Si ROIC, when compared with metal bump structures.

Figure 7 shows the temperature dependence for spontaneous polarization and the dielectric constant for BST ceramic (Hanson *et al.* 1992). As can be understood from this figure, this material is suitable for operating in a dielectric bolometer mode at ambient temperatures. **Figure 8** is a scanning electron microscope (SEM) photograph of pixels, showing reticulated BST detectors and organic mesa structures on the Si ROIC (Owen *et al.* 1992). The readout architecture for the BST ferroelectric uncooled IRFPA is shown in **Figure 9** (Hanson *et al.* 1992). The unit cell contains a high-pass filter, a gain stage, a tunable low-pass filter, a buffer, and an address switch. An NETD of 80 mK at *f*/1.0 was achieved with this IRFPA.

The hybrid structure shown in **Figure 10** was developed in the United Kingdom (Watton *et al.* 1996), which used lead scandium tantalate (PST) as a ferroelectric material. Like in the case of the

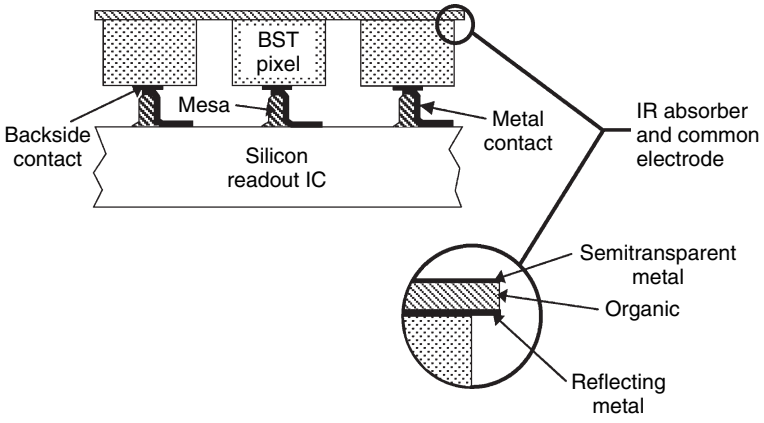


Figure 6 Hybrid pixel structure of barium strontium titanate (BST) dielectric bolometer uncooled infrared focal plane array (IRFPA). (Source: Hanson C 1993 Uncooled thermal imaging at Texas Instruments. *Proc. SPIE* **2020**, 330–9.)

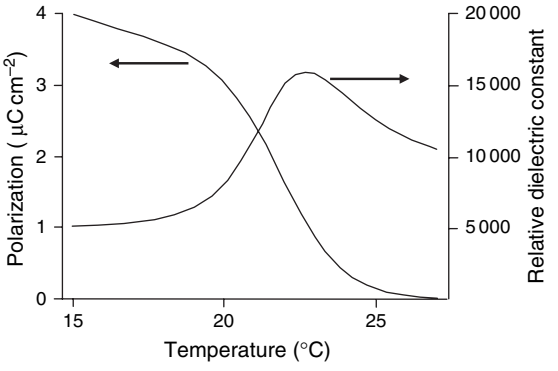


Figure 7 Temperature dependence of spontaneous polarization and relative dielectric constant of barium strontium titanate (BST) ceramic. (Source: Hanson C, Beratan H, Owen R 1992 Uncooled thermal imaging at Texas Instruments. *Proc. SPIE* **1735**, 17–26.)

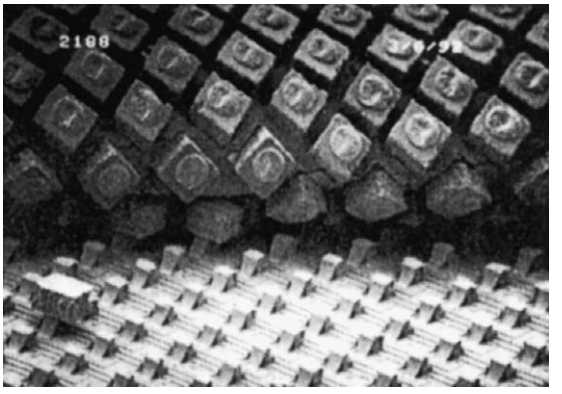


Figure 8 Scanning electron microscopy (SEM) photograph of hybrid barium strontium titanate (BST) ferroelectric pixels, showing reticulated BST pixels and organic mesa structures. (Source: Owen R, Frank S, Daz C 1992 Producibility of uncooled IRFPA detectors. *Proc. SPIE* **1683**, 74–80.)

BST uncooled IRFPA, the ferroelectric wafer is cut from a hot-pressed ceramic block and polished to around 10–15 μm thickness and is reticulated by a laser-assisted etching process. In this pixel structure, the solder bond is isolated from the Si substrate by a thick polymer layer and a narrow metal interconnection track and a small diameter of solder bond. They developed 100-μm pitch 100 × 100-element, 56-μm pitch 256 × 128-element, and 40-μm pitch 384 × 288-element ferroelectric uncooled IRFPAs (McEwen and Manning 1999, Watton and Manning 1998). Along with the device development, a three-point image-difference processing algorithm (Watton *et al.* 1996) and microscan chopper (McEwen and Manning 1999) were also implemented. The three-point image-difference processing algorithm removes errors from ramping due to leakage at the element node and eliminates thermal drift effects. The microscan chopper enhances the spatial resolution by scanning images with Ge plates on the chopper blade.

3.04.4.3 Monolithic IRFPAs with Thin-Film Ferroelectrics

Although hybrid ferroelectric uncooled IRFPAs played an important role in commercialization of IR imaging technology in the 1990s, there were serious problems when considering performance improvement and cost reduction. As can be understood from the discussion in Section 3.04.3, reduction in the thermal conductance is one of the most effective methods to enhance responsivity. By introducing the organic mesa bump structure, the thermal conductance of the hybrid uncooled IRFPA has reached a level of $4 \times 10^{-6}\,\mathrm{W/K^{-1}}$, which is sufficient to achieve an NETD of <100 mK at $f/1.0$. It is expected that extending this technology will bring further reduction in the thermal conductance to $1 \times 10^{-6}\,\mathrm{W/K^{-1}}$(Belcher *et al.* 1998). Nevertheless, this thermal conductance is still an order of magnitude larger than that of structures realized by MEMS technology. Furthermore, the hybrid process is less cost-effective than processes for monolithic structures. Reflecting these circumstances, considerable efforts have been made to realize monolithic uncooled IRFPAs with thin-film ferroelectric (TFFE) detectors (Belcher *et al.* 1998, Hanson and Beratan 2002, Hanson *et al.* 2001) or an integrated ferroelectric bolometer array (Todd *et al.* 2001, Watton and Manning 1998, Watton *et al.* 1996). The monolithic ferroelectric uncooled IRFPA offers attractive advantages over the hybrid one, such as no ferroelectric ceramic wafer preparation (cutting, thinning, and polishing), no solder bonding process, no reticulation for reducing the lateral thermal diffusion, wafer-scale processing for fully completed devices, and increased performance from superior thermal isolation provided by the microbridge (Watton *et al.* 1996).

Figure 11 illustrates a TFFE pixel structure developed by Hanson *et al.* (2001). This pixel has split top electrodes with a bottom electrode (equipotential plane), making the capacitance to be 1/4 the value of a similar capacitor whose connections are full-face electrodes on the top and the bottom. The two electrodes are transparent and together form a resonant cavity for IR absorption with a mirror on the Si ROIC. The

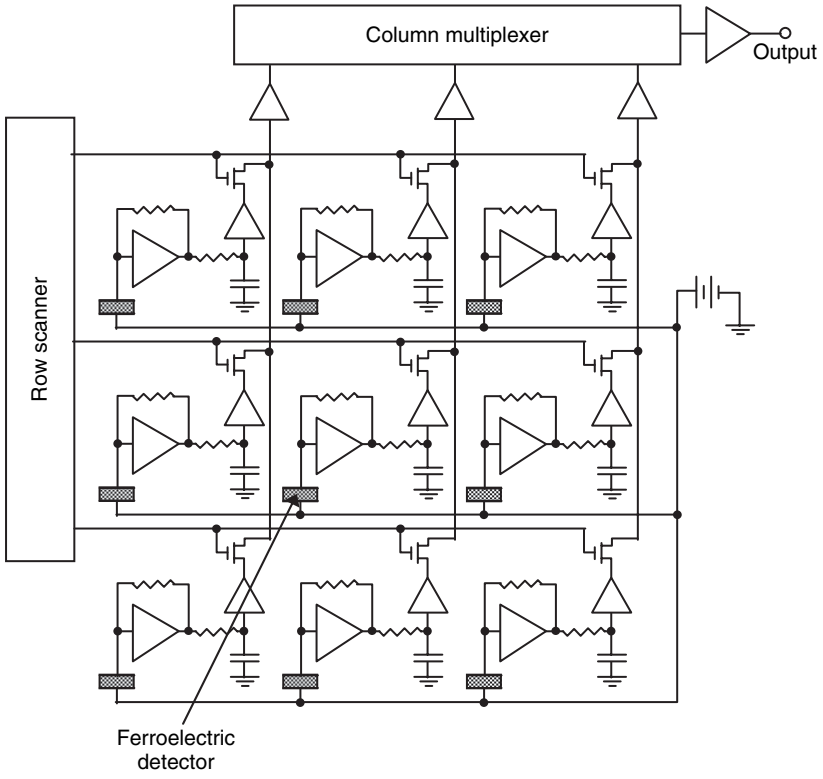


Figure 9 Readout circuit for hybrid barium strontium titanate (BST) uncooled infrared focal plane array (IRFPA). (Source: Hanson C, Beratan H, Owen R 1992 Uncooled thermal imaging at Texas Instruments. *Proc. SPIE* **1735**, 17–26.)

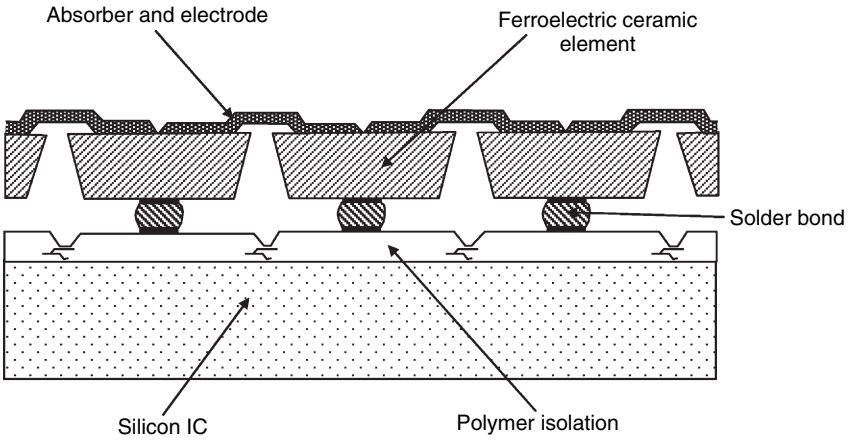


Figure 10 Hybrid pixel structure for lead scandium tantalate (PST) pyroelectric uncooled infrared focal plane array (IRFPA). (Source: Watton R, Manning P A, Perkins M J C, Gillham J P, Todd M A 1996 Uncooled IR imaging: Hybrid and integrated bolometer arrays. *Proc. SPIE* **2744**, 486–99.)

ferroelectric material is lead calcium titanate (PCT), deposited by the metal organic decomposition (MOD) of spun-on solution. They chose a pyroelectric mode rather than the dielectric bolometer mode used in their hybrid uncooled IRFPA because of its less stringent temperature control requirement.

A group in the United Kingdom also developed a monolithic ferroelectric pixel structure having full-face electrodes on the top and the bottom (Todd *et al.* 2001, Watton and Manning 1998, Watton *et al.* 1996). They tried two deposition methods with two materials: sol–gel deposited lead zirconate titanate (PZT)

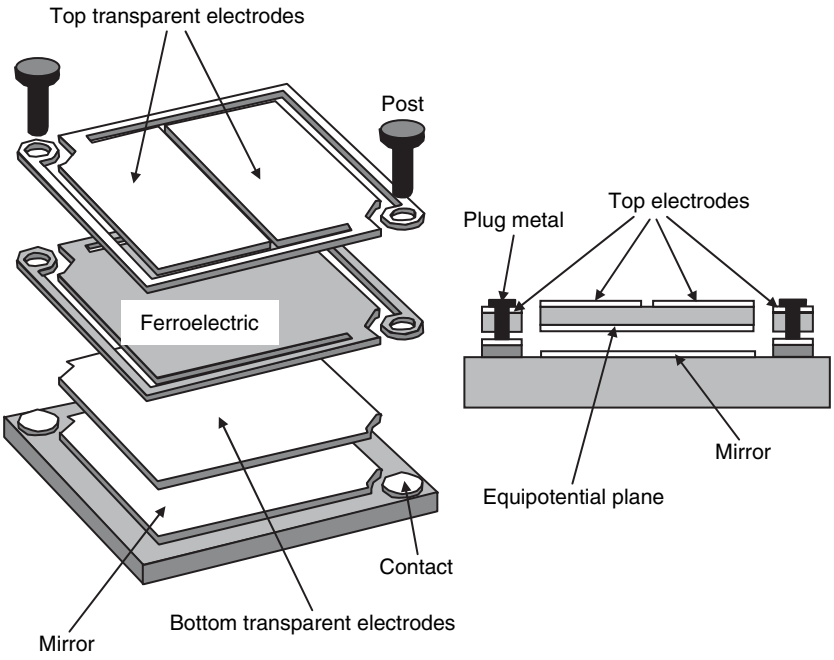


Figure 11 Monolithic thin-film ferroelectric pixel. (Source: Hanson C M, Beratan H R, Belcher J F 2001 Uncooled infrared imaging using thin-film ferroelectrics. *Proc. SPIE* **4288**, 298–303, Hanson *et al*. 2001.)

and sputtered PST, which are operated in the pyroelectric mode as well. Sol–gel is a wet chemical deposition process using the thermal decomposition of suitable metal organic precursors in solution. After several spin coating steps of the precursor solution to obtain an appropriate thickness, the film is fired just above 500°C to obtain columnar perovskite phase on a Ti/Pt electrode. Sputtered PST film was prepared from all metal targets in a dual magnetron system at 525°C. Since the figure of merit is improved by annealing at higher temperatures, rapid thermal annealing and laser annealing were studied, considering the thermal budget for the Si ROIC (Todd *et al.* 2000). Although annealing in less O_2 atmosphere increases the thermal budget, it is not sufficient to obtain a high-quality thin film suitable for high-performance uncooled IRFPAs. They, therefore, proposed a composite uncooled IRFPA structure, a cross section of which is depicted in **Figure 12** (Todd *et al.* 2000). In this structure, microbridge pixels are

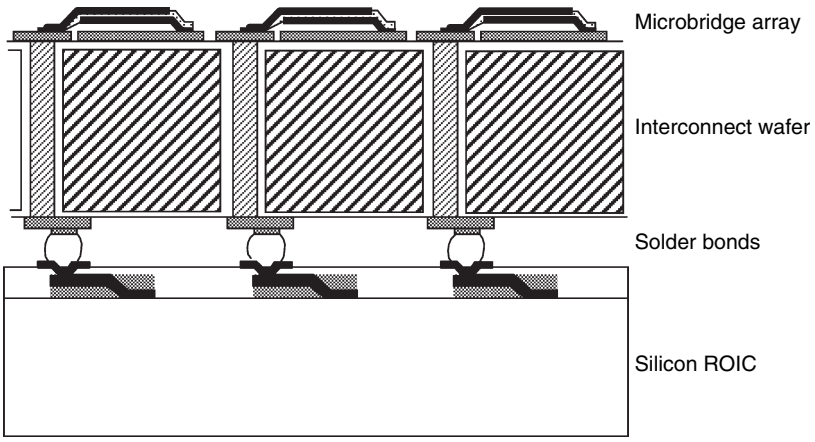


Figure 12 Composite uncooled infrared focal plane array (IRFPA) structure having microbridge pixels and interconnections through wafer. (Source: Todd M A, Manning P A, Donohue P P, Brown A G, Watton R 2000 Thin film ferroelectric materials for microbolometer arrays. *Proc. SPIE* **4130**, 128–39.)

fabricated on the front surface of a wafer, in which high-density interconnections are formed through the Si wafer. Bumps on the back surface of the wafer connect the microbridge pixels with readout circuits on the other Si wafer. Higher annealing of the detector wafer is possible before the hybridization process.

Figure 13 shows an example of a monolithic pixel structure made of BST (Xu *et al.* 1999). Anisotropic etching of a (110) Si wafer by tetramethyl ammonium hydroxide (TMAH) etchant produces the detector membrane having perpendicular cavity walls. The pulsed laser deposition (PLD) and MOD methods were employed to obtain thin BST films on Si readout circuits (Xu *et al.* 2000). By using the PLD method, BST films with good crystallographic quality are grown at relatively low temperature (520°C). This method, however, has disadvantages of poor uniformity and difficulty in enlarging the deposition area. On the other hand, larger area deposition is possible with the MOD. But, crystallizing of the spin-coated film requires a higher process temperature between 600 and 800°C. Evaluation of small linear dielectric bolometer IRFPA fabricated by the PLD and MOD methods showed a temperature coefficient of dielectric constant of around 0.5% per Kelvin, and a resulting responsivity of $100\,\mathrm{V\,W^{-1}}$ for a 200-μm-square pixel. This IRFPA uses a pixel circuit that consists of two serially connected capacitors driven by two opposite polar pulses.

Another monolithic ferroelectric uncooled IRFPA was developed using an electrospray deposited polyvinylidene fluoride (PVDF) thin film. This IRFPA also operates in a pyroelectric mode (Fujitsuka *et al.* 1997). In electrospray deposition, a PVDF solution in an organic solvent is charged by applying a high electric voltage between a needle and an electrode of the substrate, and the charged droplets of the PVDF solution are transported from the needle to the substrate by the electric field. Most solvent is vaporized during the transport process and the PVDF is deposited on the electrode, arranging its dipole due to the presence of a strong electric field. **Figure 14** shows the pixel structure of the PVDF pyroelectric uncooled IRFPA. The detector membrane is supported by four beams made of a 700-nm undoped silicate glass film. A Ti/TiN composite film, which is a LSI-compatible material, was selected for the lower electrode. After a bulk micromachining process with an ethylenediamine pyrocatechol (EDP) etchant, the PVDF film and Au-black upper electrode are deposited using a metal mask. The upper electrode also works as an IR absorber. A 16×16-element uncooled IRFPA having a pixel size of $75\,\mu\mathrm{m} \times 75\,\mu\mathrm{m}$ was fabricated by this technology. This IRFPA achieved a responsivity of $6600\,\mathrm{V\,K^{-1}}$, D^* of $1.6 \times 10^7\,\mathrm{cm\,Hz^{1/2}\,W^{-1}}$, and NETD of 0.15 K at $f/1.0$ with a 1-Hz bandwidth.

3.04.5 Resistance Bolometer IRFPAs

3.04.5.1 Operation of Resistance Bolometer Infrared Detector

The resistance bolometers detect IR radiation by their resistance change. The most important figure

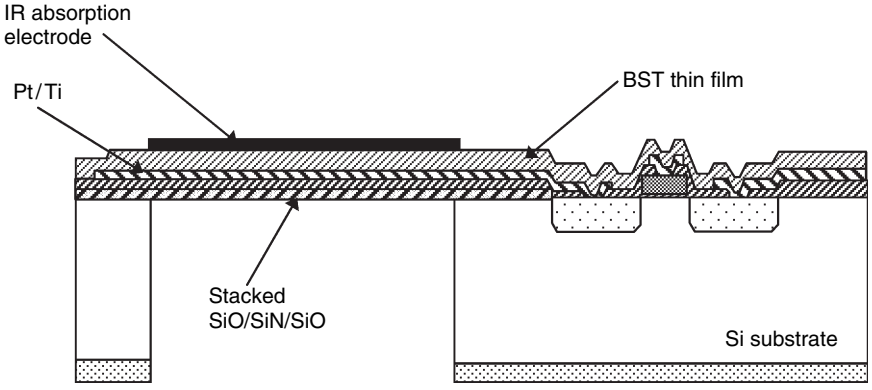


Figure 13 Thin-film ferroelectric monolithic pixel structure made of barium strontium titanate (BST). (Source: Xu H, Mukaigawa T, Hashimoto K, Kubo, R, Kiyomoto T, Zhu H, Noda M, Okuyama M 1999 Si monolithic microbolometers of ferroelectric BST thin film combined with readout FET for uncooled infrared image sensor. *Tech. Dig. 10th Int. Conf. Solid-State Sensors and Actuators (Transducers)*, Sendai, Japan, pp. 398–401.)

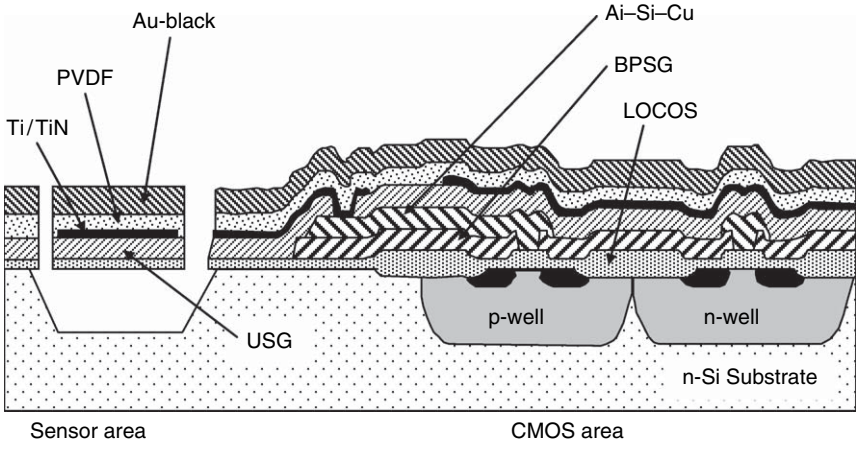


Figure 14 Monolithic ferroelectric uncooled infrared focal plane array (IRFPA) with electro-spray deposited PVDF. (Source: Fujitsuka N, Sakata J, Miyachi Y, Mizuno K, Ohtsuka K, Taga Y, Tabata O 1997 Monolithic pyroelectric infrared image sensor using PVDF thin film. *Proc. Int. Conf. Solid-State Sensors and Actuators (Transducers)*, Chicago, IL, USA, pp. 1237–1240; reproduced with permission from IEEE.)

of merit for the resistance bolometer is the TCR α, defined as:

$$\alpha = \frac{1}{R_B}\frac{dR_B}{dT} \quad [40]$$

where R_B is the resistance of the bolometer. Both metals and semiconductors are used for resistance bolometer uncooled IRFPAs. The resistance of metals increases with temperature rise, reflecting increased carrier scattering (decreased mobility) in higher temperatures. For metals near room temperatures, R_B is expressed as:

$$R_B = R_{B0}\{1 + \gamma(T - T_0)\} \quad [41]$$

where R_{B0} is the resistance at a temperature T_0 and γ is a constant. From the definition, the TCR of metals is given by:

$$\alpha = \frac{\gamma}{1 + \gamma(T - T_0)} \quad [42]$$

Metals typically have TCR values on the order of $10^{-3}\,K^{-1}$.

In contrast to metals, semiconductors exhibit negative TCRs, which means that the resistance becomes low when the temperature rises. The effect is caused by the temperature dependence of the carrier density and/or mobility. Mobile carriers in the semiconductors are generated by thermal excitation, and as a result the number of carriers increases exponentially with inverse temperature. The mobilities of some semiconductors, in which hopping over barriers at grain boundaries is the major carrier transport mechanism, also have exponential dependences on the temperature. Therefore, a resistance bolometer made of a semiconductor has characteristic represented as:

$$R_B = R_{B0}\exp\left\{\beta\left(\frac{1}{T} - \frac{1}{T_0}\right)\right\} \quad [43]$$

where β is a constant. Substituting eqn [43] into eqn [40], we obtain the TCR for a semiconductor as:

$$\alpha = -\frac{\beta}{T^2} \quad [44]$$

The TCRs of semiconductors are an order of magnitude higher than that of metals. Although eqns [42] and [44] show temperature-dependent TCRs, in designing uncooled IRFPAs we generally regard the TCR as a constant because the temperature change of the detector is very small.

When a resistance bolometer is driven with a constant bias current, then we can measure the resistance from the voltage drop across the bolometer. Assume that the IR radiation changes the detector temperature by ΔT_D, and as a consequence changes the resistance of the bolometer by ΔR_B, then the significant signal output V_S is given by:

$$V_S = I_B\ \Delta R_B = I_B \alpha R_B\ \Delta T_D \quad [45]$$

where I_B is the bias current flowing through the bolometer.

Besides the temperature fluctuation noise and background fluctuation noise, the resistance bolometer has two noise sources: Johnson noise and $1/f$

noise. The Johnson noise is associated with the random motion of carriers in a conductor, and is given by the following equation:

$$V_{JN} = (4kT_D R_B B)^{1/2} \quad [46]$$

and the $1/f$ noise V_{fN} is approximated by the following equation (Kruse 2001):

$$V_{fN} = \sqrt{\frac{(I_B R_B)^2 n}{f}} \quad [47]$$

where n is the $1/f$ noise parameter.

Referring to the discussion of the heat balance in Section 3.04.3, we obtain a frequency-dependent responsivity for the resistance bolometer as:

$$R = \frac{I_B \alpha R_B \eta}{G_T (1 + \omega^2 \tau_T^2)^{1/2}} \quad [48]$$

Responsivity is directly proportional to the bias current, the TCR, and the resistance. In the uncooled IRFPAs, the readout circuits determine the range for the acceptable resistance. Since improvement of the TCR involves time-consuming material research and it is very difficult to find higher TCR materials having appropriate resistivity and low $1/f$ noise, the TCR of bolometers used in uncooled IRFPAs has been limited to 0.02–0.03 K^{-1} for more than a decade. Although a method readily adopted to enhance the sensitivity is to increase the bias current, we have to consider the self-burnout problem and the increase in the $1/f$ noise. Here, we will discuss the self-burnout problem according to Kruse (Kruse *et al.* 1962). He analyzed the self-burnout problem with a series circuit of a bolometer and photo-insensitive load resistor with a constant voltage power supply.

We start with the condition where the only heat generation source is the Joule heating and there is no incident radiation. Then the heat balance equation is written as (Kruse *et al.* 1962):

$$C_H \frac{dT_D}{dt} = I_B^2 R_B - G_{T0}(T_{DJ} - T_S) \quad [49]$$

where G_{T0} is the thermal conductance and T_{DJ} is the temperature of the bolometer when the bias current is flowing. The first term in the right-hand side represents the heat generation by Joule heating. In the presence of an incident IR power of magnitude P_{IN}, the temperature of the bolometer rises from T_{DJ} to T_D, and the temperature difference $\Delta T_D = (T_D - T_{DJ})$ fulfills the heat balance equation (Kruse *et al.* 1962).

$$C_H \frac{d\Delta T_D}{dt} = \frac{d\left(I_B^2 R_B\right)}{dT} \Delta T_D - G_T \ \Delta T_D + P_{IN} \quad [50]$$

where the first term in the right-hand side is the increase in the Joule heating energy due to the rise in temperature. Considering the definition of the TCR and the steady-state condition for eqn [49], eqn [50] can be rewritten as (Kruse *et al.* 1962):

$$C_H \frac{d\Delta T_D}{dt} = P_{IN} - G_E \ \Delta T_D \quad [51]$$

where

$$G_E = G_T - G_{T0}(T_{DJ} - T_S)\alpha\left(\frac{R_L - R_B}{R_L + R_B}\right) \quad [52]$$

where R_L is the resistance of the load resistor. The solution of eqn [50] for a sinusoidal incident IR energy of $P_0 \exp(j\omega t)$ is given by the following equation:

$$\Delta T_D = \Delta T_{D0} \ \exp\left(-\frac{G_E}{C_H} t\right) + \frac{P_0 \ \exp(j\omega t)}{G_E + j\omega C_H} \quad [53]$$

The second term in the right-hand side is the same as that of the solution for the heat balance equation without the Joule heating that is given in Section 3.04.3. The first term represents a transient phenomenon and causes the self-burnout problem. If G_E is positive, then the transient term becomes zero with time. In this instance, the self-burnout problem never occurs. However, if G_E is negative, it increases exponentially with time, and the temperature of the bolometer rises beyond the acceptable level. It is shown that the critical temperature for self-burnout is about 50 K (Kruse *et al.* 1962) above the ambient temperature for semiconductors having a TCR of 0.02 K^{-1}. Taking this self-burnout problem into account, current high-performance resistance bolometer uncooled IRFPAs make use of a pulse-biased scheme, in which high bias current, which causes the self-burnout if continuously biased, is applied to the bolometer in a short period so as to not reach the critical temperature.

There are many possible readout architectures for resistance bolometer uncooled IRFPAs. **Figure 15** shows a typical readout architecture used in current high-performance microbolometer uncooled IRFPAs (Kruse and Skatrud 1997). In this architecture, pixels

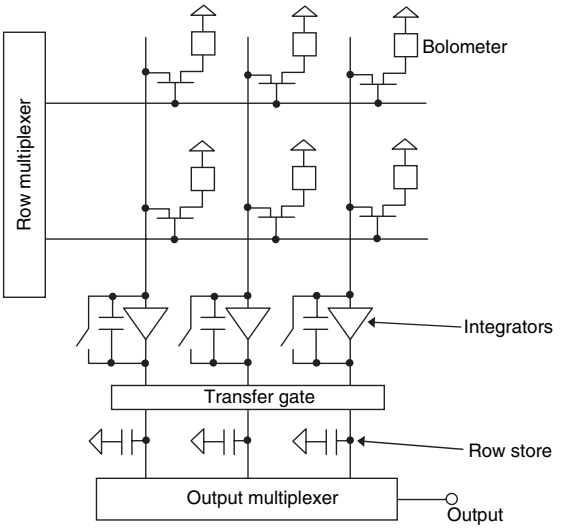


Figure 15 Typical readout architecture for resistance bolometer infrared focal plane array (IRFPA). (Source: Reprinted from *Semiconductor and Semimetals*, 47, Kruse P W, Skatrud D D, Uncooled Infrared Imaging Arrays and Systems, 94, Copyright (1997), with permission from Elsevier.)

in a row are biased simultaneously by turning on pixel selection transistors in the pixels, and outputs from the pixels are integrated by a column integrator. The integration operation not only amplifies the output signal but also restricts the bandwidth, and thus improves the signal-to-noise ratio. Therefore, a longer integration time offers a better signal-to-noise ratio. The maximum integration time for this readout architecture is a horizontal period, which equals the frame time divided by the number of rows. Integrated output signals are transferred to the sample-and-hold capacitances in the successive horizontal blanking period, and finally are read out serially.

3.04.5.2 VO_x Microbolometer IRFPAs

Wood reviewed the early works on the resistance bolometer uncooled IRFPAs (Kruse and Skatrud 1997). The development of resistance bolometer uncooled IRFPAs began with single-level pixel structures in the early 1980s. The concept and an example of the single-level pixel are shown in **Figure 16** (Kruse and Skatrud 1997). In this pixel structure, the resistance bolometer IR detector and readout electronics for the detector are arranged side by side. The bolometer film is formed in the SiN microbridge structure that is supported over an etched pit fabricated by bulk micromachining. Removal of underlying Si reduces thermal conductance and enhances sensitivity. They developed 1×16-element linear and 64×128-element 2D resistance bolometer uncooled IRFPAs with 75-μm-square single-level pixel (Kruse and Skatrud 1997). The material used in their IRFPAs was a thin-film Ni–Fe having a TCR of $+0.0023\,K^{-1}$. Other single-layer resistance bolometer detectors are also reported (Liddiard 1983). These early single-level resistance bolometer uncooled IRFPAs did not attract much attention because of their limited sensitivities caused by low TCRs of the metal bolometers and small fill factors.

Advances in surface micromachining technology made it possible to fabricate a double-level pixel structure where the bolometer film structure is supported over the readout circuits. Using the double-level structure and a semiconductor bolometer, Wood and coworkers succeeded in improving the sensitivity of resistance bolometer uncooled IRFPAs to a level better than that of ferroelectric IRFPAs (Capper and Elliott 2001, Kruse and Skatrud 1997, Wood 1993a,b, Wood *et al.* 1992). **Figure 17** shows the pixel structure of their uncooled IRFPA. A thin bolometer film of

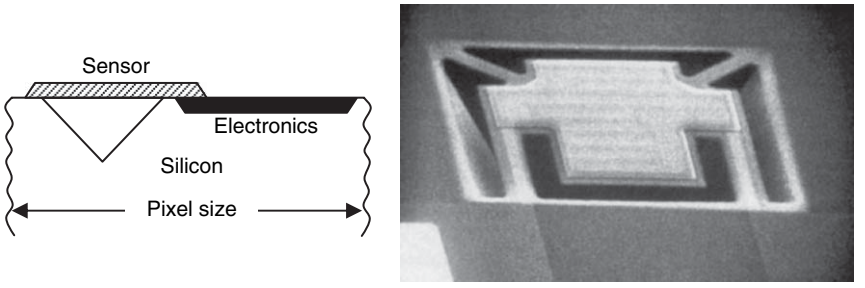


Figure 16 Single-level resistance bolometer pixel: Example of pixel photograph (*left*) and pixel layout (*right*). (Source: Reprinted from *Semiconductor and Semimetals*, 47, Kruse P W, Skatrud D D, Uncooled Infrared Imaging Arrays and Systems, 99,100, Copyright (1997), with permission from Elsevier.)

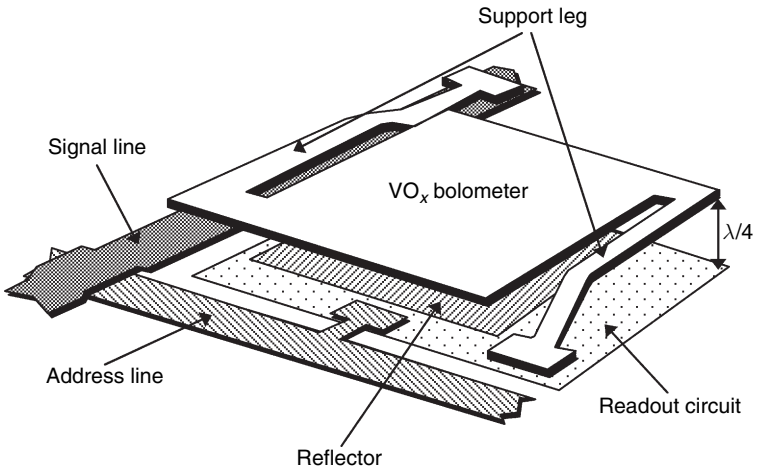


Figure 17 VO_x double-level resistance bolometer pixel. (Source: Wood R A 1993a High-performance infrared thermal imaging with monolithic silicon focal planes operating at room temperature. *Proc. IEEE IEDM*, Washington, DC, USA, pp. 175–7; reproduced with permission from IEEE.)

mixed vanadium oxide (VO_x) is deposited onto a microbridge structure that is supported by two long narrow legs. The microbridge is made of 0.5-μm-thick SiN. The supporting legs contain interconnections between the bolometer resistor and the readout circuit, which are made of 0.05-μm Ni–Cr conductive films. IR absorption is performed by a 1/4 wavelength-resonant optical cavity, consisting of an IR reflector on the substrate, IR-absorbing thin metal film on the bolometer resistor, and dielectric layer (dielectric thin films and the gap) between them.

A VO_x film can be deposited on the microbridge structure by sputtering at temperatures lower than the critical temperature for Si readout circuits. The TCR and resistivity of typical VO_x films for uncooled IRFPAs are 0.02 K^{-1} and 0.1 Ω cm, respectively. It was reported that VO_x films show no phase transition near room temperature. **Figures 18** and **19** show the characteristics of VO_x bolometers (Kruse and Skatrud 1997).

Figure 20 shows a typical fabrication sequence of a VO_x microbolometer uncooled IRFPA (Capper and Elliott 2001). After fabricating Si readout circuits, a sacrificial layer is deposited and patterned by photolithography technique. In the next step, the bottom structure of the microbridge made of SiN and VO_x film is deposited. Then, contact holes to the readout circuits are opened, and the bolometer film and readout circuits are connected by thin metal interconnections, which is covered by a SiN passivation film. The final step is a release process. After parts of the SiN layers are removed in order for an etchant to

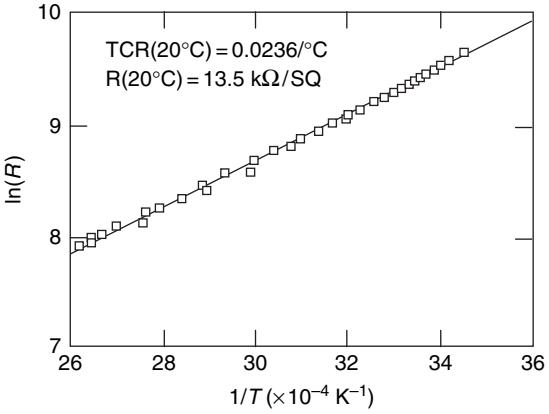


Figure 18 Temperature dependence of the resistance of typical VO_x bolometer. (Source: Reprinted from *Semiconductor and Semimetals*, 47, Kruse P W, Skatrud D D, Uncooled Infrared Imaging Arrays and Systems, 54, Copyright (1997), with permission from Elsevier.)

access the sacrificial layer, the microbridge structure is released by a selective etching process. The sacrificial layer is selected in combination with the material for the microbridge. Since the etching rate of SiO_2 to HF solution is much higher that that of SiN, SiO_2 was used as a sacrificial layer for the microbridge structure shown here in the beginning. The wet process, however, brings the sticking problem during drying, resulting in low production yield. This problem becomes serious when we reduce the thermal conductance to improve the sensitivity. Therefore, current manufacturing processes for high-performance IRFPAs use organic materials

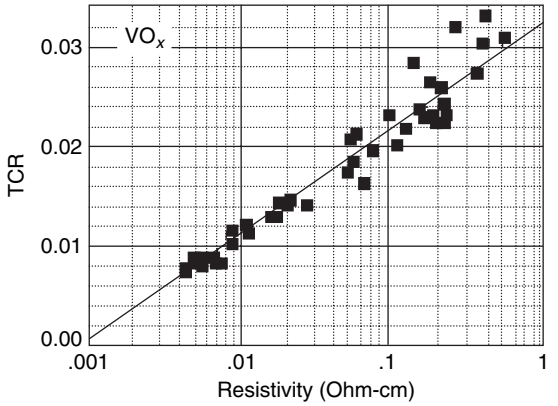


Figure 19 Temperature coefficients of resistance (TCR) versus resistivity of VO_x resistance bolometer. (Source: Reprinted from *Semiconductor and Semimetals*, 47, Kruse P W, Skatrud D D, Uncooled Infrared Imaging Arrays and Systems, 55, Copyright (1997), with permission from Elsevier.)

such as spun-on polyimides as sacrificial layers. Some organic materials for the sacrificial layer are photosensitive, and can be directly patterned by photolithography.

Using the VO_x resistance bolometer technology explained above, a 240×336-element uncooled IRFPA was developed (Wood *et al.* 1992). Thanks to the double-level structure, a fill factor of 70% and a thermal conductance of $2 \times 10^{-7}\,\mathrm{W\,K^{-1}}$ were achieved with a 50-μm-square pixel. **Figure 21** shows a SEM photograph of the pixels. The IRFPA employed a pulse-biased operation with a 5-μs pulse width and 250-μA pulse amplitude. An impressive NETD of 39 mK was reported with $f/1.0$ optics and a 30-Hz frame rate. Inspired by the success of their double-level VO_x microbolometer uncooled IRFPA, a number of IRFPAs were developed based on this technology (Herring and Howard 1996, Howard and Clarke 1999, Jerominek *et al.* 1998, Marshall *et al.* 1996, Radford *et al.* 1996, Terre *et al.* 2003, Wada *et al.* 1998). These IRFPAs have pixel sizes of around 50 μm square and array sizes from 128 μm × 128 μm to 320 μm × 240 μm.

3.04.5.3 Microbolometer IRFPAs with Other Materials

VO_x is not the only material that can be used for resistance bolometers. Si-based resistance bolometers including amorphous Si, amorphous and polycrystalline SiGe, and single-crystal Si have been studied for use in uncooled IRFPAs.

Amorphous Si is a successful example among the materials investigated for this application (Brady *et al.* 1999, Francisco 2001, Mottin *et al.* 2002, Tissot *et al.* 1998, 2000, Unewisse *et al.* 1995, Vedel *et al.* 1999). Early work on amorphous Si was done by a group in Australia (Unewisse *et al.* 1995). One of the issues in applying amorphous Si to uncooled IRFPAs is large $1/f$ noise. They prepared amorphous Si with sputtering and plasma-enhanced chemical vapor deposition

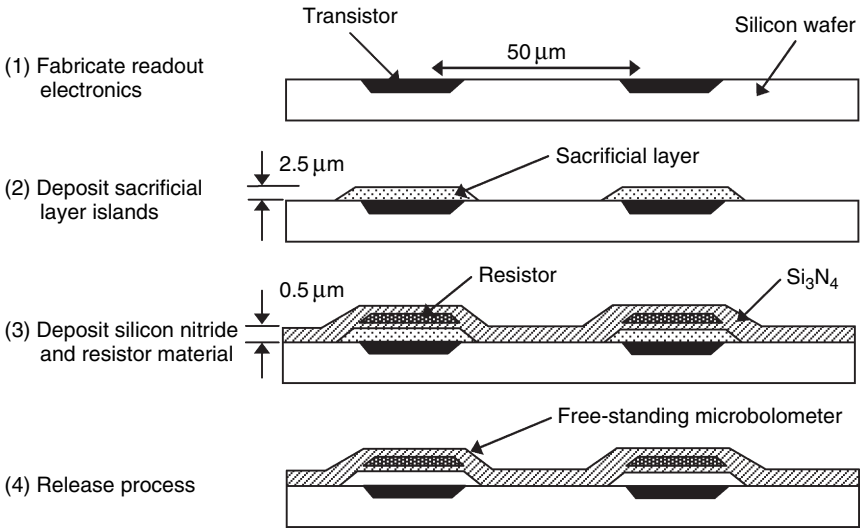


Figure 20 Fabrication sequence of double-level microbolometer pixel. (Source: Wood R A 1993a High-performance infrared thermal imaging with monolithic silicon focal planes operating at room temperature. *Proc. IEEE IEDM*, Washington, DC, USA, pp. 175–7; reproduced with permission from IEEE.)

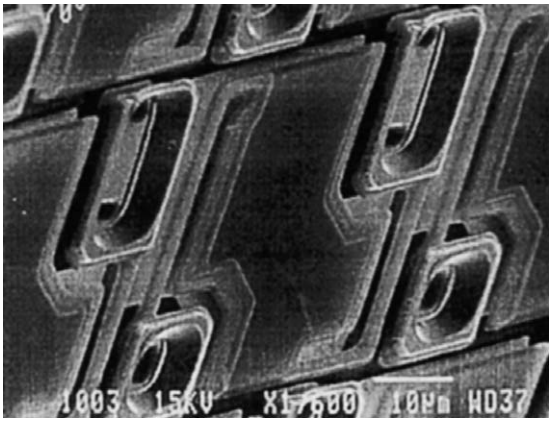


Figure 21 Scanning electron microscope (SEM) photograph of double-level microbolometer pixel. (Source: Capper P, Elliott C T 2001 *Infrared Detectors and Emitters: Materials and Devices*. Kluwer Academic Publishers, Norwell, MA.)

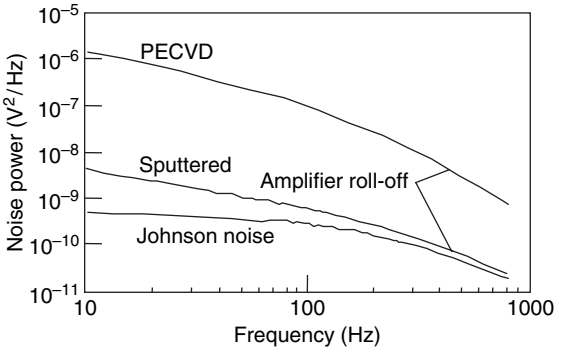


Figure 22 1/*f* noise of amorphous Si resistance bolometers prepared with plasma-enhanced chemical vapor deposition (PECVD) and sputtering. (Source: Unewisse M H, Craig B I, Watson R J, Reinholed O, Liddiard K C 1995 The growth and properties of semiconductor bolometers for infrared detection. *Proc. SPIE* **2554**, 43–54.)

(PECVD), and compared the noise characteristics of films with a similar resistivity. They found that the level of $1/f$ noise is highly dependent on the film preparation method, and amorphous Si deposited with PECVD exhibited a very high value of $1/f$ noise, as shown in **Figure 22**. They also found that amorphous silicon prepared by PECVD has a random telegraph signal noise cause by current transported along a narrow filament through the barrier of a wide-bandgap material.

Tissot *et al.* (1998) developed 256×64-element amorphous Si resistance bolometer uncooled IRFPAs with 50-μm-square pixels for a laboratory model. Their pixel has essentially the same microbridge structure as the VO_x microbolometers, but it comprises only a very thin layer (0.1 μm) of doped amorphous Si with no extra supporting layer or membrane. This makes the heat capacity of the structure very low, and offers possibility of higher frame rate operation and/or lower thermal conductance design. **Figure 23** shows the dependence of TCR on the resistivity for amorphous Si thin films (Tissot *et al.* 1998). Although a $0.05\,K^{-1}$ TCR is feasible for lightly doped films, heavily doped films with TCRs around $0.025\,K^{-1}$ are generally used for uncooled IRFPAs.

Figure 24 shows a pixel structure and a close-up photograph of the pixel near a metal stud (Mottin *et al.* 2002, Tissot *et al.* 1998). The fabrication process for this structure is depicted in **Figure 25** (Mottin *et al.* 2002). Their 256×64-element uncooled IRFPA exhibited an NETD of <50 mK with $f/1.0$ optics and high frame rate up to 100 Hz (Vedel *et al.* 1999). This technology has blossomed to a commercial product of a 320×240-element uncooled IRFPA with 45-μm pixels, which have an NETD of 70 mK at $f/1.0$ and 50 Hz (Tissot *et al.* 2000). Before the launch of the commercial product, they had to solve a reliability problem associated with the amorphous Si material. The metastable effect of amorphous Si may cause change in resistance of the bolometer after high-temperature acceleration tests if no measures against it are taken. Tissot *et al.* (2000) overcame this problem by using an original heat treatment and reported that

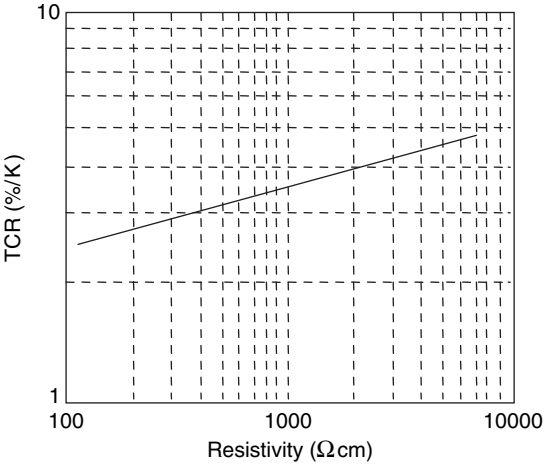


Figure 23 Temperature coefficients of resistance (TCR) versus resistivity of amorphous Si resistance bolometer. (Source: Tissot J-L, Rothan F, Vedel C, Vilain M, Yon J-J 1998 LETI/LIR's amorphous silicon uncooled microbolometer development. *Proc. SPIE* **3379**, 139–44.)

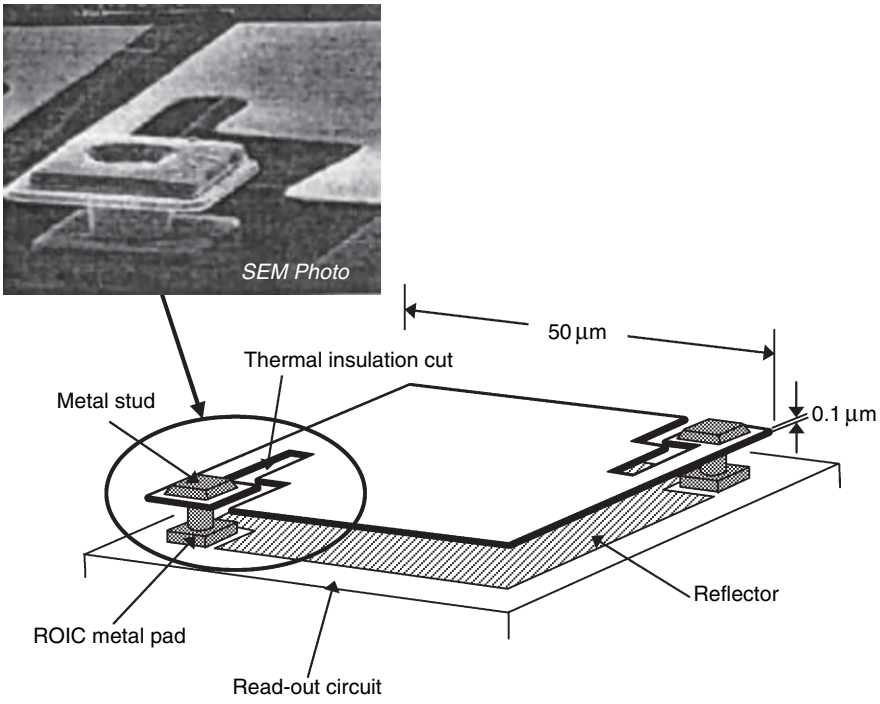


Figure 24 Self-supported amorphous Si microbolometer pixel. (Source: Tissot J-L, Rothan F, Vedel C, Vilain M, Yon J-J 1998 LETI/LIR's amorphous silicon uncooled microbolometer development. *Proc. SPIE* **3379**, 139–44; Mottin E, Bain A, Martin J L, Ouvrier-Buffet J L, Yon J J, Chatard J P, Tissot J L 2002 Uncooled amorphous silicon technology: High performance achievement and feature trends. *Proc. SPIE* **4721**, 56–63.)

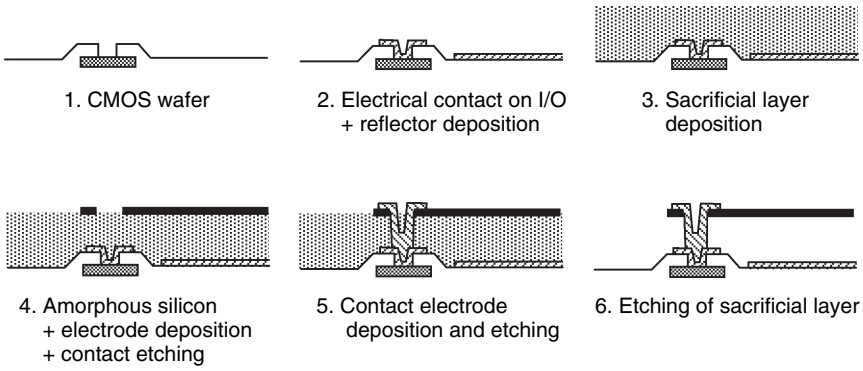


Figure 25 Fabrication process for self-supported amorphous Si microbolometer pixel. (Source: Mottin E, Bain A, Martin J L, Ouvrier-Buffet J L, Yon J J, Chatard J P, Tissot J L 2002 Uncooled amorphous silicon technology: High performance achievement and feature trends. *Proc. SPIE* **4721**, 56–63.)

no significant resistance change was observed at a storage temperature of 125°C after 1000 h.

Brady and coworkers reported a 160 × 120-element amorphous Si resistance bolometer uncooled IRFPA with a pixel size of 46.8 µm square (Brady *et al.* 1999, Francisco 2001). The high amorphous Si pixel resistance of 30 MΩ of this IRFPA allows the use of a DC detector bias. They adopted the full-frame time integration operation, which becomes possible with the high pixel resistance and a switched capacitor integrating amplifiers in a pixel.

Although the single-crystal Si has relatively low TCRs of 0.5–0.7% per Kelvin, it is attractive because of its compatibility with Si-LSI technology. Tezcan and coworkers reported results on a single-crystal Si resistance microbolometer uncooled

IRFPA (Eminoglu *et al.* 2001, Tezcan *et al.* 2001). The whole structure of the single-crystal microbolometer is fabricated using a standard CMOS process technology in a Si foundry, and the only additional process is the final release process to make a freestanding structure by anisotropic etching of the bulk Si using TMAH solution.

Polycrystalline SiGe has low tensile stress, low thermal conductivity, and TCR comparable with VO_x, and therefore it is suitable for self-supporting microbolometer uncooled IRFPAs (Henini and Razeghi 2002, Leonov *et al.* 2003, Moor *et al.* 2000, Sedky *et al.* 1998). As in the case of amorphous Si uncooled IRFPAs, the self-supported structure of polycrystalline SiGe offers possibility of lower thermal conductance and lower heat capacity than do multilayer structures. In order to make a very thin freestanding structure, Moor *et al.* (2000) developed a vapor HF surface micromachining at an elevated wafer temperature and a U-shape structure for stiffness enhancement. They noticed that water created during the reaction of the vapor HF etching process leads to sticking and it can be avoided by heating a wafer to a temperature between 30°C and 50°C. The U-shaped structure has a U-profile cross section, the idea of which is based on high stiffness of I-shape steel in the construction mechanics. The polycrystalline SiGe film is deposited by atmospheric pressure or reduced pressure chemical vapor deposition at 600°C (Henini and Razeghi 2002). The TCR of polycrystalline SiGe is $-0.015\,K^{-1}$, including the TCRs of the interconnection metals, and evaluation of a 128-element linear IRFPA with 50-μm-square pixels showed an NETD of 65 mK at $f/1.0$ (Leonov *et al.* 2003). Amorphous GeSi and GeSiO thin films are also investigated for the use of microbolometers (Rana and Butler 2005). As for amorphous GeSi, increased Si concentration gives higher resistivity and lower TCR. For GeSiO, the addition of O_2 to SiGe increases the resistivity and TCR.

High-T_c superconducting materials and colossal magneto resistance materials are expected to have high TCRs, and some of them are investigated for uncooled IRFPAs. Among several candidates, semiconducting amorphous YBaCuO has exhibited promising results (Almasri *et al.* 2000, Jahanzeb *et al.* 1997, Wada *et al.* 2000). Semiconducting YBaCuO is prepared by RF magnetron sputtering from a composite target. Monolithic structures are readily realized with this material because thin films having characteristics suitable for uncooled IRFPAs are prepared by deposition at room temperature. Almasri *et al.* (2000) produced 40-μm-square self-supported YBaCuO microbolometer pixels. The self-supporting structure was fabricated by a surface micromachining technology with a polyimide sacrificial layer. A 400-nm YBaCuO film is supported by two 10-nm Ti interconnection electrodes. Since Ti is not a good material to make a good ohmic contact to YBaCuO, Au contact pads were inserted between YBaCuO and Ti. TCR and D^* for a broadband (2.5–13 μm) radiation of their microbolometers are reported to be $-0.035\,K^{-1}$ and $10^8\,cm\,Hz^{1/2}\,W^{-1}$, respectively. Wada *et al.* (2000) developed a 320×240-element YBaCuO resistance bolometer uncooled IRFPA with 40-μm-square pixels. Their YBaCuO bolometer had a TCR of $-0.032\,K^{-1}$ and a $1/f$ noise comparable with the VO_x bolometers. The IRFPA has an NETD of 80 mK with $f/1.0$ optics, and an excellent thermal image was demonstrated. Good YBaCuO films can be deposited on various substrates, including flexible substrates such as polyimide and Kapton. A smart skin that incorporates flexible microsensors is expected using this feature of YBaCuO (Celik-Butler *et al.* 2002).

Although the TCR of metal bolometers is about an order of magnitude lower than that of semiconductor bolometers, some attempts were made with Ti bolometers (Manning *et al.* 2004, Tanaka *et al.* 1996). Ti is selected because of its full compatibility with the Si-LSI process, low thermal conductivity, and low $1/f$ noise. Tanaka and coworkers developed a 128×128-element Ti bolometer with 50-μm-square pixels. The TCR of a Ti thin film was $0.0025\,K^{-1}$, which is smaller than a bulk Ti TCR of $0.0042\,K^{-1}$. They demonstrated an NETD of 90 mK at $f/1.0$ with a 2.5-mA bias current and 5.3-μs integration time.

3.04.5.4 Small Pixel Microbolometer IRFPAs

The pixel structure of the resistance bolometer uncooled IRFPA made progress from the single level to the double level in order to enlarge the fill factor and thus to enhance the sensitivity. During the 1990s, the technology for the double-level structure was fully mastered, and many products manufactured by this technology became commercially available. Those products had pixel sizes around 50 μm square. The double-level structure worked well for relatively large pixels. Pixel size reduction, however, is mandatory in imaging device development to realize higher spatial resolution and smaller imaging systems. Further pixel size reduction from 50 μm square imposes a serious

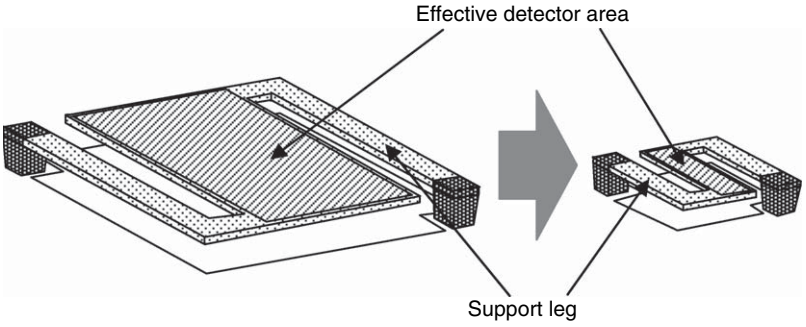


Figure 26 Difficulties in pixel size reduction for microbolometer.

problem, as illustrated in **Figure 26**. Smaller pixels collect less IR energy to elevate their temperature, and as a result have lower sensitivity. Radford *et al.* (1999) showed that lengthening the support legs is one of the most effective methods to improve sensitivity. They achieved an NETD of 8 mK at $f/1.0$ with a 50-μm-square pixel having long multibending legs. Unfortunately, the trade-off between the fill factor and the thermal conductance becomes more serious when reducing the pixel size, and therefore, their approach cannot be applied to smaller pixels. There are two approaches to overcome this problem. One is to develop new pixel structures that ensure high fill factor and low thermal conductance simultaneously. The other is to move to manufacturing processes that allow tighter design rules.

Murphy *et al.* (2002) implemented several improvements to the VO_x microbolometer pixels such as the reduction of the leg width and the introduction of dry etching micromachining process and low-density SiN film having low thermal conductivity and low specific heat. They, however, found that conventional micromachining processes suffer severe performance degradation as the unit cell size is reduced below 40 μm, and even if micromachining design rules are improved by a factor of 2, the performance of a 25-μm pixel is very poor compared with a 50-μm pixel. In order to overcome this problem, they developed a multilevel microbolometer pixel, in which the thermal isolation legs and VO_x bolometer are in different levels as shown in **Figure 27** (Murphy *et al.* 2002, 2003, 2005). This structure is fabricated by a surface micromachining process with double sacrificial layers. Using this pixel structure, 320×240 and 640×480-element uncooled IRFPAs with 25-μm-square pixels have been developed. A recent report showed that the former IRFPA has an NETD of 22 mK at $f/1.0$ and the latter IRFPA an NETD of 35 mK at $f/1.0$ (Murphy *et al.* 2005). Furthermore, a

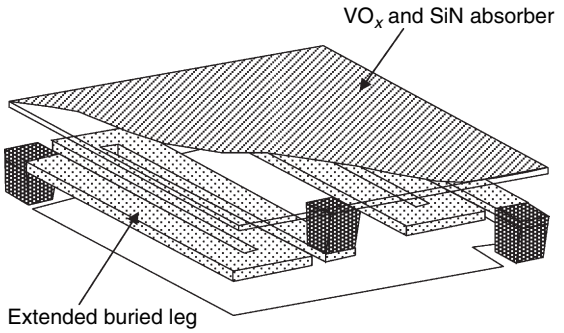


Figure 27 Multilevel VO_x microbolometer pixel having hidden support legs. (Source: Murphy D, Ray M, Wyles R, Asbrock J, Lum N, Wyles J, Hewitt C, Kennedy A, Lue D V, Anderson J, Bradley D, Chin R, Kostzewa T 2002 High sensitivity 25 μm microbolometer FPAs. *Proc. SPIE* **4721**, 99–110.)

program is progressing to develop 640×480-element uncooled IRFPAs with 20-μm-square multilevel pixels (Murphy *et al.* 2005). SEM photographs of the 20-μm multilevel microbolometer pixels are shown in **Figure 28**. Current performance of the 20-μm IRFPA was reported to be 27 mK at $f/1.0$ with a 30-Hz frame rate (Murphy *et al.* 2006). Multilevel resistance bolometer pixel structures were also reported from other organizations (Jerominek *et al.* 2000, Lee *et al.* 1998).

Another pixel structure categorized as the first approach is illustrated in **Figure 29** (Tohyama *et al.* 2004). This structure has an eaves sticking out from the freestanding bolometer that absorbs the IR radiation falling on the leg and interconnection areas. The eaves are made of SiN. Tohoyma *et al.* (2004) reported that the responsivity of a 23.5-μm-square pixel with the eaves was improved by a factor 1.3 compared with the conventional double-level pixel. Extra IR-absorbing structures are also adopted in other resistance bolometer uncooled IRFPAs (Han *et al.* 2005, Hay and Deusen 2005, Wada *et al.* 2000).

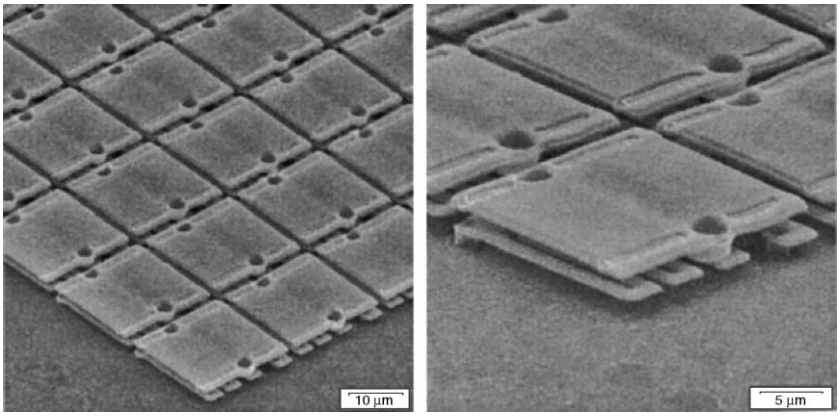


Figure 28 Scanning electron microscope (SEM) photograph of multilevel VO_x microbolometer pixel having hidden support legs. (Source: Murphy D, Ray M, Kennedy A, Wyles J, Hewit C, Wyles R, Gordon E, Sessler T, Baur S, Lue D V, Anderson S, Chin R, Gonzalez H, Pere C L, Ton S, Kostrzewa T 2005 Expand applications for high performance VO_x microbolometer FPAs. *Proc. SPIE* **5783**, 448–59.)

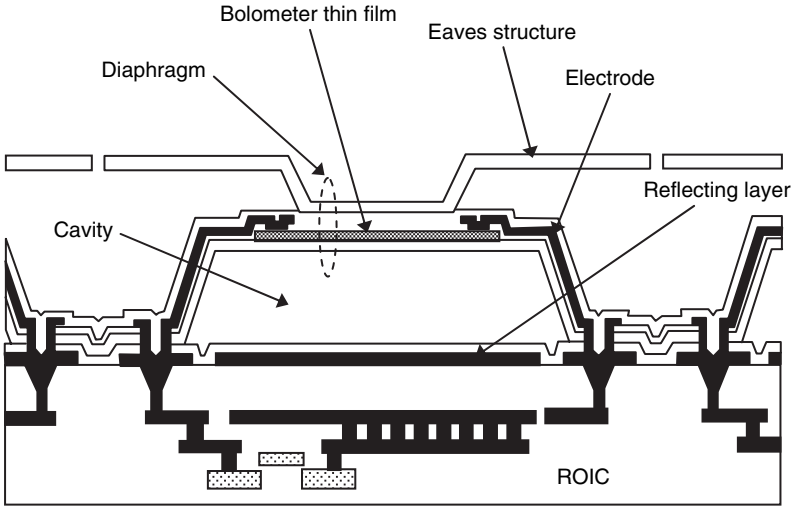


Figure 29 Microbolometer pixel with eaves structure for enhanced infrared absorption. (Source: Tohyama S, Miyoshi M, Kurashina S, Ito N, Sasaki T, Ajisawa A, Oda N 2004 New thermal isolation pixel structure for high-resolution uncooled infrared FPAs. *Proc. SPIE* **5406**, 428–36.)

The conventional VO_x double-level pixel structure is still being used in the pixel sizes from 28 to 25 μm by reducing minimum design rules (Altman *et al.* 1999, Mizrahi *et al.* 2005, Norton and Kohin 2005, Norton *et al.* 2006). An example is a 28-μm pixel designed with 0.8-μm fine pattern design rules for a 640 × 480-element IRFPA (Altman *et al.* 1999). This pixel achieved a thermal conductance of $5 \times 10^{-8}\,\mathrm{W\,K^{-1}}$ and a fill factor of 64%. In addition to the tighter design rules, a single-contact-per-pixel design contributes to fill factor improvement of IRFPAs with the smaller double-level pixel (Norton *et al.* 2006).

The amorphous Si resistance bolometer pixel is also being miniaturized while preserving the double-level pixel structure (Mottin *et al.* 2001, Tissot *et al.* 2006, Yon *et al.* 2005). In order to fabricate a 35-μm pixel and a 25-μm pixel, enhanced amorphous Si technology has been developed, which includes an improvement of thermal resistance by more than a factor of 3, reduction in the $1/f$ noise by increasing the bolometer thickness and improved process for the bolometer, and advanced design rules for stud formation (Mottin *et al.* 2001, Yon *et al.* 2005). Utilizing the enhanced amorphous Si technology, a 320 × 240-element IRFPA with a 35-μm pixel pitch achieved an NETD of 35 mK at $f/1.0$

(Mottin *et al.* 2001), and a 640 × 480-element IRFPA with a 25-μm pixel pitch 48 mK at *f*/1.0 (Tissot *et al.* 2006).

3.04.5.5 MWIR Response

In the quantum IRFPA field, multiband integration including IR detectors for LWIR and MWIR spectral bands is one of the most important themes for the third-generation thermal imager. On the other hand, until recently, uncooled IRFPAs have not been used for the detection of the MWIR spectral band because of their lower responsivity. But, recent progress in the uncooled IRFPA has made the responsivity comparable with that of quantum IRFPAs categorized in the second generation, and thus has made it possible to expand applications of the uncooled IRFPA to the MWIR spectral region. Tissot *et al.* (2004) showed that a 160 × 120-element amorphous Si uncooled IRFPA optimized for LWIR detection had a useful response (NETD = 170 mK at *f*/1.0) in the MWIR spectral band while its LWIR NETD was 30 mK at *f*/1.0. Simulations of IR absorption were performed to optimize the thickness of the bolometer bridge and the gap between the bolometer and the reflector for dual-band detection (Murphy *et al.* 2005, Norton and Kohin 2005). **Figure 30** shows an example of a simulated IR absorption that enables a MWIR NETD better than 100 mK at *f*/1.0 (Norton and Kohin 2005).

3.04.5.6 Packaging

As discussed in Section 3.04.3, the microbolometer uncooled IRFPA is encapsulated in a vacuum package because the thermal conductance of the support legs in the microbolometer is much lower than that of the air around the pixel. **Figure 31** explains how the surrounding pressure has an influence on the responsivity (Liddiard 1983). **Figure 32** shows a vacuum package for the first VO_x microbolometer uncooled IRFPA with 240 × 336 pixels (Capper and Elliott 2001). The package is constructed of brazed and soldered materials, which are carefully cleaned and baked before sealing to produce low outgassing. It has getters to absorb outgassing and a thermoelectric temperature stabilizer that controls the IRFPA temperature in order to eliminate the drift of output due to the temperature change.

Packaging using the package shown in **Figure 32** is a costly process and is not suitable for mass production. Mottin *et al.* (2001) reported a packaging concept that allows the simultaneous assembling of four IRFPAs. The package is made of 4inch Si wafer with cavities and low-temperature nonevaporable getters. Cole *et al.* (2001) developed a wafer-level packaging technology named integrated vacuum package (IVP) process. The IVP process consists of bonding a Si cap wafer to the IRFPA wafer to produce a 4inch bonded double wafer having multiple arrays. A low-oxygen floating-zone Si wafer having a thickness of 250 μm is used for a cap wafer. Sealing is performed by a PbSn solder. Since the vacuum cavity of the IVP is very small, a small amount of outgassing during wafer bonding degrades the

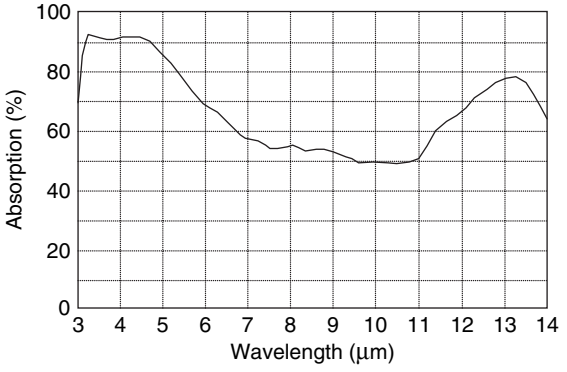


Figure 30 Simulated infrared absorption of resonant cavity absorber optimized for middle-wavelength infrared (MWIR) detection. (Source: Norton P W, Kohin M 2005 Technology and applications advancements of uncooled imagers. *Proc. SPIE* **5783**, 524–30.)

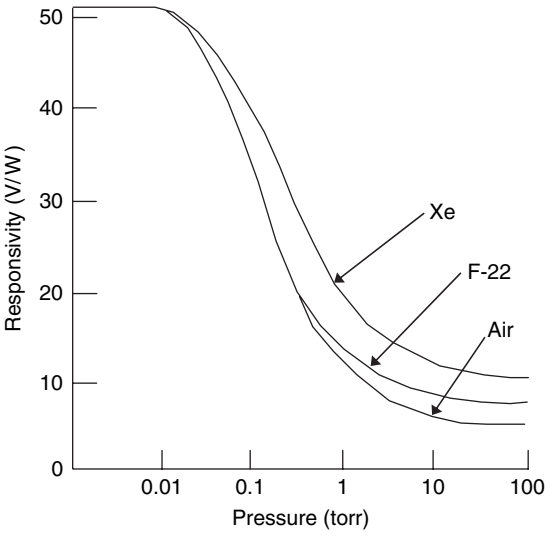


Figure 31 Dependence of responsivity on pressure. (Source: Liddiard K C 1983 Thin-film resistance bolometer IR detectors. *Infrared Phys.* **24**, 57–64.)

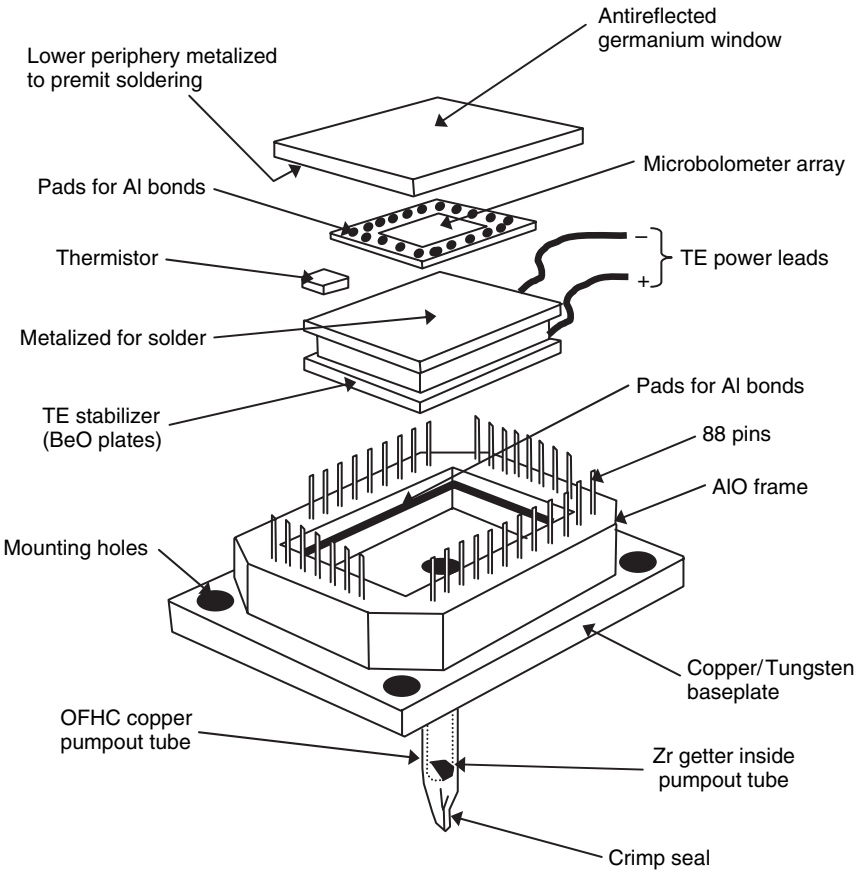


Figure 32 Vacuum package for VO_x microbolometer uncooled infrared focal plane array (IRFPA) (Source: Capper P, Elliott C T 2001 *Infrared Detectors and Emitters: Materials and Devices*. Kluwer Academic Publishers, Norwell, MA.)

vacuum level. They were able to solve this problem by using an etched pumping port in the IRFPA wafer, which is sealed by an evaporated thin metal film after a baking process. Microstructured antireflection coating is adopted for the cap. **Figure 33** illustrates a part of the IVP structure. While the wafer-level packaging technology is still under development, pixel-level packaging was also proposed for uncooled IRFPAs (Astier *et al.* 2004).

3.04.6 Thermoelectric IRFPAs

3.04.6.1 Operation of Thermoelectric Infrared Detector

When two dissimilar metals or semiconductors are connected at both ends, making a closed circuit (**Figure 34(a)**), and a temperature difference is given between the two connected points, a current flows in the circuit. If one of the conductors is cut in the middle (**Figure 34(b)**) or the circuit is cut at junction B (**Figure 34(c)**), a thermoelectric voltage appears in the opened circuit. This effect is called the Seebeck effect after its discoverer.

The thermoelectric voltage V_S is proportional to the temperature difference ΔT_D ($= T_A - T_B$), or

$$V_S = \alpha \ \Delta T_D \qquad [54]$$

where α is the Seebeck coefficient. Practically, the Seebeck coefficient is measured for a pair of materials. The Seebeck coefficient, however, can be defined as a material-inherent value. Let the Seebeck coefficient of conductor 1 and conductor 2 in **Figure 34** be α_1 and α_2, respectively, then the Seebeck coefficient of this pair is expressed as:

$$\alpha = \alpha_1 - \alpha_2 \qquad [55]$$

The thermocouple has configurations shown in **Figure 34(b)** or **34(c)**. **Figure 35** shows an example

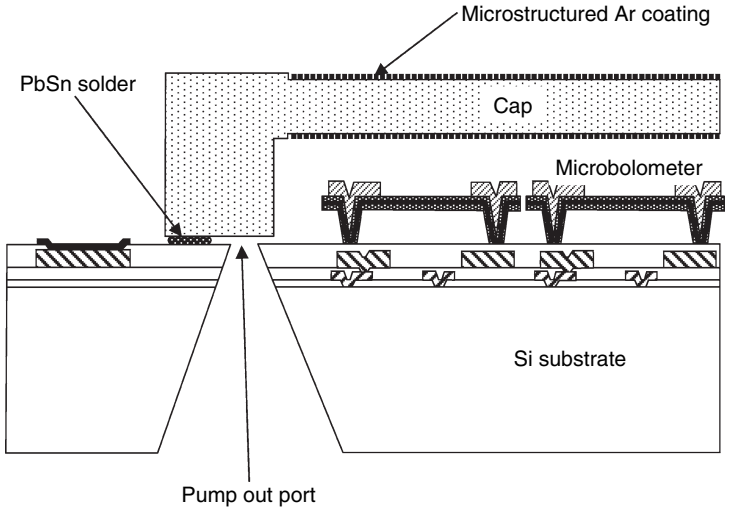


Figure 33 Integrated vacuum package for microbolometer uncooled infrared focal plane array (IRFPA). (Source: Cole B E, Higashi R E, Ridely J A, Wood R A 2001 Integrated vacuum packaging for low-cost light-weight uncooled microbolometer arrays. *Proc. SPIE* **4369**, 235–9.)

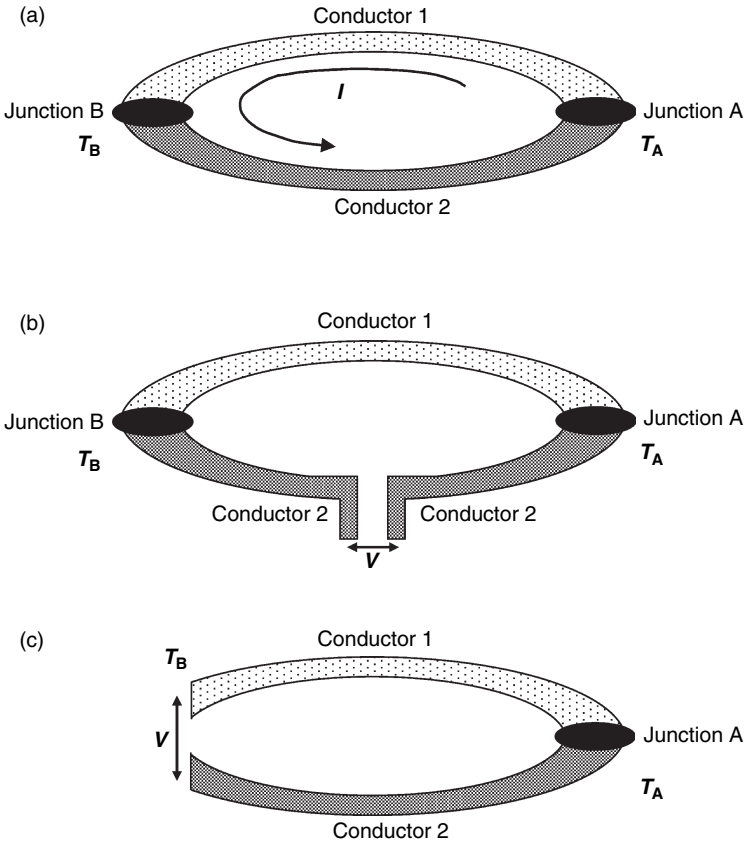


Figure 34 Generation of thermoelectric current and voltage by Seebeck effect.

of a thermoelectric IR detector, which corresponds to the configuration shown in **Figure 34(b)**. The substrate acts as a heat sink. While a cold junction is located on the substrate, a hot junction is fabricated on the thin membrane, under which the substrate is removed. IR radiation on the detector is absorbed in an IR absorber and raises the temperature at the hot junction because the membrane has a low thermal

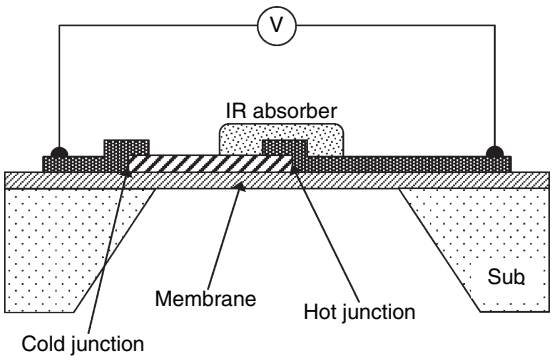


Figure 35 Structure of thermoelectric infrared detector.

conductance to the substrate. On the other hand, the temperature at the cold junction remains unchanged because it is just on the heat sink. Thus, a temperature difference between the hot and the cold junctions appears in response to IR radiation.

Since the thermoelectric voltage is additive, serial connection of hot and cold junctions of the thermocouples enhances the sensitivity. The serially connected thermocouple is called the thermopile. The thermoelectric voltage of a thermopile with m hot/cold junction pairs is given by the following equation:

$$V_S = m\alpha\Delta T_D \quad [56]$$

The electrical noise in thermoelectric IR detectors is the Johnson noise associated with the resistance of the thermocouple or thermopile R_D.

$$V_{JN} = (4kT_D R_D B)^{1/2} \quad [57]$$

Referring to the discussion in Section 3.04.4.3, one obtained the Johnson noise-limited normalized detectivity D^* of the thermoelectric IR detector as (Kruse 2001):

$$D^* = \frac{\eta m\alpha\sqrt{A_D}}{2G_T\sqrt{kT_D R_D}\sqrt{1+\omega^2\tau_T^2}} \quad [58]$$

and the NETD as (Kruse 2001):

$$\mathrm{NETD} = \frac{8F^2 G_T\sqrt{kT_D R_D B}\sqrt{1+\omega^2\tau_T^2}}{\tau_0 A_D \eta m\alpha(\Delta P_{SC}/\Delta T_{SC})} \quad [59]$$

Higher responsivity is obtained with larger Seebeck coefficient and smaller thermal conductance, and lower electrical noise is attained with smaller resistance. Considering these dependences, it is convenient to use a figure of merit for thermoelectric materials, which is defined as (Kruse 2001):

$$Z = \frac{\alpha_1 - \alpha_2}{\sqrt{\rho_1\sigma_1} + \sqrt{\rho_2\sigma_2}} \quad [60]$$

where σ_1 and σ_2 are thermal conductivity, and ρ_1 and ρ_2 are the electrical resistivity of materials 1 and 2, respectively.

3.04.6.2 Thermopile IRFPAs

Before the introduction of MEMS technology, thermoelectric IR detectors were made on thin plastic or alumina substrates. The Si-LSI-compatible MEMS technology had a great impact also on thermoelectric IR detectors from the viewpoint of sensitivity, productivity, and array integration.

The first thermopile detector fabricated by the micromachining process was reported in 1982 (Lahiji and Wise 1982). **Figure 36** shows the structure of the monolithic silicon thermopile detector. Hot junctions are formed on a thin membrane, and cold junctions on the thick silicon rim. The membrane consists of a thin dielectric film and a highly boron-diffused silicon layer. The fabrication process for this thermopile IR

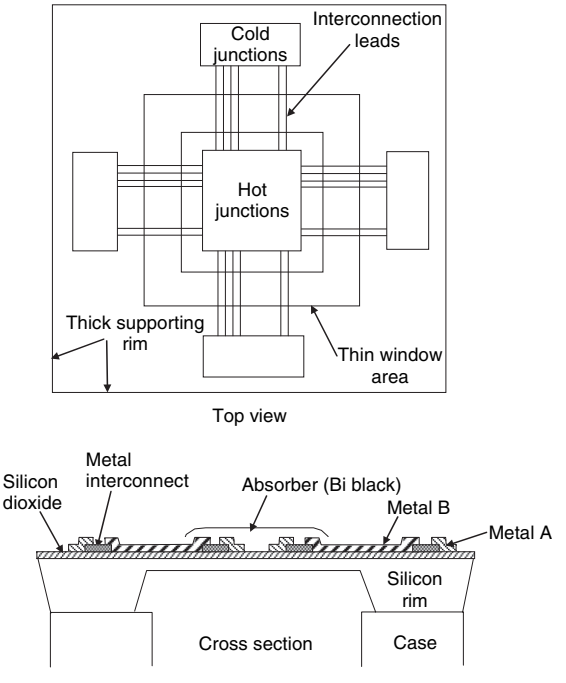


Figure 36 Structure of monolithic thermopile infrared detector. (Source: Lahiji G R, Wise K D 1982 A batch-fabricated silicon thermopile infrared detector. *IEEE Trans. Electron Devices* **ED-29**, 14–22; reproduced with permission from IEEE.)

detector is shown in **Figure 37**. The membrane is formed using an EDP etchant. As this is an anisotropic etching, and the etching rate of highly doped silicon is virtually zero, the planar shape and thickness of the membrane are precisely controlled. Bismuth black was used as an IR absorber. Using this technology, detectivities of mid-10^7 cm $Hz^{1/2}$ W^{-1} for 15-ms response times were achieved with Bi/Te, p-polysilicon/Au, and n-polysilicon/Au thermopiles. This technology for a single detector was extended to fabricate a 16×2-element staggered array with p-type polysilicon-Au thermopiles (Choi and Wise 1986).

High compatibility with commercial CMOS LSI processes is one of the most important features of thermopile uncooled IRFPAs. A group from ETH Zürich extensively made researches to develop CMOS-based thermopile uncooled IRFPAs, including the Si-LSI-compatible IR absorber. The thermopiles investigated were an aluminum/p-doped polysilicon junction and an n-doped/p-doped polysilicon junction. They started with single-element IR sensors that were fabricated by a front-accessed bulk micromachining technology (Lenggenhager *et al.* 1993). The IR sensors have IR absorbers made of dielectric films. **Figure 38** shows spectral reflectance data of dielectric films of SiN and SiO_2. Since the value of 1 minus reflectance gives the absorptivity, these data indicate that SiN and SiO_2 have useful absorption in an 8- to 14-μm spectral band. The absorptivity is enhanced by placing an aluminum reflector for double-pass absorption under the dielectric absorber and/or by using structuring absorbers (Schneeberger *et al.* 1995).

There are several possible pixel structures in realizing 2D thermopile uncooled IRFPAs (Oliver and Wise 1999). Among these pixels, IRFPAs having a single membrane by back-etched processes require pixel thermal isolation structures by which the cold junctions keep a constant temperature. Fabricating all pixels on a single membrane, the back-etched bulk micromachining process enables us to pack pixels as close as possible. **Figure 39** shows an

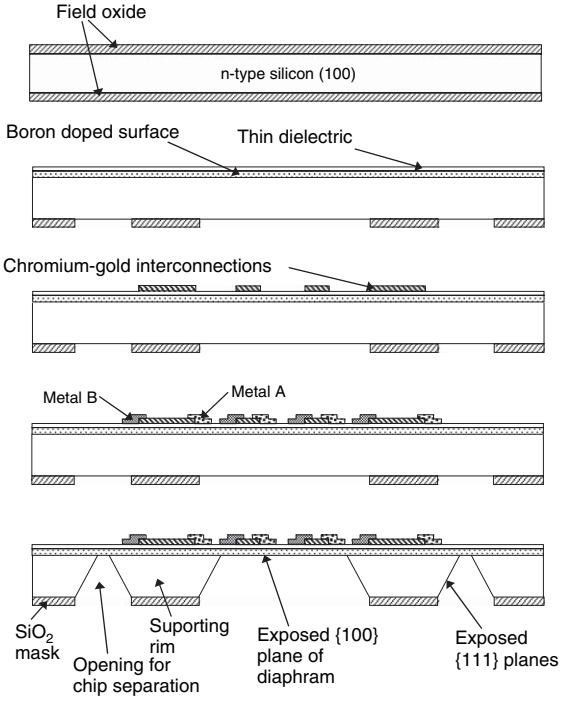


Figure 37 Process sequence for monolithic thermopile infrared detector. (Source: Lahiji G R, Wise K D 1982 A batch-fabricated silicon thermopile infrared detector. *IEEE Trans. Electron Devices* **ED-29**, 14–22; reproduced with permission from IEEE.)

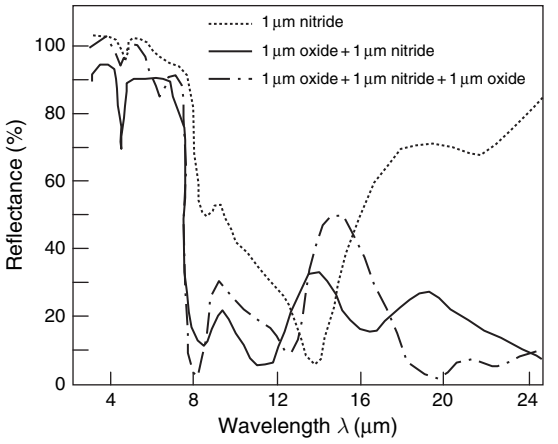


Figure 38 Spectral reflectance of SiO_2 and SiN. The value of [1-reflectance] is the absorption in the dielectric layer(s). (Source: Lenggenhager R, Baltes H, Elbel T 1993 Thermoelectric infrared sensors in CMOS technology. *Sens. Actuators A* **37–38**, 216–20.)

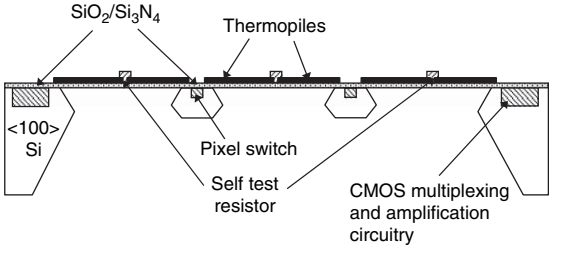


Figure 39 Thermopile uncooled infrared focal plane array (IRFPA) structure fabricated by combination of front-undercut and back-etched micromachining process. (Source: Oliver A D, Wise K D 1999 A 1024-element bulk-micromachined thermopile infrared imaging array. *Sens. Actuators* **73**, 222–31.)

example of such a structure (Oliver and Wise 1999). This structure is fabricated by the combination of a front-undercut and a back-etched micromachining technology. Etching with EDP from both sides forms the structure having heat sinks between pixels. This heat sink is made of lightly doped single-crystal silicon on which pixel switches are located. A 32×32-element IRFPA was developed with this structure. The thermopile of 32-pair p-type and n-type polysilicon thermocouples is arranged in a 375-μm-square pixel.

The other structure made by a back-etched process is shown in **Figure 40** (Münch *et al.* 1997). The pixels are thermally separated by gold lines, which are made by a tape automated bonding technology with gold bumps. A wafer-level bulk micromachining process from the back with KOH solution was developed in order to fabricate this structure. Using this technology, a 10×10-element thermoelectric uncooled IRFPA was developed (Schaufelbühl *et al.* 2001). This device integrated a low-noise chopper-stabilized amplifier. A 250-μm-square pixel has 12 thermocouples made of n-doped polysilicon and aluminum with a relative Seebeck coefficient of $108\,\mu V\,K^{-1}$. The separation gold line with 25 μm thickness and 80 μm width successfully reduces the crosstalk of <3.6%. An imaging module built with this IRFPA and an $f/1.0$ polyethylene Fresnel lens exhibited an NETD of 530 mK at a frame rate of 0.5 Hz. The sensitivity was improved by a factor of 3.3 for a successively developed 16×16-element IRFPA by employing a new pixel design and thermopile geometry optimized considering the thermal resistance and electrical resistance (Schaufelbüehl *et al.* 2001).

Front-accessed bulk-undercut processes have also been adopted to 2D thermoelectric uncooled IRFPAs. **Figure 41** shows an example of a pixel photograph fabricated by such a process (Hirota *et al.* 2001, 2003). The thermopile is made of p-doped polysilicon/n-doped polysilicon. The polysilicon layers for thermopile are doped with a boron dose of $1 \times 10^{16}\,cm^{-2}$ for p-type area and a phosphorous dose of $1 \times 10^{16}\,cm^{-2}$ for n-type area, and the junction between them is shunted by aluminum. A Au-black IR absorber is formed on the freestanding structure by lift-off process using a phosphosilicate glass (PSG) sacrificial layer (Hirota and Morita 1998). The Au-black films with thicknesses of 2–3 μm have absorptivity of >90% in a wavelength range of 8–13 μm. The developed devices are a 48×32-element IRFPA with 190-μm-square pixels and a 120×190-element IRFPA with 100-μm-square pixels. These devices are intended to operate in vacuum. When operated under atmospheric pressure, the responsivity is improved by increasing the number of thermopile pairs. The responsivity in vacuum, however, is a complicated function of the pixel structure, the materials used for support legs, the design rules for pattern layout, and the number of thermopile pairs because the dominant thermal loss mechanism is thermal conduction through the support legs. In designing pixels, they optimized the number of thermopile pairs, considering the responsivity and thermal time constant. Although the thermal time constant becomes shorter as the number of thermopile pairs increases, the dependence of the responsivity on the number of

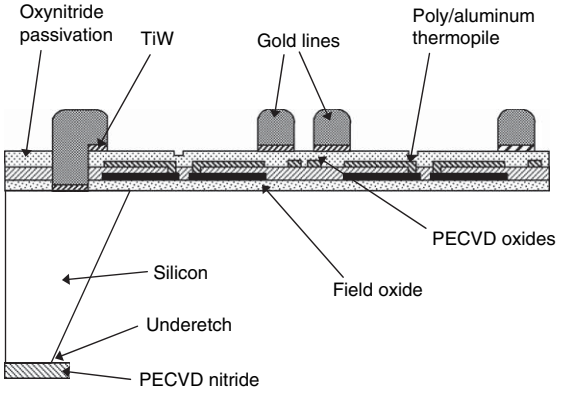


Figure 40 Thermopile uncooled infrared focal plane array (IRFPA) structure with gold bump for heat sink. (Source: Münch U, Jaeggi D, Schneeberger K, Schaufelberbühl, Paul O, Baltes H, Jasper J 1997 Industrial fabrication technology for CMOS infrared sensor arrays. *Proc. Transducers '97*, Chicago, IL, USA, pp. 205–8; reproduced with permission from IEEE.)

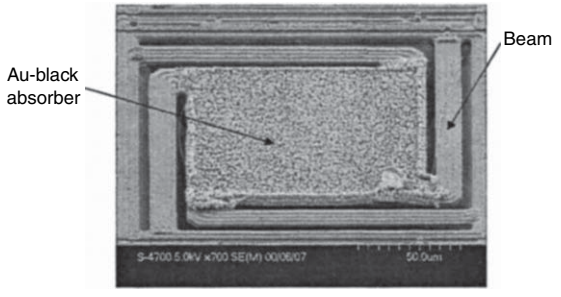


Figure 41 Scanning electron microscope (SEM) photograph of p-polysilicon/n-polysilicon thermopile pixel fabricated by front-accessed bulk-undercut process. (Source: Hirota M, Satou F, Saito M, Kishi Y, Nakajima Y, Uchiyama M 2001 Thermoelectric infrared imager and automotive applications. *Proc. SPIE* **4369**, 312–21.)

thermopile pairs generally has a peak, indicating that there is an optimum number. According to their design, the optimum numbers for the 190-μm and 100-μm-square pixels are six pairs and two pairs, respectively. For 48 × 32-element IRFPAs, the measured responsivity is 2100 V W^{-1} and an NETD of 0.4 K was obtained with *f*/0.7 optics (Hirota *et al.* 2001).

Using a surface micromachining technology, the thermopile layer can be stacked over the readout circuits. Kanno *et al.* (1994) developed a 128 × 128-element thermoelectric uncooled IRFPA with a stacked pixel. The thermopile of this IRFPA consists of p-type boron-doped polysilicon and n-type phosphorous-doped polysilicon. Doping with the impurity concentration of 10^{19} to 10^{20} cm^{-3}, the thermocouple has a Seebeck coefficient of 300–400 μV K^{-1}. The voltage generated by 32-pair thermocouples is applied to the gate electrode of an input circuit and modifies the electron injection into the accumulation capacitor. The accumulated signal charges are read out through CCDs. The stacked structure and a fine pattern of 0.6 μm width make it possible to achieve a high fill factor of 67% with a 100-μm-square pixel. An NETD of 0.5 K was reported with *f*/1.0 optics and a 120 Hz frame rate.

Although compatibility with silicon technology is one of the most important features of the thermopile uncooled IRFPA, the best performance of the thermopile uncooled IRFPA is obtained using materials impermissible in Si-LSI facilities. Compounds in the $(Bi_{1-x}Sb_x)_2(Te_{1-y}Se)_3$ family have the highest figure of merit at room temperature. Since these materials are difficult to fabricate into thin films and to be incorporated into Si-LSI processes, there are few reports on thermopile uncooled IRFPA with them. Foote *et al.* (1998) succeeded in improving the performance of thermopile uncooled IRFPAs by combining Bi–Te and Bi–Sb–Te thermoelectric materials. They fabricated uncooled linear IRFPAs with several types of pixels and achieved D^* values as high as 2.2×10^9 cm $Hz^{1/2}$ W^{-1} for 83 ms response times at a 1000 K blackbody (Foote and Jones 1998). They also developed an upheld pixel structure that is suitable for 2D IRFPAs (Foote and Gaalema 2001). **Figure 42** shows a top view and cross section of two pixels. The detectors are constructed on a silicon substrate and are fabricated by a surface micromachining process using two polyimide sacrificial layers. The IR absorber, which is connected to the isolated thermocouple junctions, is made of a silicon nitride film with a thin platinum-absorbing layer. This structure offers nearly 100% fill factor while still accommodating many thermocouples per pixel, high thermal isolation, and possibility of integrating CMOS readout circuits below the thermopile structure.

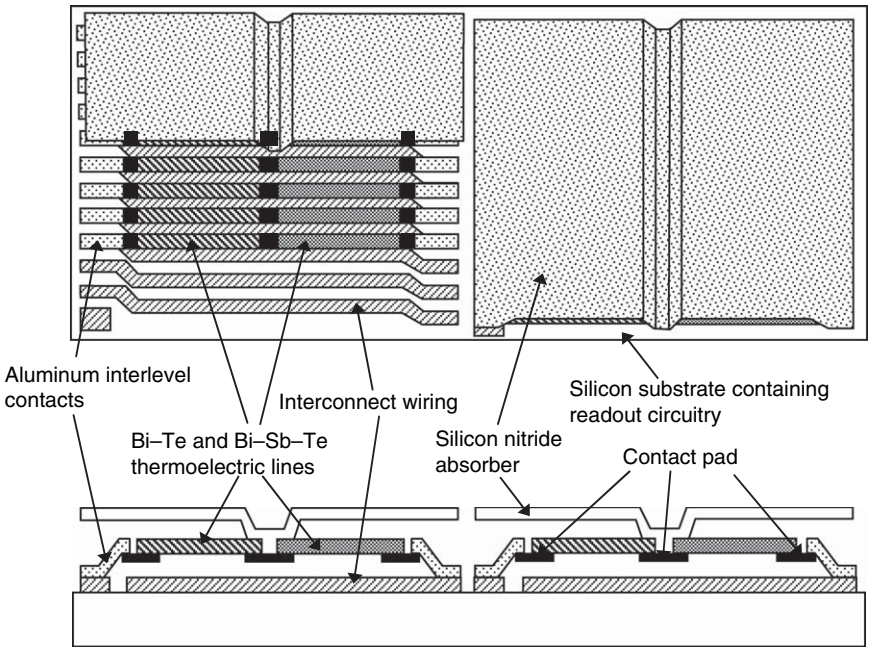


Figure 42 Upheld thermopile pixel structure having nearly 100% fill factor. (Source: Foote M C, Gaalema S 2001 Progress towards high-performance thermopile imaging arrays. *Proc. SPIE* **4369**, 350–5.)

Pursuing larger thermoelectric figure of merit, AlGaAs (Dehe *et al.* 1995) and InGaAs (Dehe *et al.* 1997) thermopiles were also investigated. Thin thermocouple membranes of AlGaAs on GaAs and InGaAs on InP were obtained by etchants with high selectivities.

3.04.7 Diode IRFPAs

Temperature-dependent characteristics of semiconductor devices sometimes bother circuit designers. They, however, can be used for temperature sensing and can be applied to uncooled IRFPAs. In fact, uncooled IRFPAs with diode temperature sensors have been successfully developed and IR cameras with some of these devices are commercially available now. We will discuss uncooled IRFPA technology with diode temperature sensors in this section.

3.04.7.1 Operation of Diode Thermal Infrared Detector

The forward-biased pn-junction diode has long been used as a temperature sensor. Assuming an ideal diffusion-limited characteristic and a sufficiently large forward-bias voltage V_F, the current I_F flowing in the diode is given by the following equation (Sze 1969):

$$I_F = A_J \mathcal{J}_S \exp\left(\frac{qV_F}{kT_D}\right) \quad [61]$$

$$\mathcal{J}_S = KT_D^{(3+\kappa/2)} \exp\left(-\frac{E_G}{kT_D}\right) \quad [62]$$

where A_J is the junction area, $\mathcal{J}_S$ is the saturation current, q is the magnitude of electronic charge, E_G is the bandgap energy, κ is a constant determined by the temperature dependence of the diffusion constant and carrier lifetime, and K is a temperature-independent constant.

When the diode is driven in a constant-current mode, the temperature sensitivity of V_F is expressed as follows.

$$\left.\frac{dV_F}{dT_D}\right|_{I_F=\text{const.}} = \frac{V_F}{T_D} - \left(3+\frac{\kappa}{2}\right)\frac{k}{q} - \frac{E_G}{qT_D} \quad [63]$$

dV_F/dT is insensitive to the process fluctuations because κ, a process-sensitive parameter, is included only in the second term of the right-hand side of eqn [63] and this term is negligible when compared with the other two terms. This feature ensures stable mass production of the diode uncooled IRFPAs. A typical forward bias condition of 0.6 V for a silicon diode gives a temperature sensitivity of about 2 mV K^{-1} at 300 K.

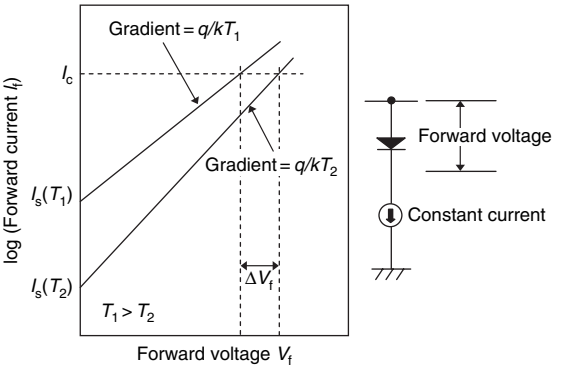


Figure 43 Operation of forward-biased diode thermometer. (Source: Kimata M, Ueno M, Takeda M, Seto T 2006 SOI diode uncooled infrared focal plane arrays. *Proc. SPIE* **6127**, 61270X-1–11.)

Figure 43 shows current–voltage characteristics of a diode for two different temperatures, explaining the operation of a diode thermometer. When the diode is operated in a constant-current mode, the forward voltage measured across the diode reflects the absolute temperature of the device. From this figure, the forward voltage is determined by $\mathcal{J}_S$, the intersection with the current axis, and the gradient q/kT_D. The dominant temperature dependence of $\mathcal{J}_S$ is derived from the exponential term in eqn [62]. The forward voltage becomes smaller as the temperature rises. In diode uncooled IRFPAs, the pixel-by-pixel difference of the forward voltage is displayed as the temperature difference.

The shot noise V_{SN} is a source of noise in the diode. It appears in statistical processes where electrical carriers flow over a potential barrier, and is given by the following equation (Ishikawa *et al.* 1999):

$$V_{SN} = \sqrt{2qI_F}\frac{dV_F}{dI_F} \quad [64]$$

In addition to the shot noise, we have to take into account the Johnson noise associated with the resistive component in the diode. In some cases, the $1/f$ noise might be one of the important noise sources.

3.04.7.2 Si Diode IRFPAs

A single-element IR detector with a polysilicon diode was reported in 1990 (Suzuki *et al.* 1990). This research is connected to the development of 16-element linear (Tanaka *et al.* 1992) and 16 × 16-element

2D (Asahi *et al.* 1992) uncooled IRFPAs with 400 μm × 400 μm pixels. **Figure 44** shows the schematic cross section of the polysilicon diode uncooled IRFPA. The diodes are made on a SiN membrane under which a cavity is formed by bulk micromachining with a polysilicon sacrificial layer. The IRFPAs integrate NMOS readout circuits, including pixel switches and shift registers. Although an image taken with 2D IRFPAs was demonstrated, preliminary results of these uncooled IRFPAs show a relatively low normalized detectivity of 6×10^5 cm $Hz^{1/2}$ W^{-1} and large fixed pattern noise (FPN) of ±20%.

Although the pixel structure with the polysilicon diodes can be fabricated with relatively simple process technology, this technology did not evolve further because of their poor performance. In its place, IR detectors with silicon single-crystal diodes were developed using back-accessed (Xu *et al.* 1993) and front-accessed (Reay *et al.* 1994) bulk micromachining processes.

Ishikawa *et al.* (1999) proposed another diode uncooled IRFPA. By using a silicon-on-insulator (SOI) wafer, they developed three freestanding pixel structures that contained single-crystal temperature-sensing diodes. **Figure 45(a)** shows a cross section of the original SOI diode pixel structure (type I) (Ishikawa *et al.* 1999, Kimata *et al.* 2006). As the figure illustrates, the pn-junction diodes are held over a cavity by two thin support legs. The thermal conductance of the support legs is designed to be as small as possible to obtain higher sensitivity. The support legs include electrical interconnections between the diodes and the readout circuits. Since the diode temperature sensor, the support legs, and the other components in the pixel have to be placed on the same level, only a small fraction of the area can be allocated to the freestanding effective IR-absorbing area if a single-level structure is employed. To overcome this problem, they devised a novel IR-absorbing structure that covers almost the entire pixel area. This IR-absorbing structure is an interference IR absorber that is used in other IRFPAs. It consists of a metal reflector, a dielectric film having a thickness that is 1/4 of the wavelength of interest, and an ultrathin metal film IR absorber. The IR-absorbing structure is thermally connected only to the freestanding diode structure with a high thermal conductance, and thus the temperature of the IR-absorbing structure is essentially the same as that of the diode structure. While conventional two-level microbolometer pixels have fill factors of around 60%, a 90% fill factor is feasible with this structure.

Although a useful sensitivity was obtained with the type I SOI diode pixel, it had a serious problem in the thermal time constant. Since the IR absorber of the type I pixel requires a thick dielectric layer to meet the tuning condition, this structure has a relatively long thermal time constant. An improved pixel structure shown in **Figure 45(b)** (type II) (Ishikawa *et al.* 2000, Kimata *et al.* 2006) was developed in order to reduce the thermal time constant. In this structure, the reflector is removed from the IR-absorbing structure, and moves to the lower level, and a vacuum gap between the thermal absorber and the reflector is used as a part of the dielectric layer in the interference IR-absorbing structure. This structure makes it possible to thin the thermal absorber to a thickness suitable for mechanical considerations.

The type II SOI diode pixel works well for relatively large pixels around 40–50 μm square, though a serious problem arises with this pixel structure when we reduce the pixel size. IR energy falling on a pixel

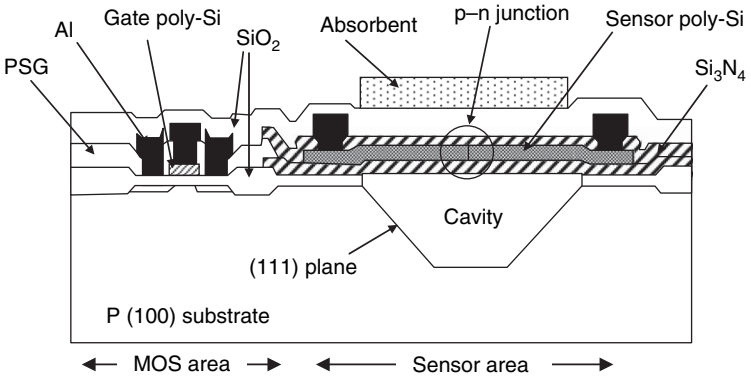


Figure 44 Polysilicon diode uncooled infrared focal plane array (IRFPA) pixel. (Source: Reprinted from *Infrared Phys.*, **33**, Tanaka A *et al.*, Infrared linear image sensor using a poly Si pn junction diode array, 229–36, Copyright (1992), with permission from Elsevier.)

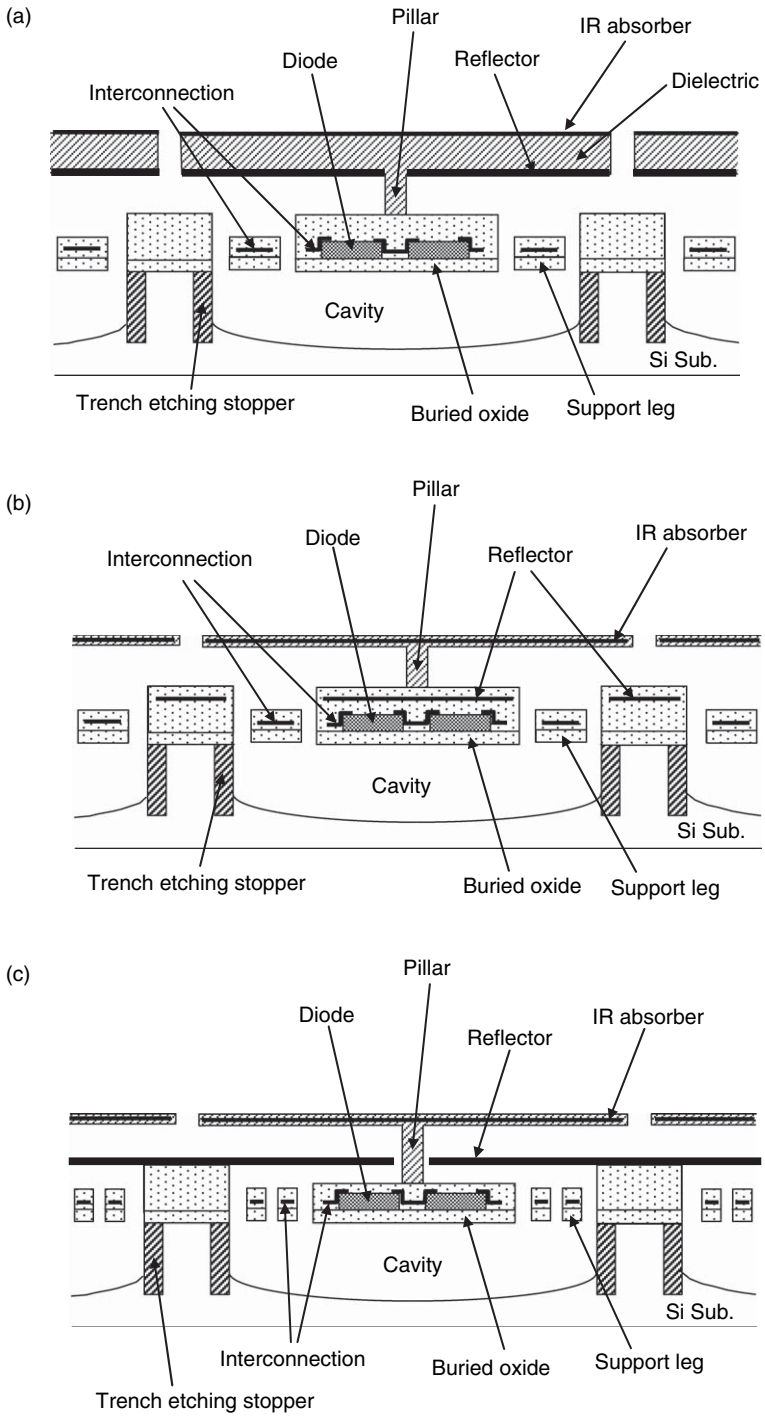


Figure 45 Three types of silicon-on-insulator (SOI) diode uncooled infrared focal plane array (IRFPA) pixels. (Source: Kimata M, Ueno M, Takeda M, Seto T 2006 SOI diode uncooled infrared focal plane arrays. *Proc. SPIE* **6127**, 61270X-1–11.)

is proportional to the pixel size, as discussed in Section 3.04.5. We, therefore, have to reduce the thermal conductance in designing small pixels if the sensitivity degradation is not acceptable. Lower thermal conductance claims more area in a pixel due to the need for longer support legs. As can be expected, the support leg area does not work as a perfect reflect or because of the slits around the

support legs, and thus reduces IR absorption. Together, these make it difficult for us to design smaller pixels with high sensitivity. Thus, a third SOI diode pixel structure shown in **Figure 45(c)** (type III) (Kimata *et al.* 2006, Kosasayama *et al.* 2004) was proposed. This pixel has an independent metal reflector between the IR absorber and the lower structure that includes the diode structure. The thermal properties of the pixel are not affected by adding this reflector structure because the reflector does not thermally connect with the IR absorber or diode structure. Since removing the reflector from the diode structure reduces the heat capacity, the type III SOI diode pixel structure offers a shorter thermal time constant than does the type II pixel structure if the thermal conductance of the two structures is the same.

The fabrication process for the SOI diode uncooled IRFPA includes bulk and surface micromachining processes (Kimata *et al.* 2006, Nakaki *et al.* 2003). It is clear that the full dry MEMS process is indispensable for fabricating the fragile structures discussed in Section 3.04.6 because the sticking problem drastically reduces the production yield if wet MEMS processes are employed. Two full dry MEMS processes have been developed for the SOI diode uncooled IRFPA. One is a single sacrificial layer process for the pixel structures of types I and II, the other a double sacrificial layer process for the pixel structures of type III.

Figure 46 illustrates a MEMS process for the type III SOI diode pixel. After completion of the process for CMOS circuits and temperature-sensing diodes, the first organic sacrificial layer is spin-coated and

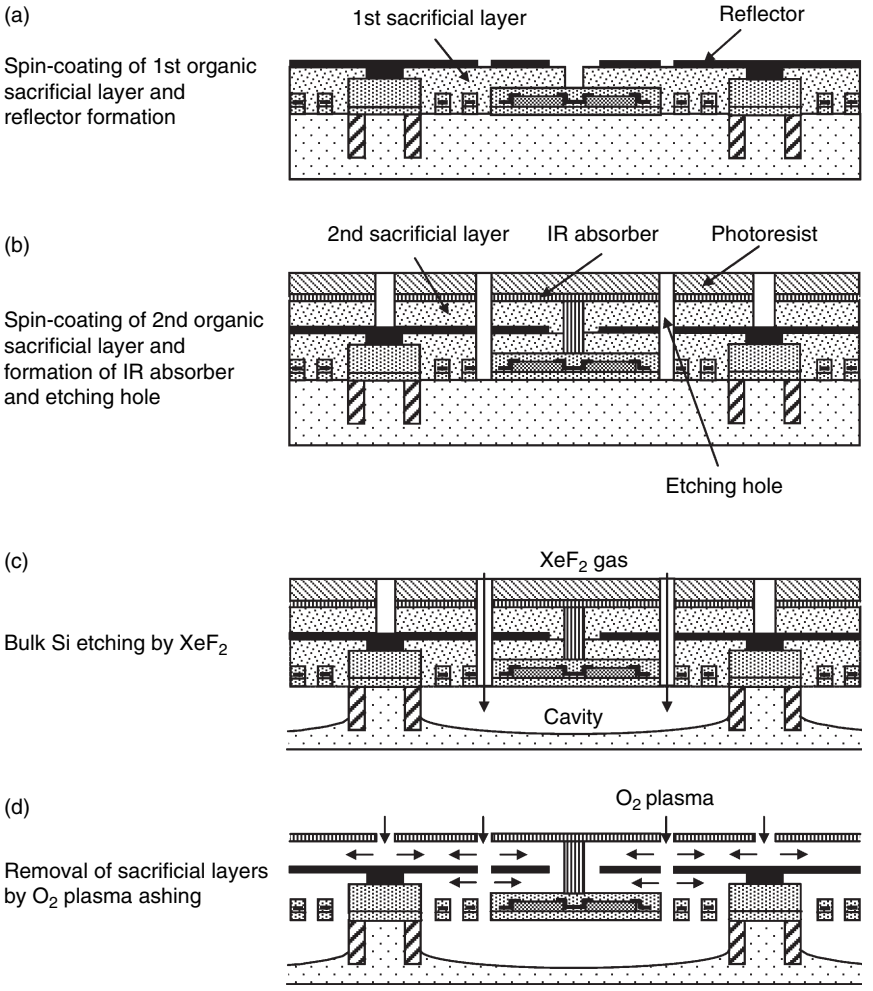


Figure 46 Fabrication process for multilevel silicon-on-insulator (SOI) diode uncooled infrared focal plane array (IRFPA) pixel. (Source: Kimata M, Ueno M, Takeda M, Seto T 2006 SOI diode uncooled infrared focal plane arrays. *Proc. SPIE* **6127**, 61270X-1–11.)

hardened by baking. A very flat surface is obtained with spin coating if we choose proper materials and suitable process conditions. A part of the sacrificial layer is removed in order to form anchors of the reflector to the substrate structure. **Figure 46(a)** shows the structure just after patterning of the metal reflector on the first organic sacrificial layer. The second sacrificial layer is then spin-coated, and the IR absorber is formed on it. Following IR absorber formation, etching holes for the bulk silicon etching are opened through the IR absorber and two sacrificial layers (**Figure 46(b)**). After that, bulk silicon etching by XeF_2 is performed (**Figure 46(c)**). XeF_2 etching is a dry etching process in which the etching proceeds isotropically. Therefore, a deep trench-etching stopper surrounding the pixel cavity is needed to avoid the connection between neighboring cavities. In this case the buried oxide of the SOI substrate plays the role of etching stopper to protect the SOI diodes during the XeF_2 etching process. Finally, as shown in **Figure 46(d)**, all organic layers, including the two sacrificial layers and the photoresist for etched-hole formation, are removed by the final O_2 plasma ashing process. It is important to note that the XeF_2 etching process is not the final process, but is done before the sacrificial layer etching. By this process flow, XeF_2 gas enters from only the central area of the cavity, enlarging the etching margin.

Figure 47 shows a SEM photograph of a type II SOI diode pixel array fabricated by the single sacrificial layer process, in which both surface and cross-sectional structures can be observed. The pixel size is 40 μm square, and the IR absorber covers 90% of the pixel area. A 25-μm-square pixel has successfully been fabricated with the double sacrificial layer process. An example of the lower part of the three-level structure is shown in **Figure 48(a)**. Two long meandering support legs effectively reduce the thermal conductance down to $1 \times 10^{-8}\,\mathrm{W\,K^{-1}}$. **Figure 48(b)** shows a close-up photograph near the IR absorber pillar.

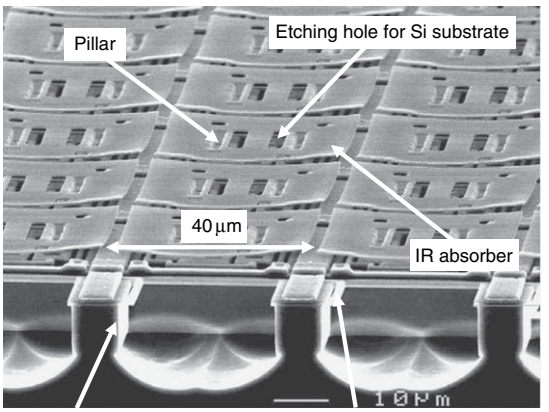


Figure 47 Scanning electron microscope (SEM) photograph of double-level silicon-on-insulator (SOI) diode uncooled infrared focal plane array (IRFPA) pixels, showing cross-sectional structure and infrared absorber covering pixel. (Source: Kimata M, Ueno M, Takeda M, Seto T 2006 SOI diode uncooled infrared focal plane arrays. *Proc. SPIE* **6127**, 61270X-1–11.)

Figure 49 shows a block diagram of the SOI diode uncooled IRFPA. The temperature-sensing diodes are active in a forward-biased condition. This design eliminates the selection switch in the pixel because unselected diodes are reverse-biased and are automatically cut off from the signal line. Each column has a gate modulation integrator for signal amplification and bandwidth restriction. The output from a pixel connects to the gate of an integration MOS transistor, which discharges the charge at an integration capacitor. The pixel diodes are biased row by row during a horizontal period. One integration time starts from a horizontal blanking period and ends at the next blanking period. After integration during one horizontal period, the voltages at the integration capacitors are transferred to sample-and-hold capacitors and are read out through an on-chip amplifier. Since the circuits for readout operation need a relatively high voltage, they are fabricated not on the SOI layer but on the bulk silicon.

Several SOI diode uncooled IRFPAs listed in **Table 2** have been reported. While the first two 320×240-element IRFPAs with 40-μm and 28-μm-square pixels employ the type II pixel structure, the type III pixel structure is applied to the IRFPAs with 25-μm-square pixels. Despite pixel size reduction, the sensitivity has been drastically improved in two IRFPAs with 25-μm-square pixels. The NETD achieved with 25-μm-square SOI diode pixels is better than 50 mK with $f/1.0$ optics (Ueno *et al.* 2005). The sensitivity improvement became possible mainly due to the reduction in the thermal conductance. Examples of thermal images taken with a 25-μm 640×480 SOI diode uncooled IRFPA are shown in **Figure 50**.

The other uncooled IRFPA with single-crystal silicon diodes was developed using the electrochemical etch-stop process technology (Eminoglu *et al.* 2002). While the SOI diode uncooled IRFPA requires some dedicated processes in the pre-MEMS process, the pixel of this uncooled IRFPA is realized

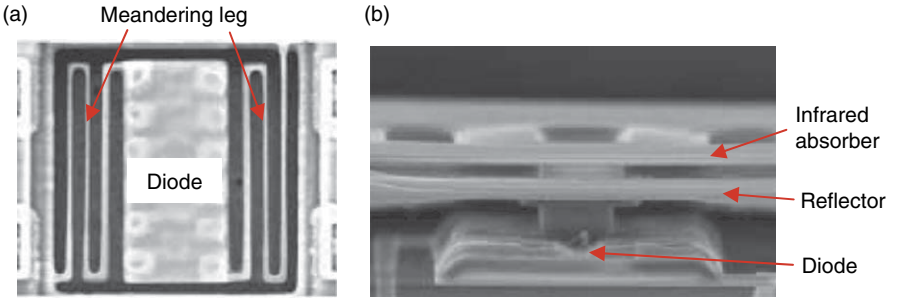


Figure 48 Scanning electron microscope (SEM) photographs of three-level silicon-on-insulator (SOI) diode structure. (a) shows the lower level having two meandering support legs and a diode temperature sensor; and (b) shows the three-level structure near the infrared absorber pillar. (Source: Kimata M, Ueno M, Takeda M, Seto T 2006 SOI diode uncooled infrared focal plane arrays. *Proc. SPIE* **6127**, 61270X-1–11.)

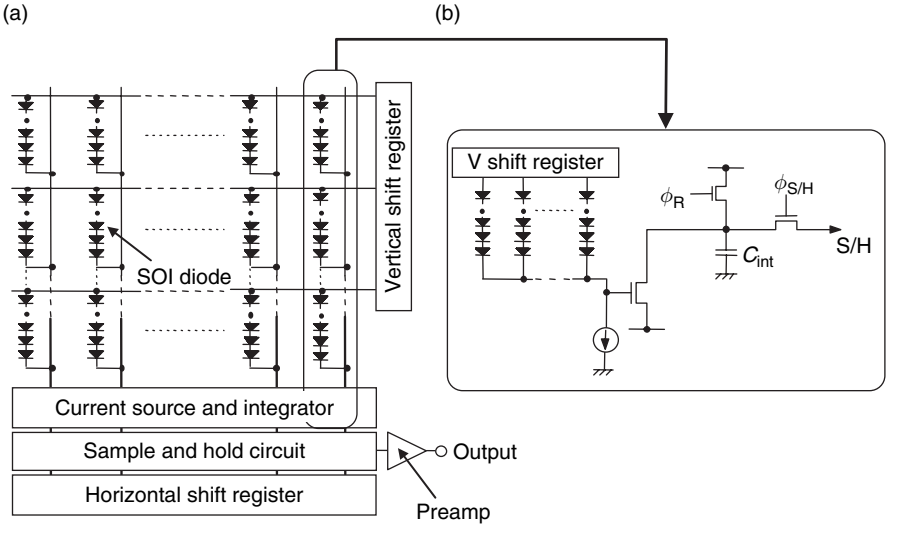


Figure 49 Readout circuit for silicon-on-insulator (SOI) diode uncooled infrared focal plane array (IRFPA). (Source: Kimata M, Ueno M, Takeda M, Seto T 2006 SOI diode uncooled infrared focal plane arrays. *Proc. SPIE* **6127**, 61270X-1–11.)

Table 2 Specifications and performances of SOI diode uncooled IRFPAs

Array size	320×240	320×240	320×240	640×480
Pixel size	$40\,\mu m \times 40\,\mu m$	$28\,\mu m \times 28\,\mu m$	$25\,\mu m \times 25\,\mu m$	$25\,\mu m \times 25\,\mu m$
Chip size	$17.0\,mm \times 17.0\,mm$	$13.5\,mm \times 13.0\,mm$	$12.5\,mm \times 13.5\,mm$	$20.0\,mm \times 19.0\,mm$
Number of diodes	8	6	6	6
Thermal conductance	$1.1 \times 10^{-7}\,W\,K^{-1}$	$4.0 \times 10^{-8}\,W\,K^{-1}$	$1.6 \times 10^{-8}\,W\,K^{-1}$	$1.6 \times 10^{-8}\,W\,K^{-1}$
Sensitivity	$930\,\mu V\,K^{-1}$	$801\,\mu V\,K^{-1}$	$2842\,\mu V\,K^{-1}$	$2064\,\mu V\,K^{-1}$
Noise	$110\,\mu V_{rms}$	$70\,\mu V_{rms}$	$102\,\mu V_{rms}$	$83\,\mu V_{rms}$
Nonuniformity	1.46%	1.25%	1.45%	0.90%
NETD (at *f*/1.0)	110 mK	87 mK	36 mK	40 mK

Source: Kimata M, Ueno M, Yutani N, Iwade S, Tsubouchi N 1987 A 512×512 element PtSi Schottky-barrier infrared image sensor. *IEEE JSSC* SC-22, 1124-9.

by a standard CMOS process and post-CMOS etching steps. A maskless post-CMOS process becomes possible by making required etch openings to reach the silicon substrate with dry etching process while CMOS metal layers are used as protection mask. A 128×128-element uncooled IRFPA with 40-μm-square pixels was reported with the CMOS-compatible single-crystal diode technology. Using a

Figure 50 Examples of thermal images with 640 × 480 silicon-on-insulator (SOI) diode uncooled infrared focal plane array (IRFPA). (Source: Kimata M, Ueno M, Takeda M, Seto T 2006 SOI diode uncooled infrared focal plane arrays. *Proc. SPIE* **6127**, 61270X-1–11.)

0.35-μm CMOS foundry process, a fill factor of 44% and thermal conductance of 1.8×10^{-7} W K^{-1} were achieved (Eminoglu *et al.* 2003). As a single-diode pixel design yields a relatively low sensitivity, an NETD of 1 K is expected for the current design of these IRFPAs.

Murguia *et al.* (1999) proposed a Schottky-barrier diode bolometer named the thermoionic thermal detector. The sensing mechanism of their device is the modulation of Schottky diode dark current (the reverse current) with temperature. The operating temperature depends on the barrier height, and thus this detector needs some cooling if a Schottky diode having lower barrier height is used. Although there are no operating thermoionic thermal detector devices reported so far, a simulation shows that the thermoionic thermal detectors optimized for operation at room temperature have a 6% per Kelvin temperature coefficient and have a potential of achieving an NETD of 6 mK (Murguia *et al.* 1999).

3.04.8 Bimaterial and Thermo-Optical IRFPAs

All the thermal detectors discussed above are based on operation mechanisms that convert the change in temperature to some electrical outputs. Old technology, however, teaches us that we can use mechanisms in which some mechanical displacements occur due to the temperature change. Microcantilever structures constructed from materials exhibiting dissimilar thermal expansion properties can be used to detect visible and IR radiation. A feasibility study of this concept was performed by Oden *et al.* (1996). They made use of commercially available piezolever for electrical deflection monitoring and atomic force microscope (AFM) probe tips for optical deflection monitoring, and found that the former microcantilever has an NETD of 90 mK and the latter has a D^* of 3.6×10^7 cm Hz$^{1/2}$ W^{-1}.

Amantea *et al.* (1997) and Wang *et al.* (2006) extended the microcantilever optical detector concept and proposed a bimaterial uncooled IRFPA in which mechanical displacement of a freestanding structure due to temperature change is detected as capacitance change. **Figure 51** explains the operation

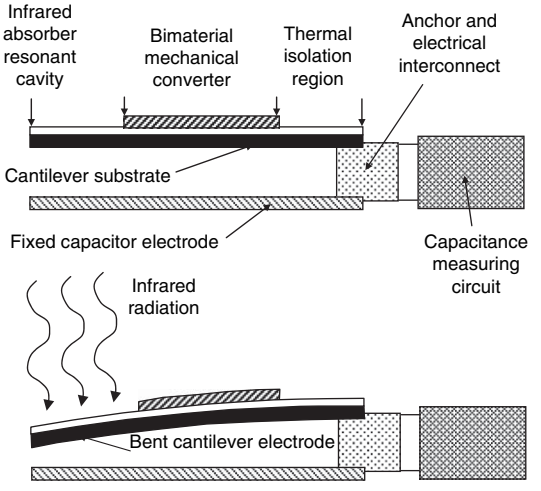


Figure 51 Operation of bimaterial uncooled infrared focal plane array (IRFPA) with electrical readout. (Source: Hunter S R, Amante R A, Goodman L A, Kharas D B, Gershtein S, Matey J R, Perna S N, Yu Y, Maley N, White L K 2003 High sensitivity uncooled microcantilever infrared imaging arrays. *Proc. SPIE* **5074**, 469–80.)

mechanism of the bimaterial thermal IR detector with capacitive readout (Hunter *et al.* 2003). The detector consists of three elements: the IR-absorbing element, the bimaterial element, and the thermal isolation support element. The bimaterial element is made of two thin films that have different thermal expansion coefficients. The IR-absorbing element absorbs the IR radiation and coverts it to heat, raising the temperature of the IR-absorbing element and the bimaterial element. Since two layers in the bimaterial element adhere to one another, they must expand by equal amounts, and tensile stress builds up in one film having a smaller thermal expansion coefficient while compressive stress builds up in the other film with a larger thermal expansion coefficient. Thus, the stress gradient across the bimaterial element developed by IR radiation causes the bimaterial element to bend, and as a result, causes the change in the electrical capacity between the cantilever electrode and the fixed electrode. They tried two bimaterial pairs; one is a SiC/Al pair and the other is a SiC/Au pair. In both case, amorphous hydrogenated SiC was used because it has a low thermal expansion coefficient (4×10^{-6} per Kelvin) when compared to Al and Au, and has a lower thermal conductivity ($0.35\ \mathrm{W\,m^{-1}\,K^{-1}}$) when compared to other dielectric materials used in Si-LSI processes.

Figure 52 shows a cross-sectional structure and pattern layout of the pixel that uses the SiC/Al bimaterial structure. A simulation shows that the SiC/Al bimaterial with a thickness of 0.2 μm and length of 50 μm exhibits a mechanical sensitivity of $0.18\ \mu\mathrm{m\,K^{-1}}$, resulting in an electrical (capacitance) sensitivity of 36% per Kelvin if the initial gap is 0.5 μm. Since the electrical sensitivity in this device corresponds to TCR for the resistance bolometer, the typical value of which is a few percent per Kelvin for semiconductor bolometers, the bimaterial uncooled IRFPA is expected to have higher sensitivity than the resistance bolometer uncooled IRFPAs. The NETD of the bimaterial uncooled IRFPA with 50-μm-square pixels was estimated as 5 mK with $f/1.0$ optics. After proposal of the bimaterial uncooled IRFPA, considerable efforts were made to overcome various problems, such as material stress control, anchor formation, sacrificial materials and release process, IR absorber design, vibration caused by a pulse readout operation (Amantea *et al.* 1998). Although the predicted performance was appealing, the best NETD measured with a 320×240-element bimaterial uncooled IRFPA with 50-μm-square pixels was 1.8 K, and the NETD and its distribution depend on the operating temperature (Amantea *et al.* 1997). Despite the disappointing results obtained until now, the development of uncooled IRFPAs based on this bimaterial concept is being continued by another group (Hunter *et al.* 2006).

The bimaterial deflection can also be measured optically. Several optical techniques to measure the individual pixel deflection have been proposed (Datskos *et al.* 2000, Ishizuya *et al.* 2001, Zhao *et al.* 2003). Ishizuya *et al.* (2001) developed a

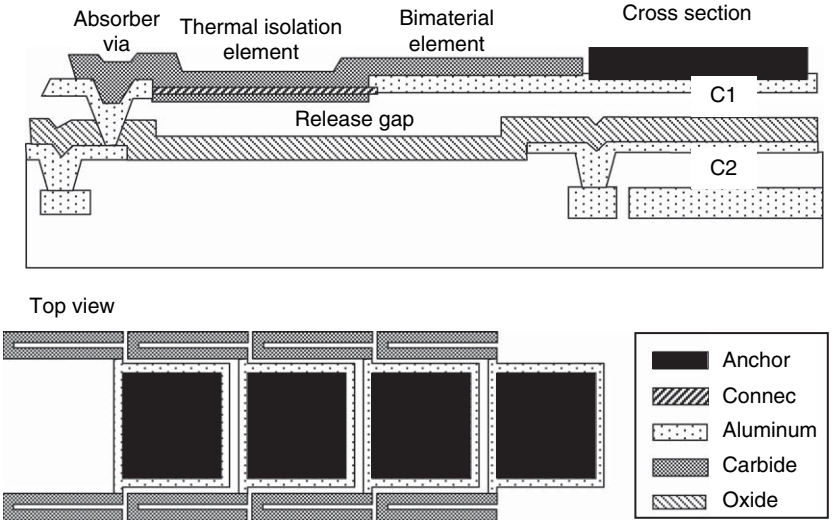


Figure 52 Pixel of bimaterial uncooled infrared focal plane arrays (IRFPAs). (Source: Amantea R, Knoedler C M, Pantuso F P, Patel V K, Sauer D J, Tower J R 1997 An uncooled IR imager with 5 mK NETD. *Proc. SPIE* **3061**, 210–22.)

266 × 194-element optical readout bimaterial uncooled IRFPA with 55-μm-square pixels. The pixel structure and schematic diagram of the optical readout system are depicted in **Figure 53**. The optical readout eliminates the metal interconnection between the freestanding thermometer and the readout circuit on the substrate, and simplifies the pixel structure and manufacturing process. In this IRFPA, IR rays are irradiated from the back of the IRFPA. The backside illumination is possible because silicon is almost transparent in the wavelength range of interest. IR energy absorbed in the IR absorber elevates the temperature of the IR absorber and the bimaterial beam, and deflects the bimaterial beam in the same way as bimaterial uncooled IRFPA with capacitance readout. To measure the deflecting angle of the bimaterial structure, a probe visible light beam is irradiated from the front side. The probe light is reflected back by the bimaterial pixels, passes through a pinhole, and is eventually imaged on a visible CCD. If the IRFPA is exposed to a uniform background, the fraction of reflected light that passes through the pinhole is the same for all the pixels. On the other hand, when the intensity of the incident IR radiation is not uniform across the IRFPA, the reflected light intensity passing through the pinhole becomes different pixel by pixel, reflecting the deflecting angle of the bimaterial structure. Thus, CCD reproduces the IR image falling on the IRFPA as visible light gray-scale image. If the CCD is replaced by a human eye, direct viewing of IR images is possible with this type of IRFPA.

Zhao *et al.* (2003) also developed an optical readout bimaterial uncooled IRFPA. Although the IRFPA has basically the same structure as that developed by Ishizuya and coworkers, they devised a different readout technique. Their readout technique requires a laser probe light source, and relies on the local interference between light scattered by the pixel edge and that scattered by the substrate. They reported that a 300 × 300-element optical readout bimaterial IRFPA with 65-μm-square pixels exhibited an NETD of 200 mK with $f/1.0$ optices and a 10-Hz frame rate.

Optical readout for uncooled IRFPAs without mechanical movement is also possible. Examples of such IRFPAs are devices that use a temperature-sensitive Fabry–Perot tunable optical filter (Flusberg and Deliwala 2006, Wu *et al.* 2005) and an electrooptical crystal operated in the paraelectric phase (Secundo *et al.* 2005). The former (Wu *et al.*

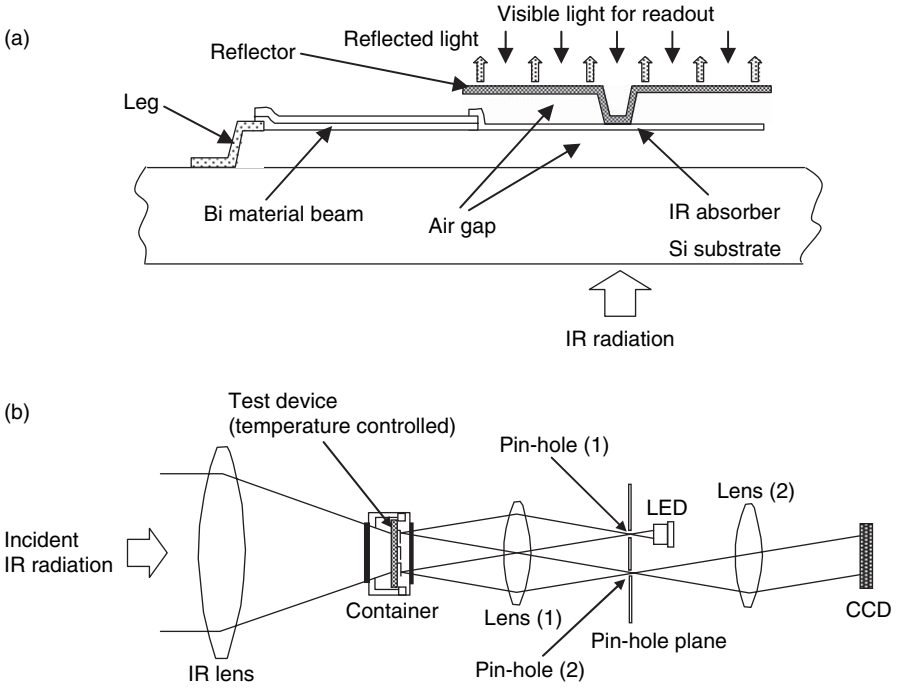


Figure 53 Bimaterial uncooled infrared focal plane array (IRFPA) with optical readout. (a) Shows the pixel structure; and (b) illustrates the readout scheme. (Source: Ishizuya T, Suzuki J, Akagawa K, Kazama T 2001 Optically readable bi-materials infrared detector. *Proc. SPIE* **4369**, 342–9.)

2005) is explained in detail here. Changing the index of refraction of filter materials with temperature gives the narrow band optical filter wavelength tunability. They developed a temperature-sensitive tunable filter using amorphous silicon and silicon nitride, and found that it has a thermo-optical coefficient (the differential temperature dependence of the refraction index) of $2.3 \times 10^{-4}\,\mathrm{K}^{-1}$ at 300 K and a tunability of $0.06\,\mathrm{nm\,K}^{-1}$. The FPA is simply constructed by arranging thermally isolated small filters on a substrate that is transparent in the near-IR wavelength region. **Figure 54** is a schematic illustration of an IR camera system with this optical readout uncooled IRFPA. IR image focused on the IRFPA produces a temperature distribution in the pixel array, depending on the intensity distribution of the IR image. A CCD or CMOS image sensor captures near-IR images through the IRFPA that is generated by a near-IR probe light introduced by a beam. The sensitivity is maximized when the emission spectrum of the probe light is narrower than the filter transmission spectrum and its peak wavelength is located at the wavelength at which the transmission spectrum has the largest gradient. A hexagonal-arranged 160×120-element IRFPA was developed with 50-μm pitch tunable filters supported by hollow post structures. This IRFPA exhibited an NETD of 280 mK at $f/0.86$ optics and 20-Hz frame rate.

3.04.9 Conclusions

IR imaging started with mechanical scanning systems, in which single-element detectors or small-format linear arrays cover all the field of view by moving the line of sight in X and Y directions. Such systems required IR detectors having a high responsivity and a fast response speed. Therefore, thermal IR detectors did not find their place in mechanical scanning IR imaging systems. Quantum IR detectors were the inevitable choice. Evolution from mechanical scanning systems to electrical scanning systems enabled the full-frame integration operation for IRFPAs, and thus alleviated the requirements for responsivity and response speed. This shift shed light on quantum IR detectors with low quantum efficiencies, such as Si-based and GaAs-based IR detectors, for their capability of large-scale

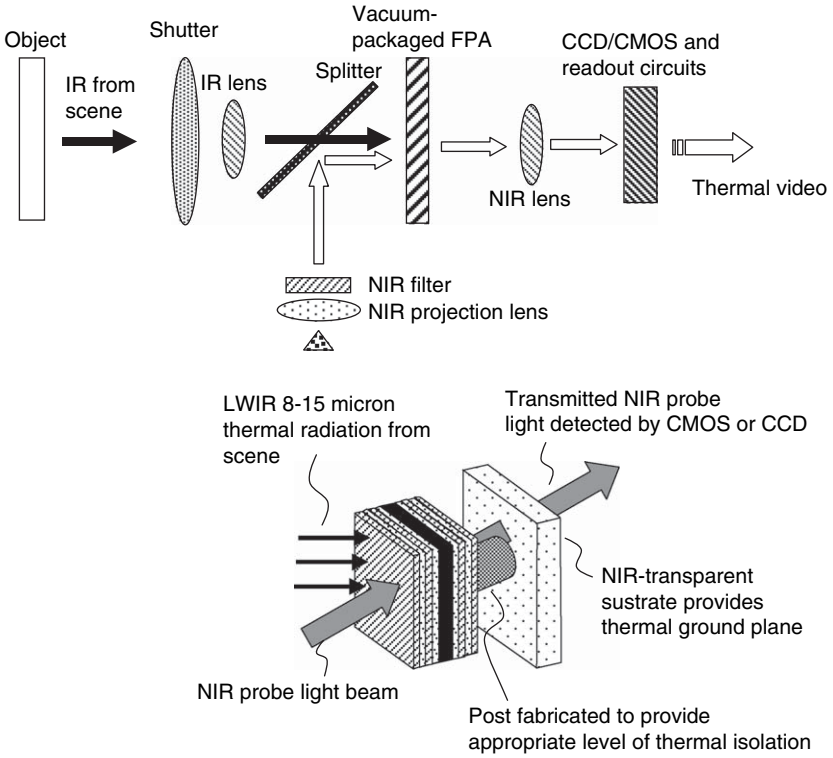


Figure 54 Thermo-optical uncooled infrared focal plane array (IRFPA) based on temperature-sensitive narrowband near-infrared (IR) filter. (Source: Wu M, Cook J, Vito R D, Li J, Ma E, Murano R, Nemchuk N, Tabasky M, Wagner M 2005 Novel low-cost uncooled infrared camera. *Proc. SPIE* **5783**, 496–505.)

integration and high uniformity. It also activated development of uncooled IRFPAs with thermal detectors.

MEMS technology, which made remarkable progress in the 1980s and in the beginning of the 1990s, accelerated performance improvement of the uncooled IRFPAs and made it possible to manufacture microbolometer uncooled IRFPAs with an NETD of <100 mK at $f/1.0$. In the uncooled IRFPAs, the MEMS technology has played an important role to break through the limitation of the thermal conductance in the hybrid structure based on the bump-bonding technology that had been used before the MEMS era. The thermal conductance achievable by the MEMS technology is one to two orders of magnitude smaller than those by the hybrid technology. The first-generation uncooled IRFPAs had pixel sizes around 50 μm square and array sizes from 160 μm × 120 μm to 320 μm × 240 μm. Currently, advanced technology for uncooled IRFPAs has reached a level that achieves NETDs of <50 mK at $f/1.0$ with pixel sizes from 25 μm square to 20 μm square. Such technology has also enabled large-scale integration up to 640 × 480 pixels. Those second-generation uncooled IRFPAs have more complicated pixel structures and/or finer design rules that require more sophisticated MEMS fabrication technology, and importance of MEMS technology is growing.

Although more than 10 years have passed since uncooled IRFPA development became active, various types, including resistance bolometer, pyroelectric, ferroelectric bolometer, thermoelectric, and diode uncooled IRFPAs, are still being used with distinct advantages. The next high-end target seems to be megapixel uncooled IRFPAs with pixels smaller than 20 μm square that have NETDs of around 10 mK at $f/1.0$. On the other hand, low-end markets need uncooled IR cameras whose prices are less than $1000. It is difficult to forecast the future of uncooled IRFPAs, but several technologies will survive in response to diverse demands from various application sectors.

References

Almasri M, Butler D P, Celik-Butler Z 2000 Semiconducting YBCO bolometers for uncooled IR detection. *Proc. SPIE* **4028**, 17–26

Altman M, Backer B, Kohin M, Blackwell R, Butler N, Cullen J 1999 Lockheed Martin's 640 × 480 uncooled microbolometer camera. *Proc. SPIE* **3698**, 137–43

Amantea R, Knoedler C M, Pantuso F P, Patel V K, Sauer D J, Tower J R 1997 An uncooled IR imager with 5 mK NETD. *Proc. SPIE* **3061**, 210–22

Amantea R, Goodman L A, Pantuso F, Sauer D J, Varghese M, Villani T S, White L K 1998 Progress towards an uncooled IR imager with 5 mK NETD. *Proc. SPIE* **3436**, 647–59

Asahi R, Tabata O, Suzuki F, Sugiyama S, Suzuki M, Tanaka A 1992 An infrared imaging sensor using poly-silicon p–n junction diodes. *Tech. Dig. 11th Sensor Symposium* Tokyo, Japan. pp. 99–102.

Astier A, Arnaud A, Ouvrier-Buffet J-L, Yon J-J, Motin E 2004 Advanced packaging developed for very low cost uncooled IRFPA. *Proc. SPIE* **5406**, 412–21

Belcher J F, Hanson C M, Beratan H R, Udayakumar K R, Soch K L 1998 Uncooled monolithic ferroelectric IRFPA technology. *Proc. SPIE* **3436**, 611–22

Beratan H, Hanson C, Meissner E G 1994 Low-cost uncooled ferroelectric detector. *Proc. SPIE* **2274**, 147–56

Boyle W S, Smith G E 1970 Charge-coupled semiconductor devices. *Bell Syst. Tech. J.* **49**, 587–93

Brady J, Schimert T, Ratcliff D, Gooch R, Ritchey B, McCardel P, Rachels K, Ropson S, Wand M, Weinstein W, Wynn J 1999 Advances in amorphous silicon uncooled IR systems. *Proc. SPIE* **3698**, 161–7

Capper P, Elliott C T 2001 *Infrared Detectors and Emitters: Materials and Devices*. Kluwer Academic Publishers, Norwell, MA

Celik-Butler Z, Butler D P, Yildiz A 2002 Room-temperature YBaCuO infrared detectors on a flexible substrate. *Proc. SPIE* **4721**, 260–8

Choi I H, Wise K D 1986 A silicon-thermopile-based infrared sensing array for use in automated manufacturing. *IEEE Trans. Electron Devices* **ED-32**, 72–9

Cole B E, Higashi R E, Ridely J A, Wood R A 2001 Integrated vacuum packaging for low-cost light-weight uncooled microbolometer arrays. *Proc. SPIE* **4369**, 235–9

Datskos P G, Rajic S, Senesac L R, Earl D D, Evans B M, Corbeil J L, Datskou I 2000 Optical readout of uncooled thermal detectors. *Proc. SPIE* **4230**, 185–97

Dehe A, Fricke K, Hartnagel H L 1995 Infrared thermopile sensor based on AlGaAs–GaAs micromachining. *Sens. Actuators A* **46–47**, 432–6

Dehe A, Pavlidis D, Hong K, Hartnagel H L 1997 InGaAs/InP thermoelectric infrared sensor utilizing surface bulk micromachining technology. *IEEE Trans. Electron Devices* **44**, 1052–9

Dereniak E L, Crowe D G 1984 *Optical Radiation Detectors*. John Wiley and Sons, New York

Eminoglu S, Tezcan D S, Akin T 2001 A CMOS *n*-well microbolometer FPA with temperature coefficient enhancement circuitry. *Proc. SPIE* **4369**, 240–9

Eminoglu S, Tanrikulu M Y, Tezcan D S, Akin T 2002 A low-cost small pixel uncooled infrared detector for large focal plane arrays using a standard CMOS process. *Proc. SPIE* **4721**, 111–21

Eminoglu S, Tanrikulu M Y, Akin T 2003 Low-cost uncooled infrared detector arrays in standard CMOS. *Proc. SPIE* **5074**, 425–36

Flusberg A, Deliwala S 2006 Highly sensitive infrared imager with direct optical readout. *Proc. SPIE* **6206,** 62061E1–E18, Orlanda, FL, USA

Foote M C, Gaalema S 2001 Progress towards high-performance thermopile imaging arrays. *Proc. SPIE* **4369**, 350–5

Foote M C, Jones E W 1998 High performance micromachined thermopile linear arrays. *Proc. SPIE* **3379**, 192–7

Foote M C, Jones E W, Caillat T 1998 Uncooled thermopile infrared detector linear arrays with detectivity greater than 10^9 cmHz$^{1/2}$/W. *IEEE Trans. Electron Devices* **45**, 1896–902

Francisco G L 2001 Amorphous silicon bolometer for fire/rescue. *Proc. SPIE* **4360**, 138–48

Fujitsuka N, Sakata J, Miyachi Y, Mizuno K, Ohtsuka K, Taga Y, Tabata O 1997 Monolithic pyroelectric infrared image sensor using PVDF thin film. *Proc. Int. Conf. Solid-State Sensors and Actuators (Transducers),* Chicago, IL, USA, pp. 1237–40

Han C J, Rawlings R, Sweeney M, Whicker S, Peysha D, Clarke J E, Sullivan B, Li C, Howard P 2005 320 × 240 and 640 × 480 UFPAs for TWS and DVE applications. *Proc. SPIE* **5783**, 559–65

Hanson C 1993 Uncooled thermal imaging at Texas Instruments. *Proc. SPIE* **2020**, 330–9

Hanson C M, Beratan H R 2002 Thin-film ferroelectrics: Breakthrough. *Proc. SPIE* **4721**, 91–9

Hanson C, Beratan H, Owen R 1992 Uncooled thermal imaging at Texas Instruments. *Proc. SPIE* **1735**, 17–26

Hanson C M, Beratan H R, Belcher J F 2001 Uncooled infrared imaging using thin-film ferroelectrics. *Proc. SPIE* **4288**, 298–303

Hay K A, Deusen D V 2005 Uncooled focal plane array detector development at Infrared Vision Technology Corporation. *Proc. SPIE* **5783**, 514–523

Henini M, Razeghi M 2002 *Handbook of Infrared Detection Technologies*. Elsevier Science Ltd, Oxford

Herring R, Howard P E 1996 Design and performance of the ULTRA 320 × 240 uncooled focal plane array and sensor. *Proc. SPIE* **2746**, 2–12

Hirota M, Morita S 1998 Infrared sensor with precisely patterned Au-black absorption layer. *Proc. SPIE* **3436**, 623–35

Hirota M, Satou F, Saito M, Kishi Y, Nakajima Y, Uchiyama M 2001 Thermoelectric infrared imager and automotive applications. *Proc. SPIE* **4369**, 312–21

Hirota M, Nakajima Y, Saito M, Satou F, Uchiyama M 2003 120 × 90 element thermopile array fabricated with CMOS technology. *Proc. SPIE* **4820**, 239–429

Howard P E, Clarke J E 1999 Advanced high-performance 320 × 240 VO_x microbolometer uncooled IR focal plane. *Proc. SPIE* **3698**, 131–7

Hudson R D 1969 *Infrared System Engineering*. John Wiley and Sons, New York

Hunter S R, Amante R A, Goodman L A, Kharas D B, Gershtein S, Matey J R, Perna S N, Yu Y, Maley N, White L K 2003 High sensitivity uncooled microcantilever infrared imaging arrays. *Proc. SPIE* **5074**, 469–80

Hunter S R, Mauer G S, Simelgor G, Jiang L 2006 High-sensitivity uncooled microcantilever infrared imaging arrays. *Proc. SPIE* **6206,** 62061J-1-1–12, Orlando, FL, USA

Ishikawa T, Ueno M, Endo K, Nakaki Y, Hata H, Sone T, Kimata M, Ozeki T 1999 Low-cost 320 × 240 uncooled IRFPA using conventional silicon IC process. *Proc. SPIE* **3698**, 556–64

Ishikawa T, Ueno M, Nakaki Y, Endo K, Ohta Y, Nakanishi J, Kosasayama Y, Yagi H, Sone T, Kimata M 2000 Performance of 320 × 240 uncooled IRFPA with SOI diode detectors. *Proc. SPIE* **4130**, 152–9

Ishizuya T, Suzuki J, Akagawa K, Kazama T 2001 Optically readable bi-materials infrared detector. *Proc. SPIE* **4369**, 342–9

Jahanzeb A, Traverse C M, Celik-Butler Z, Butler D P, Tan S G 1997 A semiconductor YBaCuO microbolometer for room temperature IR imaging. *IEEE Trans. Electron Devices* **44**, 1795–801

Jerominek H, Pope T D, Alain C, Zhang R, Lehoux M, Picard F, Fuchs R W, Grenier C, Rouleau Y, Cayer F, Savard S, Bilodeau G, Couillard J-F, Larouche C, Phong L N 1998 128 × 128 pixel uncooled bolometric FPA for IR detection and imaging. *Proc. SPIE* **3436**, 585–92

Jerominek H, Pope T D, Alain C, Zhang A, Picard F, Lehoux M, Cayer F, Savard S, Larouche C, Crenier C 2000 Miniature VO_2-based bolometric detectors for high-performance uncooled FPAs. *Proc. SPIE* **4028**, 47–56

Kanno T, Saga M, Matsumoto S, Uchida M, Tsukamoto N, Tanaka A, Itoh S, Nakazato A, Endoh T, Tohyama S, Yamamoto Y, Murashima S, Fujimoto N, Teranishi N 1994 Uncooled infrared focal plane array having 128 × 128 thermopile detector elements. *Proc. SPIE* **2269**, 450–9

Kimata M, Denda M, Yutani N, Iwade S, Tsubouchi N 1987 A 512 × 512 element PtSi Schottky-barrier infrared image sensor. *IEEE JSSC* **SC-22**, 1124–9

Kimata M, Ueno M, Takeda M, Seto T 2006 SOI diode uncooled infrared focal plane arrays. *Proc. SPIE* **6127**, 61270X-1–11

Kosasayama Y, Sugino T, Nakaki Y, Fujii Y, Inoue H, Yagi H, Hata H, Ueno M, Takeda M, Kimata M 2004 Pixel scaling for SOI diode uncooled infrared focal plane arrays. *Proc. SPIE* **5406**, 504–11

Kruse P W 1995 Uncooled IR focal plane array. *Proc. SPIE* **2552**, 556–63

Kruse P W 2001 *Uncooled Thermal Imaging Arrays, Systems and Applications*. SPIE, Bellingham, MA

Kruse P W, Skatrud D D 1997 *Uncooled Infrared Imaging Arrays and Systems*. Academic Press, San Diego, CA

Kruse P W, McGlauchlin L D, McQuistan R B 1962 *Elements of Infrared Technology: Generation, Transmission, and Detection*. John Wiley and Sons, New York

Lahiji G R, Wise K D 1982 A batch-fabricated silicon thermopile infrared detector. *IEEE Trans. Electron. Devices* **ED-29**, 14–22

Lawson W D, Nielson S, Putley E H, Young A S 1959 Preparation and properties of HgTe and mixed crystals of HgTe–CdTe. *J. Phys. Chem. Solids* **9**, 325–9

Lee H-K, Yoon J-B, Yoon E, Ju S-B, Yong Y-J, Lee W, Kim S-G 1998 A high fill-factor IR bolometer using multi-level electrothermal structures. *Tech. Dig. IEEE Int. Electron Device Meeting,* San Francisco, CA, USA, 436–66.

Lenggenhanger R, Baltes H, Peer J, Forster M 1992 Thermoelectric infrared sensors by CMOS technology. *IEEE Electron. Device Lett.* **13**, 454–6

Lenggenhager R, Baltes H, Elbel T 1993 Thermoelectric infrared sensors in CMOS technology. *Sens. Actuators A* **37–38**, 216–20

Leonov V N, Greten Y, Moor P D, Bois B D, Goessens C, Grietens B, Merken P, Perova N A, Puttens G, Hoof C V, Verbist A, Veermeiren J 2003 Small two-dimensional and linear arrays of polycrystalline SiGe microbolometers at IMEC-XenICs. *Proc. SPIE* **5074**, 446–57

Liddiard K C 1983 Thin-film resistance bolometer IR detectors. *Infrared Phys.* **24**, 57–64

Liddiard K.C 1993 Application of interferrometric enhancement to self-absorbing thin film thermal IR detectors. *Infrared Phys.* **34**, 379–84

Lim H C, Tsao S, Taguchi M, Zhang W, Quivy A A, Razeghi M 2006 InGaAs/InGaP quantum-dot infrared photodetectors with a high detectivity. *Proc. SPIE* **6127**, 61270N-1–6

Manning P A, Gillham J P, Parkinson N J, Kaushal T P 2004 Silicon foundry micro-bolometers – The route to the mass-market thermal imager. *Proc. SPIE* **5406**, 465–72

Marshall C, Butler N, Blackwell R, Murphy R, Breen T 1996 Uncooled infrared sensor with digital focal plane array. *Proc. SPIE* **2746**, 23–31

McEwen R K, Manning P A 1999 European uncooled thermal imaging sensors. *Proc. SPIE* **3698**, 322–37

Mizrahi U, Fraenkel A, Bykov L, Giladi A, Adin A, Ilan E, Shiloah N, Malkinson E, Zabar Y, Seter D, Nakash R, Kopolovich Z 2005 Uncooled detector development program at SCD. *Proc. SPIE* **5783**, 551–8

Moor D P, John J, Sedky S, Hoof C V 2000 Linear arrays of fast uncooled poly SiGe microbolometers for IR detection. *Proc. SPIE* **4028**, 27–34

Mottin E, Martin J-L, Ouvrrier-buffet J-L, Vilain M, Bain A, Yon J-J, Tissot J-L, Chatard J-P 2001 Enhanced amorphous silicon technology for 320 × 240 microbolometer arrays with a pitch of 35 μm. *Proc. SPIE* **4369**, 250–256

Mottin E, Bain A, Martin J L, Ouvrier-Buffet J L, Yon J J, Chatard J P, Tissot J L 2002 Uncooled amorphous silicon technology: High performance achievement and feature trends. *Proc. SPIE* **4721**, 56–63

Münch U, Jaeggi D, Schneeberger K, Schaufelberbühl, Paul O, Baltes H, Jasper J 1997 Industrial fabrication technology for CMOS infrared sensor arrays. *Proc. Transducers '97,* Chicago, IL, USA, pp. 205–8

Murguia J E, Tedrow P K, Shepherd F D, Leahy D, Weeks M M 1999 Performance analysis of a thermoionic thermal detector at 400 K, 300 K, and 200 K. *Proc. SPIE* **3698**, 361–75

Murphy D, Ray M, Wyles R, Asbrock J, Lum N, Wyles J, Hewitt C, Kennedy A, Lue D V, Anderson J, Bradley D, Chin R, Kostzewa T 2002 High sensitivity 25 μm microbolometer FPAs. *Proc. SPIE* **4721**, 99–110

Murphy D, Kennedy A, Ray M, Wyles J, Asbrock J, Hewitt C, Lue D V, Sessler T, Anderson J, Bradley D, Chin R, Gonzalez H, Pere C L, Kostrzewa T 2003 Resolution and sensitivity improvements for VO_x microbolometer FPAs. *Proc. SPIE* **5074**, 402–13

Murphy D, Ray M, Kennedy A, Wyles J, Hewit C, Wyles R, Gordon E, Sessler T, Baur S, Lue D V, Anderson S, Chin R, Gonzalez H, Pere C L, Ton S, Kostrzewa T 2005 Expand applications for high performance VO_x microbolometer FPAs. *Proc. SPIE* **5783**, 448–59

Murphy D, Ray M, Kennedy A, Wyles J, Hewitt C, Gordon E, Sessler T, Baur S, Lue D V, Anderson S, Chin R, Gonzalez H, Pere C L, Ton S 2006 High sensitivity 640 × 512 (20 μm pitch) microbolometer FPAs. Proc. SPIE **6206**, 62061A-1-1–14, Orlanda, FL, USA

Nakaki Y, Hata H, Yagi H, Inoue H, Sugino T, Ueno M, Takeda M, Kimata M 2003 Dry micromachining process for uncooled IR FPA with SOI diode detectors. *SENSOR 2003 Proc.* Nuremburg, Germany, pp. 179–184

Noble P J W 1968 Self-scanned silicon image detector arrays. *IEEE Trans. Electron Devices* **ED-15**, 202–9

Norton P W, Kohin M 2005 Technology and applications advancements of uncooled imagers. *Proc. SPIE* **5783**, 524–30

Norton P W, Cox S, Murphy B, Grealish K, Joswick M, Denley B, Feda F, Elmali L, Kohin M 2006 Uncooled thermal imaging sensor and application advances. *To be Published in Proc. SPIE* **6206**, 620617-1–7, Orlando, FL, USA

Oden P I, Wachter E A, Datskos P G, Thundat T, Warmack R J 1996 Optical and infrared detection using microcantilevers. *Proc. SPIE* **2744**, 345–54

Oliver A D, Wise K D 1999 A 1024-element bulk-micromachined thermopile infrared imaging array. *Sens. Actuators* **73**, 222–31

Owen R, Frank S, Daz C 1992 Producibility of uncooled IRFPA detectors. *Proc. SPIE* **1683**, 74–80

Radford W, Murphy D, Ray M, Propst S, Kennedy A, Kojiro J, Woolaway J, Soch K, Coda R, Lung G, Moody E, Gleichman D, Baur S 1996 320 × 240 silicon microbolometer uncooled IRFPAs with on-chip offset correction. *Proc. SPIE* **2746**, 82–92

Radford W, Murphy D, Finch A, Hay K, Kennedy A, Ray M, Sayed A, Wyles J, Varesi J, Moody E, Cheung F 1999 Sensitivity improvements in uncooled microbolometer FPAs. *Proc. SPIE* **3698**, 119–30

Rai-Choudhury P 2000 *MEMS and MOEMS Technology and Applications*. SPIE, Bellingham, MA, USA

Rana M, Butler D P 2005 Amorphous Ge_xSi_{1-x} and $Ge_xSi_{1-x}O_y$ thin films for uncooled microbolometers. *Proc. SPIE* **5783**, 597–606

Reay R, Klaassen E H, Kovacs G T A 1994 Thermally and electrically isolated single crystal silicon structures in CMOS technology. *IEEE Electron Device Lett.* **15**, 399–401

Rogalski A 1995 *Infrared Photon Detectors*. SPIE, Bellingham, MA, USA

Schaufelbüehl A, Münch U, Menolfi C, Brand O, Paul O, Huang Q, Baltes H 2001 256-pixel CMOS-integrated thermoelectric infrared sensor array. *Proc. 14th IEEE Int. Conf. Micro Electro Mechanical Systems,* Interlaken, Switzerland, pp. 200–3

Schaufelbühl A, Scheeberger N, Münch U, Waelti M, Paul O, Brand O, Baltes H, Menolfi C, Huang Q, Doering E, Loepfe M 2001 Uncooled low-cost thermal imager based on micromachined CMOS integrated sensor array. *IEEE J. MEMS* **10**, 503–510

Schneeberger N, Paul O, Baltes H 1995 Optimization structured absorbers for CMOS infrared detectors. *Proc. Transducers '95 and Eurosensors IX,* Stockholm, Sweden, pp. 25–9

Secundo L, Lubianiker Y, Agranat A J 2005 Uncooled FPA with optical reading: Reaching the theoretical limit. *Proc. SPIE* **5783**, 483–95

Sedky S, Fiorini P, Caymax M, Baert C, Hermans L, Mertens R 1998 Characterization of bolometers based on polycrystalline silicon germanium alloys. *IEEE Eelctron Device Lett.* **19**, 376–378

Suzuki M, Makino K, Tanaka A, Asahi R, Tabata O, Sugiyama S, Takigawa M 1990 An infrared detector using poly-silicon p–n junction diode. *Tech. Dig. 9th Sensor Symposium,* Tokyo, Japan, pp. 71–4

Sze S M 1969 *Physics of Semiconductor Devices*. John Wiley and Sons, New York

Tanaka A, Suzuki M, Asahi R, Tabata O, Sugiyama S 1992 Infrared linear image sensor using a poly-Si pn junction diode array. *Infrared Phys.* **33**, 229–36

Tanaka A, Natsumoto S, Tsukamoto B, Itoh S, Chiba K, Endoh T, Nakazato A, Okuyama K, Kumazawa Y, Hijikawa M, Gotoh H, Tanaka T, Teranishi N 1996 Infrared focal plane array incorporating silicon IC process compatible bolometer. *IEEE Trans. Electron Devices* **43**, 1844–50

Terre B, Cannata B, Franklin P, Gonzalez A, Kurth E, Ly H, Parrish B, Peters K, Romeo T, VanYsseldyk B 2003 Microbolometer development and production at Indigo Systems. *Proc. SPIE* **5074**, 518–26

Tezcan D S, Eminoglu S, Akae O S, Akin T 2001 An uncooled microbolometer infrared focal plane array in standard CMOS. *Proc. SPIE* **4288**, 112–21

Tissot J-L, Rothan F, Vedel C, Vilain M, Yon J-J 1998 LETI/LIR's amorphous silicon uncooled microbolometer development. *Proc. SPIE* **3379**, 139–44

Tissot J-L, Martin J-J, Mottin E, Viain M, Yon J-J, Chatard J P 2000 320 × 240 microbolometer uncooled IRFPA development. *Proc. SPIE* **4130**, 473–9

Tissot J-L, Crastes A, Trouilleau C, Fieque B, Tinnes S 2004 Multipurpose high performance 160 × 120 uncooled IRFPA. *Proc. SPIE* **5406**, 550–6

Tissot J-L, Fieque B, Trouilleau C, Robert R, Crasters A, Minssian C, Legras O 2006 First demonstration of 640 × 480 uncooled amorphous silicon IRFPA with 25 μm pixel-pitch. Proc. SPIE **6206**, 620618-1–4, Orlando, FL, UK

Todd M A, Manning P A, Donohue P P, Brown A G, Watton R 2000 Thin film ferroelectric materials for microbolometer arrays. *Proc. SPIE* **4130**, 128–39

Tohyama S, Miyoshi M, Kurashina S, Ito N, Sasaki T, Ajisawa A, Oda N 2004 New thermal isolation pixel structure for high-resolution uncooled infrared FPAs. *Proc. SPIE* **5406**, 428–36

Ueno M, Kosasayama Y, Sugino T, Nakaki Y, Fujii Y, Inoue H, Kama K, Seto T, Takeda M, Kimata M 2005 640 × 480 pixel

uncooled infrared FPA with SOI diode detectors. *Proc. SPIE* **5783**, 567–577

Unewisse M H, Craig B I, Watson R J, Reinholed O, Liddiard K C 1995 The growth and properties of semiconductor bolometers for infrared detection. *Proc. SPIE* **2554**, 43–54

Vedel C, Martin J-L, Ouvrier-Buffet J L, Tissot J-L, Vilain M, Yon J-J 1999 Amorphous silicon based uncooled microbolometer IRFPA. *Proc SPIE* **3698**, 276–83

Wada H, Nagashima M, Oda N, Sasaki T, Kawahara A, Kanzaki M, Tsuruta Y, Mori T, Matsumoto S, Sima T, Hijikawa M, Tsukamoto N, Gotoh H 1998 Design and performance of 256×256 bolometer-type uncooled infrared detector. *Proc. SPIE* **3379**, 90–100

Wada H, Sone T, Hata H, Nakaki Y, Kaneda O, Ohta Y, Ueno M, Kimata M 2000 YBaCuO uncooled microbolometer IRFPA. *Sens. Mater.* **12**, 315–25

Wang W, Ypadhyay V, Munoz C, Bumgarner J, Edwards O 2006 FEA simulation, design and fabrication of uncooled MEMS capacitive thermal detector for infrared FPA imaging. Proc. SPIE **6206**, 62061L-1–12, Orlando, FL, USA

Watton R, Manning P 1998 Ferroelectrics in uncooled thermal imaging. *Proc. SPIE* **3436**, 541–54

Watton R, Manning P A, Perkins M J C, Gillham J P, Todd M A 1996 Uncooled IR imaging: Hybrid and integrated bolometer arrays. *Proc. SPIE* **2744**, 486–99

Wood R A 1993a High-performance infrared thermal imaging with monolithic silicon focal planes operating at room temperature. *Proc. IEEE IEDM,* Washington, DC, USA, pp. 175–7

Wood R A 1993b Uncooled thermal imaging with monolithic silicon focal planes. *Proc. SPIE* **2020**, 322–9

Wood R A, Han C J, Kruse P W 1992 Integrated uncooled infrared detector imaging array. *Tech. Dig. IEEE Solid-State Sensor and Actuator Workshop,* Hilton Head Island, SC, USA, pp. 132–5

Wu M, Cook J, Vito R D, Li J, Ma E, Murano R, Nemchuk N, Tabasky M, Wagner M 2005 Novel low-cost uncooled infrared camera. *Proc. SPIE* **5783**, 496–505

Xu Y-P, Huang R-S, Rigby A 1993 A silicon-diode-based infrared thermal detector array. *Sens. Actuators A* **37–38**, 226–30

Xu H, Mukaigawa T, Hashimoto K, Kubo R, Kiyomoto T, Zhu H, Noda M, Okuyama M 1999 Si monolithic microbolometers of ferroelectric BST thin film combined with readout FET for uncooled infrared image sensor. *Tech Dig. 10th Int. Conf. Solid-State Sensors and Actuators (Transducers),* Sendai, Japan, pp. 398–401

Xu H, Hashimoto K, Mukaigawa T, Zhu H, Kubo R, Usuki T, Kishihara H, Noda M, Suzuki Y, Okuyama M 2000 Development of Si monolithic $(Ba,Sr)TiO_3$ thin-film ferroelectric microbolometers for uncooled chopperless infrared sensing. *Proc. SPIE* **4130**, 140–51

Yon J J, Astier A, Bisotto S, Chamings G, Durand A, Martin J L, Mottin E, Ouvrier-Buffet J L, Tissot J L 2005 First demonstration of 25 μm pitch uncooled amorphous silicon microbolometer IRFPA at LETI-LIR. *Proc. SPIE* **5783**, 432–40

Yutani N, Yagi H, Kimata M, Nakanishi J, Nagayoshi S, Tsubouchi N 1991 1040×1040 element PtSi Schottky-barrier IR image sensor. *Tech. Digest IEDM,* San Francisco, CA, USA, pp. 175–8

Zhao Y, Choi J, Horowtz R, Majumdar A, Kitching J, Norton P 2003 Characterization and performance of optomechanical uncooled infrared imaging system. *Proc SPIE* **4820**, 164–74

Biography

Masafumi Kimata received his BS and MS degrees in electronic engineering from Nagoya University in 1974 and 1976, respectively, and received his Ph.D. degree in faculty of engineering science from Osaka University in 1992. He joined Mitsubishi Electric Corporation in 1976, where he was involved in the research and development of silicon-based infrared focal plane arrays, including Schottky-barrier cooled infrared focal plane arrays and SOI diode uncooled focal plane arrays. In 2004, he retired from Mistubishi Electric, and presently, he is a professor of Ristumeikai University, where he continues his research on MEMS-based uncooled infrared focal plane arrays. He was awarded the Prime Minister Prize of the Japan National Invention Awards in 1993 for the invention of high-resolution Schottky-barrier infrared focal plane arrays. He has been serving as a program committee member of SPIE's conference on Infrared Technology and Applications since 1992. He is a fellow member of SPIE.

3.05 Integrated Optics

Toshiaki Suhara, Department of Electrical, Electronic, and Information Engineering, Graduate School of Engineering, Osaka University, Suita, Osaka, Japan

Nomenclature

a_{TE}, a_{TM}	Asymmetry parameters	**D**	Electric flux density of optical wave
$A(z)$, $B(z)$	Amplitudes of coupled waves	**E**	Electric field of optical wave
b	Normalized propagation constant	E^e_z	External electric field amplitude
c	Light velocity in vacuum	f	Surface acoustic wave frequency
		$g(z)$	Output field profile function

$h(z)$	Input field profile function	V	Normalized waveguide thickness (normalized frequency)
H	Magnetic field of optical wave	W	Waveguide width
h	Grating groove depth	α_r, $\alpha_q^{(i)}$	Radiation decay factors
$I(g,h)$	Overlap integral	β	Propagation constant
j	Imaginary unit	β	Optical wave vector
k	Wave number of optical wave	γ_a, γ_c	Evanescent factors
K	Wave number of periodic structures	$\delta_{mm'}$	Cronecker's delta
K	Grating vector of periodic structures	$\delta(x)$	Dirac's delta function
L	Device length or coupling (interaction) length	$\delta\lambda$	Wavelength deviation
		Δ	Phase mismatch parameter
L_c	Complete coupling length	$\Delta\beta$	Propagation constant deviation
m	Mode number	ε_0	Dielectric permittivity of vacuum
N	Effective index of refraction (mode index)	$[\varepsilon]$	Dielectric tensor
		$\varepsilon(x,y)$	Relative dielectric permittivity profile
N_I, N_{II}	Effective indexes	θ	Incidence angle
$n(x,y)$	Refractive index profile	θ_B	Bragg angle
n_e	Extraordinary refractive index	η	Coupling efficiency or diffraction efficiency
n_o	Ordinary refractive index		
$P_q^{(i)}$	Power distribution ratio	κ	Coupling coefficient
[p]	Photoelastic tensor	κ_c	Lateral wave number
P_s	Surface acoustic wave power	κ_{SHG}	Nonlinear optic coupling coefficient (SHG coupling coefficient)
q	Order of coupling or diffraction		
Q	Grating thickness parameter	λ	Optical wavelength in vacuum
[r]	Electrooptic (Pockels) tensor	Λ	Period of periodic structures
S	Complex Poynting vector	μ_0	Magnetic permittivity of vacuum
[S]	Strain tensor	Φ_a,Φ_s	Phase shifts associated with the total internal reflection
T	Waveguide of thickness		
[T]	Stress tensor	$\Phi(y,z)$	Phase distribution function
v	Velocity of surface acoustic wave	ω	Angular frequency of optical wave

3.05.1 Introduction

Optical techniques using coherent laser light combined with electronic techniques offer tremendous possibilities for applications in wide-ranging areas of storing, transmitting, processing, and sensing of information. Although optical setups and instruments consisting of conventional optical elements may be used in laboratories, they are not suitable for practical use in industries, offices, and consumers' lives. Compact optical systems can be developed by miniaturizing conventional elements such as lenses, prisms, and mirrors and assembling them to construct a module for a specific application. An example of such systems, called microoptics, is optical disk pickup heads, which are now widely used in CD and DVD systems. It is hoped that further integration of optical components, or realization of optical analog of electronic integrated circuits (ICs) and large-scale integrations (LSIs), would significantly enhance the possibilities of applications.

The concept of integrated optics (IO) was proposed based on such an idea (Miller 1969, Shubert and Harris 1968). IO uses optical waveguides as the fundamental structure to confine coherent optical waves. Various active and passive optical components are realized in the waveguide, and they are integrated to construct a device of higher function for a specific application (Coldren and Corzine 1995, Lifante 2003, Murphy 1999, Nishihara *et al.* 1989, Tamir 1975, 1988). The resultant devices are called IO devices, integrated photonic (IP) devices, optical integrated circuits (optical ICs), and photonic integrated circuits (photonic ICs). Devices consisting of optical and electronic components are called optoelectronic integrated circuits (OEICs). Although these terms are used with selection often being based

on the features to be emphasized, the nomenclature is not necessarily clear. They may be classified under microoptics in a broader sense.

The IO technologies has several advantages and outstanding features:

(1) Properties of guided waves can be used to realize components and devices based on new working principles different from those of bulk devices and to achieve higher performances.
(2) High performances, in terms of a high conversion efficiency, a low driving power (voltage), or a high speed, can be obtained for each component, as optical field is confined in narrow waveguides, and the electric (or other) fields for control are concentrated in a narrow space.
(3) The devices are compact and robust and are of low weight, and they do not require critical alignments, as these devices are rugged solid-state structures.
(4) Advanced functions can be achieved by integrating large number of components or different functions of components.
(5) The devices can be mass-produced by the cost-effective planar lithographic fabrication technologies.

Although it took rather a long time, various IO devices were developed in several areas of applications, and many IO devices have been commercialized and put into practical use. The application of IO covers almost all areas of information technology, i.e., optical signal processing, optical communications, optical memories, optical interconnection, and displays. Another important area is optical sensing. A number of components, devices, modules, and systems have been developed. Practical devices include integrated laser light sources, optical modulators, switches, attenuators, filters, wavelength multi/demultiplexers, and integrated receivers for optical network; IO disk pickup heads; and IO interferometer displacement sensors.

This chapter presents an outline of the IO technologies. After discussing theoretically the fundamental characteristics of optical waveguides and coupling of guided waves, passive and active components in waveguides are presented. Then examples of IO devices and their applications are presented to highlight the IO technology including the recent R&D activities.

3.05.2 Optical Waveguides

Optical waveguides are the fundamental structure for the implementation of IO devices. This section describes the concept and the basic characteristics of the waveguides by using simple models (Marcuse 1974, Snyder and Love 1983). Although many waveguides used for IO are prepared in anisotropic crystals, isotropic materials are assumed for simplicity in the theoretical discussions.

3.05.2.1 Waveguide Structures and Optical Waves

Optical waveguides are structures to confine and guide optical waves by the total internal reflection (TIR) at dielectric interfaces. They are classified into planar waveguides (slab waveguides) (**Figure 1(a)**) and channel waveguides (**Figure 1(b)**). Planar waveguides for confinement of optical field in a guiding layer can easily be fabricated and are used as a simple model for discussing the waveguide characteristics. Channel waveguides for 3D confinement are more suitable for implementation of many IO devices. It is required that the refractive index of the guiding (core) region be higher than that of the surrounding (cladding) region. In ray optics, the waveguiding can simply be interpreted in terms of the zigzag pass of the ray at an angle greater than the critical angle, representing successive TIRs at the core/cladding interfaces (**Figure 2**). However, wave optics analysis is required to clarify the important waveguide characteristics.

The wave optics analysis can be made based on the electromagnetic theory using Maxwell's equations (Marcuse 1974, Nishihara *et al.* 1989, Snyder and Love 1983). We consider optical waves of angular

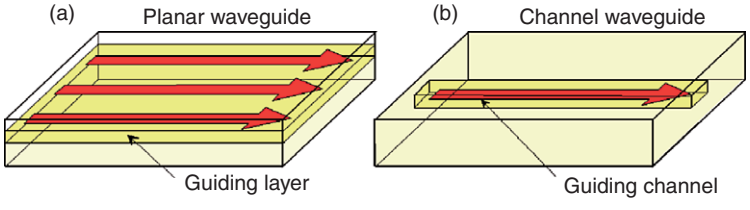


Figure 1 Schematic of optical waveguides.

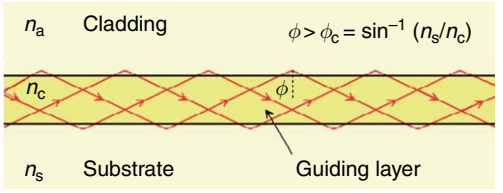


Figure 2 Ray optics presentation of optical waveguiding.

frequency ω. The complex descriptions $\mathbf{E}$ and $\mathbf{H}$ are used for the electric and the magnetic field, respectively, and the time-dependent factor $\exp(j\omega t)$ is omitted.

A lossless waveguide structure, in general, can be described by the distribution of relative permittivity ε over the cross-section. Taking the z-axis of the coordinate system along the waveguide (propagation) axis, the distribution can be written as $\varepsilon = \varepsilon(x,y) = [n(x,y)]^2$, where $n(x,y)$ is the refractive index profile. The fields of optical waves in the structure may be written as

$$\begin{aligned}\mathbf{E}(x,y,z) &= \mathbf{E}(x,y)\exp(-j\beta z),\ \mathbf{H}(x,y,z)\\ &= \mathbf{H}(x,y)\exp(-j\beta z)\end{aligned} \quad [1]$$

where β denotes the propagation constant, and β is often displayed as a wave vector β along the propagation axis. Substituting eqn [1] into Maxwell's equations and eliminating H yield a wave equation:

$$\left[\nabla_t^2 + \left\{k^2 n^2(x,y) - \beta^2\right\}\right]\mathbf{E}(x,y) = 0 \quad [2]$$

where $\nabla_t = (\partial/\partial x, \partial/\partial y, 0)$ is a transverse differential operator, $k = \omega(\varepsilon_0\mu_0)^{1/2} = \omega/c = 2\pi/\lambda$ is the wave number, c the velocity of light, and λ the wavelength in vacuum. The above equation is valid for cases where the medium is homogeneous ($\nabla\varepsilon = 0$) in sections and is valid approximately under an assumption of $\nabla\varepsilon \cong \mathrm{O}$ The propagation constant can be written as

$$\beta = Nk \quad [3]$$

where N is the effective index of refraction or mode index.

Among many solutions of eqn [2] with appropriate boundary conditions are those such that $\mathbf{E}(x, y)$ is large near the waveguide axis and decays to 0 away from it, called guided modes, as they give optical waves that are confined in the waveguide and propagate along the axis. There are a limited number of guided modes, and they are characterized by discrete mode profiles $\mathbf{E}(x, y)$ and $\mathbf{H}(x, y)$ with discrete propagation constants. Solutions such that $\mathbf{E}(x, y)$ does not decay with the distance from the axis are called radiation modes, as the optical wave radiates far away from the axis. There are an infinite number of radiation modes, and they are characterized by continuous mode profiles and propagation constants.

An important property of optical waves in a waveguide is mode orthogonality, which can be deduced from Maxwell's equations. Let the two modes in a waveguide be $\{\mathbf{E}_m, \mathbf{H}_m\}$ and $\{\mathbf{E}_{m'}, \mathbf{H}_{m'}\}$. The power flow of a mode is given by the integral over the cross section of the real part of the z component of the complex Poynting vector $\mathbf{S} = (1/2)\mathbf{E} \times \mathbf{H}^*$. It is convenient to normalize the field so as to have unity power flow. Then the orthonormal relation for guided modes is given by

$$\iint \mathrm{Re}\left(\frac{\mathbf{E}_{tm} \times \mathbf{H}^*_{tm'}}{2}\right)\mathrm{d}x\,\mathrm{d}y = \pm\delta_{mm'} \quad [4]$$

where $\delta_{mm'} = 0$ for $m \neq m'$, $\delta_{mm'} = 1$ for $m = m'$, and the plus and minus signs in the RHS represent the forward ($\beta_m > 0$) and backward ($\beta_m < 0$) modes, respectively. For radiation modes, the Cronecker's δ in the RHS of eqn [4] should be replaced by Dirac's δ function. The modal fields also possess the completeness, which means that an arbitrary field can be expressed in the form of a series expansion using modal fields.

3.05.2.2 Planar Waveguides

The simplest model of waveguides is a planar waveguide of thickness T with a step index profile as shown in **Figure 3(a)**. The index distribution is given by

$$n(x,y) = n(x) = \begin{cases} n_a, & (0 < x) \\ n_c, & (-T < x < 0)\ (n_c > n_s > n_a) \\ n_s, & (x < -T) \end{cases} \quad [5]$$

where n_a, n_c, and n_s are the refractive indexes of the (upper) cladding layer, guiding (core) layer, and lower cladding layer (substrate), respectively. It can be shown from Maxwell's equations that the modes in the planar waveguide structure can be classified into two polarization modes.

3.05.2.2.1 Transverse electric modes

One of the polarization modes is transverse electric (TE) modes, where the electric field has the transverse component E_y only and $E_x = E_z = 0$. The

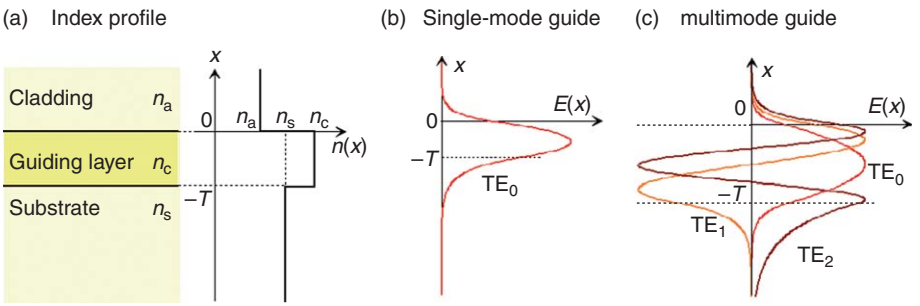


Figure 3 Planar waveguide of step index distribution and the mode field profiles. TE, transverse electric.

dominant field components $E_y(x)$ satisfies a wave equation

$$\left[\frac{d^2}{dx^2} + \left\{k^2 n^2(x) - \beta^2\right\}\right] E_y(x) = 0 \qquad [6]$$

and the magnetic field is given by $H_x = (-\beta/\omega\mu_0)E_y$, $H_z = (\mathrm{j}/\omega\mu_0)\,(dE_y/dx)$, and $H_y = 0$.

The guided mode solution of eqn [6] can be written as

$$E^y(x) = \begin{cases} E_a \exp\{-\gamma_a x\} & (0 < x) \\ E_c \cos\{\kappa_c x + \Phi_a\} & (-T < x < 0) \\ E_s \exp\{+\gamma_s(x+T)\} & (x < -T) \end{cases} \qquad [7]$$

where E_a, E_c, and E_s are constants and

$$\gamma_a = k\sqrt{N^2 - n_a^2}, \gamma_s = k\sqrt{N^2 - n_s^2}, \kappa_c = k\sqrt{n_c^2 - N^2} \qquad [8]$$

It should be noted that for guided modes the mode index N is in $n_s < N < n_c$. Using the continuity of $E_y(x)$ and $H_z(x)$, required by the boundary conditions at $x = 0$, T, we obtain a characteristic equation:

$$\begin{aligned} &\kappa_c T - \Phi_a - \Phi_s = m\pi \,(m = 0, 1, 2, \cdots) \\ &\Phi_a = \tan^{-1}\left(\frac{\gamma_a}{\kappa_c}\right), \Phi_s = \tan^{-1}\left(\frac{\gamma_s}{\kappa_c}\right) \end{aligned} \qquad [9]$$

which determines the mode index N. The integer m is the mode order number, and Φ_a and Φ_s are the phase shifts associated with the TIR. The constants E_a, E_c, and E_s are correlated by $E_a = E_c \cos\Phi_a$ and $E_s = E_c(-1)^m \cos\Phi_s$.

3.05.2.2.2 Transverse magnetic modes

Another polarization mode is transverse magnetic (TM) modes, where the magnetic field has the transverse component H_y only and $H_x = H_z = 0$. The dominant field component $H_y(x)$ satisfies the same wave equation as eqn [6], and the electric field is given by $E_x = (\beta/\omega\varepsilon_0 n^2)H_y$, $E_z = (-\mathrm{j}/\omega\varepsilon_0 n^2)$ $(\mathrm{d}H_y/\mathrm{d}x)$ and E_y=0. The guided mode solution for $H_y(x)$ can be written as

$$H_y(x) = \begin{cases} H_a \exp\{-\gamma_a x\} & (0 < x) \\ H_c \cos\{\kappa_c x + \Phi_a\} & (-T < x < 0) \\ H_s \exp\{+\gamma_s(x+T)\} & (x < -T) \end{cases} \qquad [10]$$

Using the continuity of $H_y(x)$ and $E_z(x)$, required by the boundary conditions at $x = 0$, T, we obtain a characteristic equation:

$$\begin{aligned} &\kappa_c T - \Phi_a - \Phi_s = m\pi \,(m = 0, 1, 2, \cdots) \\ &\Phi_a = \tan^{-1}\left\{\left(\frac{n_c}{n_a}\right)^2 \left(\frac{\gamma_a}{\kappa_c}\right)\right\} \\ &\Phi_s = \tan^{-1}\left\{\left(\frac{n_c}{n_s}\right)^2 \left(\frac{\gamma_s}{\kappa_c}\right)\right\} \end{aligned} \qquad [11]$$

The constants H_a, H_c,and H_s are correlated by $H_a = H_c \cos\Phi_a$ and $H_s = H_c(-1)^m \cos\Phi_s$.

The dominant field components are described by the same formula for the TE and TM modes. The field profiles of the fundamental ($m = 0$) mode and a few higher-order modes are illustrated in **Figure 3(b)** and **3(c)**, respectively. They are characterized by a sinusoidal variation in the core and the exponential decay in the cladding. The propagation constant and the mode index depend on the index profile, optical wavelength, polarization, and mode order. Although an explicit expression for N cannot be derived, it can be given in a convenient graphic form using normalized parameters. We define a normalized guide thickness (normalized frequency) by

$$V = kT\sqrt{n_c^2 - n_s^2} = \left(\frac{\omega}{c}\right) T\sqrt{n_c^2 - n_s^2} \qquad [12]$$

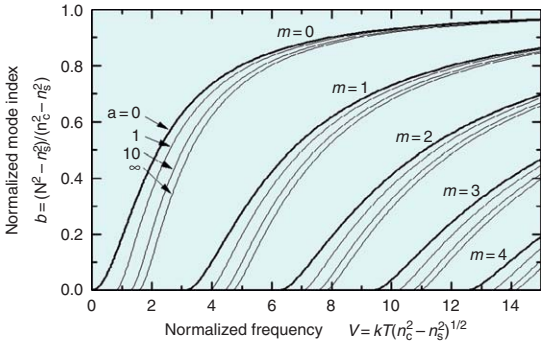


Figure 4 Mode dispersion curves of planar waveguides of step index distribution.

and an asymmetry parameter by $\alpha_{TE} = \left(n_s^2 - n_a^2\right)/\left(n_c^2 - n_a^2\right)$. The mode index N is represented by

$$b = \frac{N^2 - n_s^2}{n_c^2 - n_s^2} \qquad [13]$$

Then the characteristic equation [9] can be expressed by using V, α_{TE}, and b only. **Figure 4** shows the mode dispersion curves obtained by solving the characteristic equation. The mode number increases with V. The mode index N increases monotonously with increasing V and approaches n_f. With decreasing V, N decreases to n_s at the cutoff point. Waveguides that support only the fundamental ($m = 0$) mode are single-mode (monomode) waveguides, and those that support higher-order modes are multimode waveguides. For many applications, single-mode waveguides with small V (small index difference and small core thickness) are advantageous, as conversion and interference between modes may give rise to complication and deterioration of the device performances. By putting $N \rightarrow n_s$ in eqn [9], the cutoff condition for the mth mode is obtained as $V_{cm} = m\pi + \tan^{-1}(\alpha_{TE})^{1/2}$. For symmetrical waveguides ($\alpha_{TE} = 0$), there is no cutoff for the fundamental ($m = 0$) mode, and the single-mode condition is given by $V < \pi$. Similar treatment can be made for TM modes with an asymmetry parameter $\alpha_{TM} = (n_c/n_a^4)\left(n_s^2 - n_a^2\right)/\left(n_c^2 - n_a^2\right)$, and in many cases the expressions for TE modes can be used as a good approximation for TM modes.

Radiation modes can also be expressed in the form of eqn [1]. They are classified into substrate radiation modes and substrate–air radiation modes. The former is the superposition of two plane waves in the core and substrate regions with evanescent tail in the air (upper cladding) region, and $n_a < |N| < n_s$. The latter has propagating plane wave component(s) in all core, substrate, and air regions, and the effective index $|N|$ is $< n_a$. They are of continuous spectrum. The spectrum of the propagation constants of guided and radiation modes are illustrated in **Figure 5**.

Waveguides having a graded index profile can also be used for IO device implementation. For example, the index profiles of metal-indiffused $LiNbO_3$ waveguides and annealed proton-exchanged (APE) $LiNbO_3$ waveguides can be expressed approximately by a half Gauss function or a complementary error function. Exact analysis of such waveguides requires numerical calculations. The Wentzel–Kramers–Brillouin (WKB) method (Coldren and Corzine 1995, Nishihara *et al.* 1989, Snyder and Love 1983, Tamir 1975) and multilayer decomposition method (Nishihara *et al.* 1989, Snyder and Love 1983) are used. The characteristics, however, can be illustrated well with the above results, and the step index model can be used for design with reasonable accuracy, if appropriate values are used for refractive indexes and thickness to represent the graded index profile.

3.05.2.3 Channel Waveguides

Figure 6 illustrates various types of channel waveguides, and **Figure 7** shows the cross section of a channel waveguide with step index profile. The core thickness and width are T and W, respectively. The wave equation is given by eqn [2]. Although the exact field expressions for channel waveguides cannot be obtained in an analytical form, approximate but analytical solutions can be used to illustrate the characteristics and design the structure.

One of the methods is the effective index method (Coldren and Corzine 1995, Nishihara *et al.* 1989,

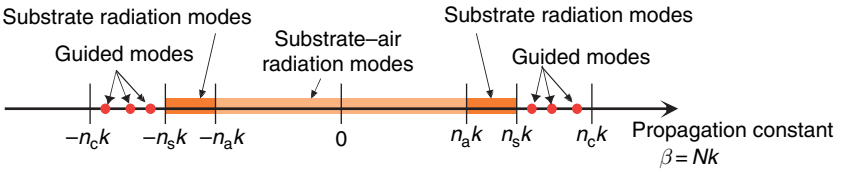


Figure 5 Spectrum of optical wave modes in a planar waveguide.

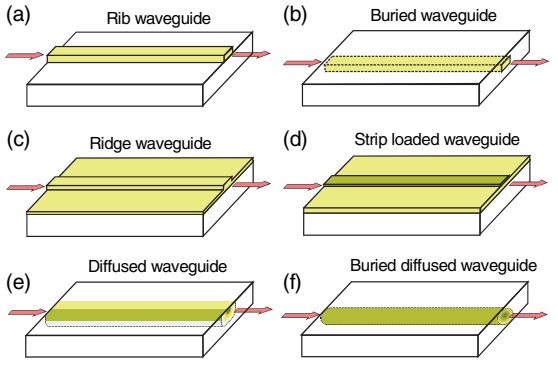


Figure 6 Various types of channel waveguides.

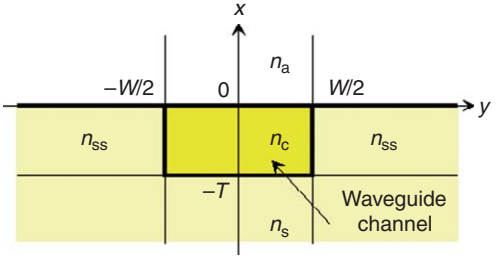


Figure 7 Cross section of a channel waveguide of step index distribution.

Tamir 1975) that applies with good accuracy for guides of $W >> T$. The channel structure is converted into two imaginary planar waveguides, and the field distribution is written approximately in the form of a product of profiles in the depth and width directions as, for example, $E_y(x,y) = E_y(x)E_y(y)$. In the first step, the depth profile $E_y(y)$ and the effective index N_I for the channel region are determined by using an imaginary planar guide model of thickness $T(W \to \infty)$. In the second step, an imaginary planar waveguide of core thickness W with an effective core index N_I and an effective cladding index $N_{II} = \left(n_{ss}^2 - n_c^2 + N_I^2\right)^{1/2}$ is considered to determine the width profile $E_y(x)$ and the mode index N. Thus the mode profile and the mode index N of guided mode in channel waveguide are obtained by repeating the analysis of planar waveguide. The normalized graphic data of **Figure 4** can be used in each step.

Another approximate method is the Marcatili's (1969) method, which gives better accuracy for cases including narrow channels. The field profile of guided modes, e.g., $E_y(x,y)$, is expressed analytically by using sinusoidal and exponential functions for five regions (core region, and upper, lower, left, and right cladding regions), and from the boundary

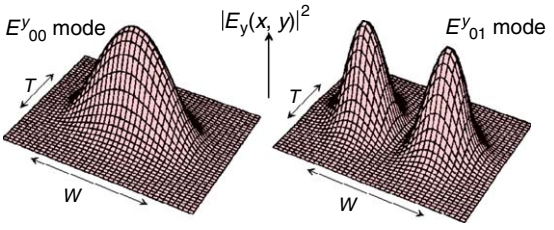


Figure 8 Optical intensity distributions of guided modes in a channel waveguide.

conditions at the upper, lower, left, and right walls of the core, characteristic equations are obtained. Integers p and q are used as mode order numbers for the depth and width directions. The propagation constant β can be obtained by solving these equations jointly. **Figure 8** illustrates the examples of the intensity distributions given by $\left|E_y(x,y)\right|^2$.

An important feature of the above channel guided modes is that, although the dominant electric field component is E_y, there are nonzero E_z and H_z. This means that these modes are not exactly TE modes but hybrid modes (Nishihara *et al.* 1989, Snyder and Love 1983). They are designated as E^y_{pq} modes. There also exist guided modes with $E_y = 0$ and dominant electric field component $E_x(x,y)$. These modes are designated as E^x_{pq} modes. The E^y_{pq} and E^x_{pq} modes are also called TE-like and TM-like modes, respectively, and are often abbreviated as TE_{pq} and TM_{pq} modes. Analysis of channel waveguides of arbitrary index profile can be performed by numerical method based on the finite element method (FEM) (Koshiba *et al.* 1992) and the finite difference method (FDM) (Stern 1988).

3.05.2.4 Waveguide Materials and Fabrication Techniques

Waveguides are fabricated by using materials selected depending on the optical wavelength and the function to be implemented. Although a large number of materials was studied, only some of them survived as materials for developing practical IO devices. They are classified into glasses, polymers, ferroelectric crystals, and semiconductors. Typical waveguide materials are listed in **Table 1** with the fabrication techniques and applications.

Typical waveguide material for passive components is optical glasses. Optical waveguides can be fabricated in multicomponent glasses by ion exchange. For example, a soda lime glass substrate is immersed in molten salt such as KNO_3, $AgNO_3$,

Table 1 Various kinds of optical waveguide materials

Category	*Material*	*Fabrication technology*	*Application*
Multicomponent glasses	Borosilicate glasses[a]	RF sputtering	PC, TO devices
	Soda lime glasses	Ion exchange	PC, TO devices
	Chalcogenide glasses[a]	Vacuum evaporation	AO devices
Silica glasses	SiO_2 (TiO_2 doped)[a]	FHD	PC, TO devices
	SiO_xN_y[a]	PCVD	PC, TO devices
	Er:SiO_2[a]	Rare-earth doping	Lasers
Ferroelectric crystals	$LiNbO_3$	Ti indiffusion, proton exchange	EO, AO, NLO devices
	$LiTaO_3$	Proton exchange	EO, AO, NLO devices
	Er:$LiNbO_3$	Rare-earth doping	Lasers
Polymers	Polymers[a]	Spin coating, dip coating, casting	PC, TO devices
	EO polymers[a]	Spin coating, dip coating, casting	EO devices
Silicon	SOI ($Si/SiO_2/Si$)	Wafer direct bonding, ion implantation	IR devices
Group III–V semiconductors	AlGaAs	MBE	Lasers
	InGaAsP	MOVPE	Optical amplifiers
	InGaAs		EA devices
	AlGaInP		Monolithic IO devices

[a]On SiO_2/Si substance prepared by thermal oxidation of Si substrate.
AO, acoustooptic; EA, electroabsorption; EO, electrooptic; FHD, flame hydrolysis deposition; IR, infrared; MBE, molecular beam epitaxy; MOVPE, metal–organic vapor phase epitaxy; NLO, nonlinear optic; PC, passive component; PCVD, plasma chemical vapor deposition; RF, radio frequency; SOI, silicon on insulator; TO, thermooptic.

and $TlNO_3$ to replace Na^+ in the glass matrix by K^+, Ag^+, and Tl^+ in the salt. The refractive index is increased as a result of the exchange. Channel guides can be fabricated by using a patterned mask layer on the substrate. Fabrication techniques of glass waveguides include deposition of borosilicate glass (e.g., Corning #7059) by radio frequency (RF) sputtering and deposition of chalcogenide glasses by vacuum evaporation or RF sputtering.

An important structure for various thin-film waveguides is SiO_2/Si substrate, which is prepared by growing an SiO_2 layer on a Si wafer by thermal oxidation. The advantages are high quality and low cost, and possibilities of integration of photodiodes (PDs) and electronics. The SiO_2 layer with low refractive index serves as an optical buffer layer to isolate the guided mode field confined in the guiding layer from the high-index Si substrate. The appropriate buffer layer thickness is a few times as large as the penetration depth of the evanescent tail of the guided mode.

One of the high-quality ($<0.05\,dB\,cm^{-1}$ loss) waveguides now most widely used is buried silica (SiO_2 doped with GeO_2 or TiO_2) waveguides on Si substrate (Himeno *et al.* 1998). **Figure 9** illustrates the fabrication process of silica channel waveguides by flame hydrolysis deposition (FHD) and reactive ion etching (RIE) combined with photolithography or electron beam lithography. Silica waveguides and silicon oxynitride (SiO_xN_y) waveguides with SiO_2

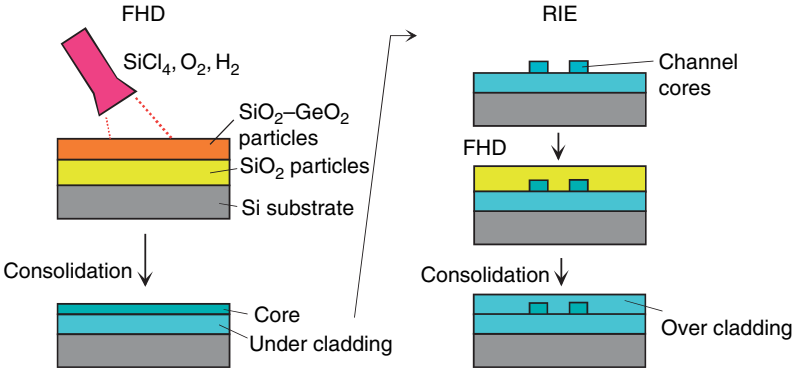


Figure 9 Fabrication process of silica channel waveguides by the flame hydrolysis deposition (FHD) and reactive ion etching (RIE).

cladding can also be fabricated by plasma-enhanced chemical vapor deposition (PCVD) and RIE.

Another important advantage of using a Si substrate is that various micromachining processes originally developed for very large-scale integration (VLSI) fabrication, including various lithography, anisotropic etching, surface micromachining, and replication process, can be employed to form 3D microstructures (Gabriel 1998). The structures can find applications as a Si optical bench. Microelectromechanical systems (MEMS) consisting of mechanical moving parts and micro actuators can also be constructed. Optical systems using MEMS structures are called optical MEMS or microoptoelectromechanical systems (MOEMS). A number of MOEMS have been developed (Giles *et al.* 1999, Lin and Goldstein 2002, Motamedi 1994).

Important waveguide materials for electrooptic (EO), acoustooptic (AO), and nonlinear optic (NLO) devices are ferroelectric crystals such as $LiNbO_3$ and $LiTaO_3$. High-quality waveguides ($<0.2\,dB\,cm^{-1}$ loss) can be fabricated by Ti indiffusion in $LiNbO_3$ and proton exchange/annealing in $LiNbO_3$ and $LiTaO_3$. The fabrication processes are illustrated in **Figure 10**. Ti:$LiNbO_3$ waveguides are fabricated by deposition and pattering of Ti film by RF sputtering and liftoff, followed by heat treatment in a quartz tube by an electric furnace at a temperature $\sim$1000°C in an appropriate atmosphere (wet N_2 to suppress LiO_2 outdiffusion). APE waveguides are fabricated by immersing the crystal with patterned mask layer in molten benzoic acid at $\sim$200°C for exchange of Li^+ in the crystal with H^+ (proton) in the acid. The exchange gives rise to increase in the extraordinary index of the crystal. Then the crystal is annealed at $\sim$350°C to obtain index profile for single-mode guiding.

Waveguides of increasing interest include Si waveguides for infrared light. Silicon-on-insulator (SOI) structures with appropriate Si thickness, prepared by wafer direct bonding, can be used as a planar waveguide (Jalali 1997). Channel guides can be fabricated by standard Si IC technologies. They exhibit low propagation losses for wavelength longer than 1.3 μm. Because of the high index of refraction, the Si layer is suitable for the fabrication of photonic crystal waveguides. They have the advantages that well-established Si IC and MEMS processes can be employed and offer the possibility of implementing devices for optical wave control by carrier or thermooptic (TO) effect and Raman optical amplifiers.

Organic polymers are a material with unique features. Waveguide layers are prepared from polymer solutions by spin coating or dip coating on various substrates, and channel guides can readily be fabricated by RIE and buried with another polymer with lower index. They are suitable for many passive devices including IO chemical and biomedical sensors. The TO coefficient is much larger than that of glass materials. EO polymers with EO coefficients comparable or even higher than that of $LiNbO_3$ have been developed.

Group III–V compound semiconductors, such as GaAs and InGaAsP, are important materials for monolithic IO devices including laser diodes (LDs) as light sources. Double-hetero (DH) and separately confinement hetero (SCH) structures for bulk and

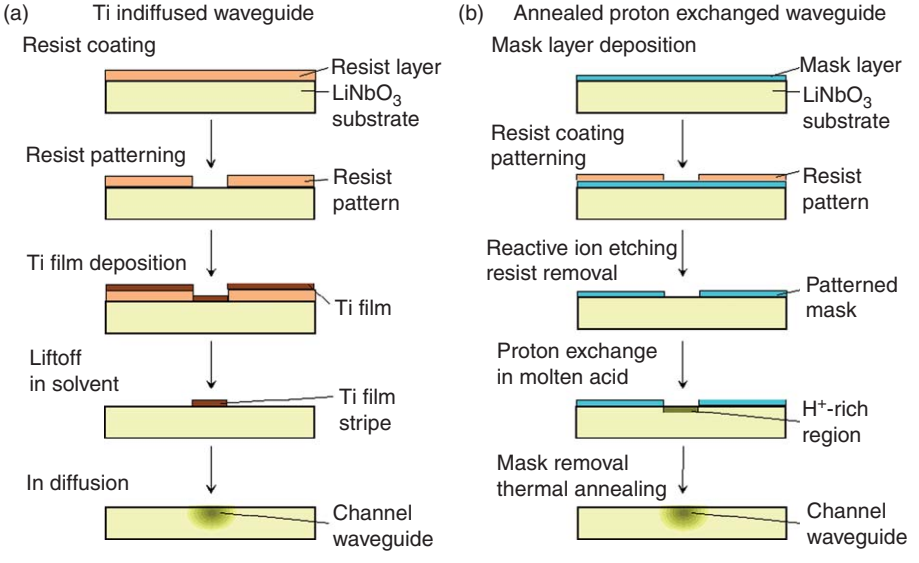


Figure 10 Fabrication process of $LiNbO_3$ channel waveguides.

quantum-well (QW) lasers can be used also as passive waveguides. They are fabricated by molecular beam epitaxy (MBE) or metal–organic vapor phase epitaxy (MOVPE). The passive waveguides without current injection, however, exhibit large absorption loss at the lasing wavelength. An important requirement for the implementation of monolithic IO devices is reduction in the passive waveguide loss. To solve the problem, area-selective MOVPE growth technique using SiO_2 mask layer (Kudo *et al.* 1997) or area-selective QW disordering (intermixing) technique (Li 2000) is employed to modify the bandgap energy distribution.

Propagation losses of guided waves has several origins. They include absorption loss due to absorption of guided wave power by the core and cladding materials (insufficient transparency). Imperfection of the interface between core and cladding and inclusion of defects in the structure give rise to scattering loss. Although the losses depend on wavelength, materials, and fabrication processes, the losses are reduced to values (typically $< 0.2\,\mathrm{dB\,cm^{-1}}$) appropriate for practical device implementation.

3.05.2.5 Excitation of Guided Mode by End Coupling

The simplest and practical method for guided mode excitation in a waveguide is end-fire coupling as shown in **Figure 11(a)**. A laser beam is focused by a lens into a small spot, and the input end of the waveguide is positioned at the focus. The input face is prepared by cutting and polishing the waveguide. For some waveguide materials, precision dicing method using a disk saw can be used. For semiconductor waveguides, a waveguide facet prepared by cleaving the waveguide with substrate can be used as the input facet. The coupling efficiency η is calculated by expanding the field distribution on the waveguide input end $E_{\mathrm{in}}(x, y, 0)$ with field profiles of the modes of the waveguide $E_m(x,y)(m = 0, 1, 2, \cdots)$ and by taking the absolute square of the coefficient for the mode to be excited and is given by

$$\eta = \frac{\left| \iint E_m^*(x,y) E_{\mathrm{in}}(x,y,0)\,\mathrm{d}x\,\mathrm{d}y \right|^2}{\iint |E_m(x,y)|^2\,\mathrm{d}x\,\mathrm{d}y \iint |E_{\mathrm{in}}(x,y,0)|^2\,\mathrm{d}x\,\mathrm{d}y} \quad [14]$$

The numerator in the RHS of eqn [14] is called overlap integral. The value of η takes the maximum of 1 when $E_{\mathrm{in}}(x,y,0)$ is proportional to $E_m(x,y)$ and is <1 otherwise. To obtain a high coupling efficiency, the beam should be focused on the input face, the focus spot should be aligned with the waveguide core, and the spot size should be matched with the mode size by optimizing the combination of input beam diameter and focal length of the lens. In many cases of appropriately designed channel waveguides, it is not difficult to achieve $\eta > 0.7$ using micromanipulator and an appropriate lens.

Laser light emitted from a semiconductor laser can be coupled directly into a waveguide by aligning the laser to the waveguide with a narrow gap between them as shown in **Figure 11(b)**. This method is called butt coupling, and the coupling efficiency is given by eqn [14] with the laser mode profile $E_{\mathrm{in}}(x,y)$ instead of $E_{\mathrm{in}}(x,y,0)$. High coupling efficiency is obtained by precise alignment and mode matching. Laser light transmitted through an optical fiber can also be coupled to a waveguide by similar butt coupling. The end of the fiber is bonded to the waveguide end using optical glue. This is called fiber pigtailing and is widely used for implementing practical devices.

3.05.3 Distributed Coupling of Guided Waves

Distributed coupling of guided waves in two or more waveguides and that in a waveguide with periodic structures or gratings provide important working principles of IO components and devices. Waveguide

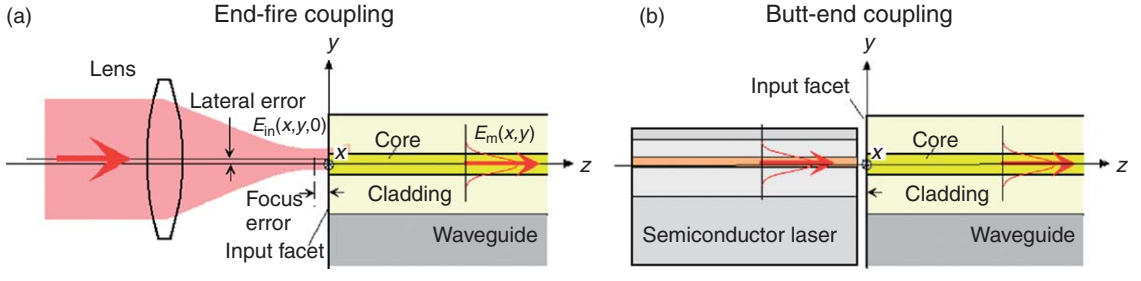


Figure 11 Excitation of guided mode by end coupling.

gratings perform various passive functions, such as deflection, reflection, input/output coupling, mode conversion, wavelength filtering, wavelength dispersion, and phase matching. Periodic modulation in refractive index produced through EO and AO effects, i.e., dynamic gratings, can provide an effective means for guided wave control. This section presents distributed coupling in waveguides and waveguide gratings based on the coupled mode theory (Nishihara *et al.* 1989, Suhara *et al.* 1986, Yariv 1973).

3.05.3.1 Description of Periodic Structures and Phase Matching

Consider a waveguide described by a cross-sectional distribution of the relative dielectric permittivity $\varepsilon(x,y)$. Expression of a waveguide grating is given by the modification in the permittivity $\Delta\varepsilon(x,y,z)$ superimposed on $\varepsilon(x,y)$. Because $\Delta\varepsilon$ is periodic, it can be written in a Fourier series as

$$\Delta\varepsilon(x,y,z)=\sum_q \Delta\varepsilon_q(x,y)\exp(-\mathrm{j}qKz),\quad K=2\pi/\Lambda \qquad [15]$$

where Λ is the grating period, and $\Delta\varepsilon_q(x,y)$ the qth Fourier amplitude. A grating of periodicity along the z-axis is assumed for simplicity. Note that a waveguide structure modified without grating can be described by $\Delta\varepsilon_0(x,y)$ and $\Delta\varepsilon_q(x,y)=0$ for $q\neq 0$.

Optical functions of the structures can be interpreted in terms of coupling between modes. When a guided mode characterized by a phase factor $\exp(-\mathrm{j}\beta_a z)$ with a propagation constant β_a is incident in the interaction region, space harmonics of phase factors β_a+qK are induced. The harmonic can propagate as a mode, provided that the β_a+qK coincides at least approximately with propagation constant β_b of a mode in the structure. This means that effective qth-order coupling between modes a and b takes place when

$$\beta_b=\beta_a+qK,\quad q=0,\pm1,\pm2,\cdots \qquad [16]$$

Eqn [16] is the phase-matching condition, and more general relation, $\boldsymbol{\beta}_b=\boldsymbol{\beta}_a+q\mathbf{K}$ with wave vectors $\boldsymbol{\beta}_a$, $\boldsymbol{\beta}_b$, and grating vector $\mathbf{K}$ is called Bragg condition and can be depicted as a wave vector diagram.

3.05.3.2 Coupled Mode Equations

The optical field in the waveguide grating structure may be written as

$$\mathbf{E}(x,y,z)=A(z)\mathbf{E}_a(x,y)\exp(-\mathrm{j}B_a z)+B(z)\mathbf{E}_b(x,y)\exp(-\mathrm{j}B_b z) \qquad [17]$$

The mode fields $\mathbf{E}_a(x,y)\exp(-\mathrm{j}\beta_a z)$ and $\mathbf{E}_b(x,y)\exp(-\mathrm{j}\beta_b z)$ satisfy Maxwell's equations with waveguide permittivity $\varepsilon(x,y)$, and the total field $\mathbf{E}(x,y,z)$ satisfies those with $\varepsilon(x,y)+\Delta\varepsilon(x,y,z)$. From these Maxwell's equations, and using the orthonormal relation of the modes, one can deduce the equations that describe the spatial evolution of the mode amplitudes $A(z)$ and $B(z)$:

$$\pm\frac{\mathrm{d}}{\mathrm{d}z}A(z)=-\mathrm{j}\kappa^* B(z)\exp(-\mathrm{j}2\Delta z) \qquad [18a]$$

$$\pm\frac{\mathrm{d}}{\mathrm{d}z}B(z)=-\mathrm{j}\kappa A(z)\exp(+\mathrm{j}2\Delta z) \qquad [18b]$$

$$2\Delta=\beta_b-(\beta_a+qK) \qquad [19]$$

As for $\pm$ in eqn [18a] (eqn [18b]), $+$ and $--$ should be taken for $\beta_a>0$ and $\beta_a<0$ ($\beta_b>0$ and $\beta_b<0$), respectively. Equations [18] are coupled mode equations, and 2Δ describes the deviation from the phase-matching condition. The parameter κ, called coupling coefficient, is given by

$$\kappa=\frac{\omega\varepsilon_0}{4}\iint \mathbf{E}_a^*(x,y)\Delta\varepsilon_q(x)\mathbf{E}_b(x,y)\mathrm{d}x\,\mathrm{d}y \qquad [20]$$

3.05.3.3 Coupling between Guided Modes

3.05.3.3.1 Collinear coupling

Coupling between guided modes propagating along the same axis (collinear coupling) is classified into codirectional coupling and contradirectional coupling. They are illustrated in **Figure 12(a)–(c)** with the wave vector diagrams.

Codirectional coupling: Consider coupling between two different modes propagating in the same direction ($\beta_a>0,\beta_b>0$). The solution of eqns [18a] and [18b] for boundary conditions $A(0)=1,B(0)=0$ indicates periodic transfer of the guided mode power from mode a to b. The efficiency of the power transfer in a coupling (grating) length L is given by

$$\eta=\left|\frac{B(L)}{A(0)}\right|^2=\frac{\sin^2\sqrt{|\kappa|^2+\Delta^2}L}{1+\Delta^2/|\kappa|^2} \qquad [21]$$

The efficiency under exact phase matching ($\Delta=0$) is given by $\eta_0=\sin^2|\kappa|L$. For coupling of guided waves of the same mode ($\beta_a=\beta_b$) in two parallel identical waveguides (a directional coupler (DC)) (**Figure 12(a)**), the phase matching is satisfied

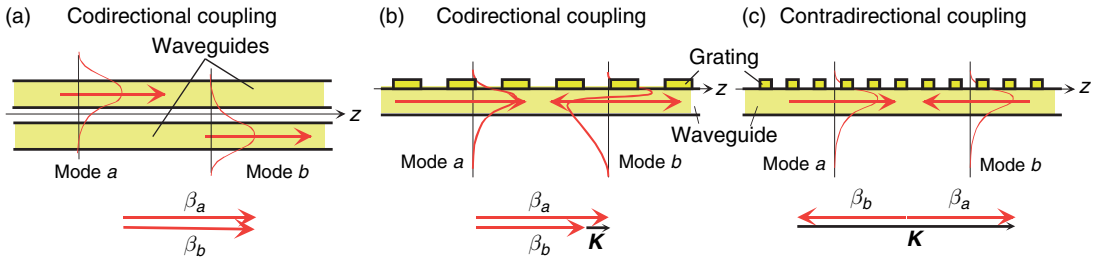


Figure 12 Collinear coupling of guides modes in waveguides or in a waveguide with grating.

automatically without grating ($q = 0$). Coupling of two different modes ($\beta_a \neq \beta_b$) in a waveguide may take place when the phase matching is accomplished by a grating of appropriate period (**Figure 12(b)**). This implies that the grating functions as a mode converter. In either case, complete power transfer takes place when L equals odd integer multiple of the complete coupling length $L_c = \pi/2|\kappa|$. The dependence of η on κL and that of η/η_0 on ΔL are plotted in **Figure 13**.

Contradirectional coupling: Consider coupling between modes propagating in the opposite directions ($\beta_a > 0$, $\beta_b < 0$) in a waveguide with grating of length L (**Figure 12(c)**). Equations [18] with $A(0) = 1, B(L) = 0$ give a solution that indicates monotonous power transfer. This implies that distributed reflection of forward mode a into backward mode b takes place. The mode b may be the same lateral mode as a ($\beta_b = -\beta_a$, reflection without mode conversion) or may be a different mode ($\beta_b \neq -\beta_a$, reflection associated with mode conversion). The grating is called distributed Bragg reflector (DBR). The conversion efficiency, or the reflectivity, is given by

$$\eta = \left|\frac{B(0)}{A(0)}\right|^2 = \left[1 + \frac{1 - \Delta^2/|\kappa|^2}{\sinh^2\sqrt{|\kappa|^2 - \Delta^2}L}\right]^{-1} \qquad [22]$$

Under exact phase matching ($\Delta = 0$), the efficiency is given by $\eta_0 = \tanh^2|\kappa|L$, and most of the power is transferred ($\eta > 0.84$) when $L > \pi/2|\kappa|$. The dependence of η on κL and that of η/η_0 on ΔL are plotted in **Figure 14**.

Equations [21] and [22] indicate that a high efficiency is obtained only when phase matching is satisfied, and the mismatch gives rise to a reduction in the efficiency. Therefore, for a given waveguide grating, a high efficiency is obtained at or near a phase-matching wavelength λ_0, and the efficiency is reduced with deviation of the wavelength from λ_0. Thus the mode conversion and the reflection exhibit wavelength selectivity, and the grating can be used as a wavelength filter. Let $\delta\lambda = \lambda - \lambda_0$ be the wavelength deviation. Assuming $N(\lambda) \approx N(\lambda_0)$, we have $2\Delta \approx (N_a - N_b)(2\pi\delta\lambda/\lambda^2)$ for codirectional coupling, and $2\Delta \approx 2N(2\pi\delta\lambda/\lambda^2)$ for contradirectional coupling of $\beta_b = -\beta_a$. Combining these relations with the bandwidths in terms of ΔL, the wavelength bandwidths are given by $2\delta\lambda/\lambda \approx 2.5\Lambda/\pi|q|L$ and $2\delta\lambda/\lambda \approx 5\Lambda/\pi|q|L$, for co- and contradirectional coupling, respectively. Because the grating period Λ for contradirectional coupling is much shorter (as can be seen from the phase-matching condition [16]), contradirectional coupling (DBR)

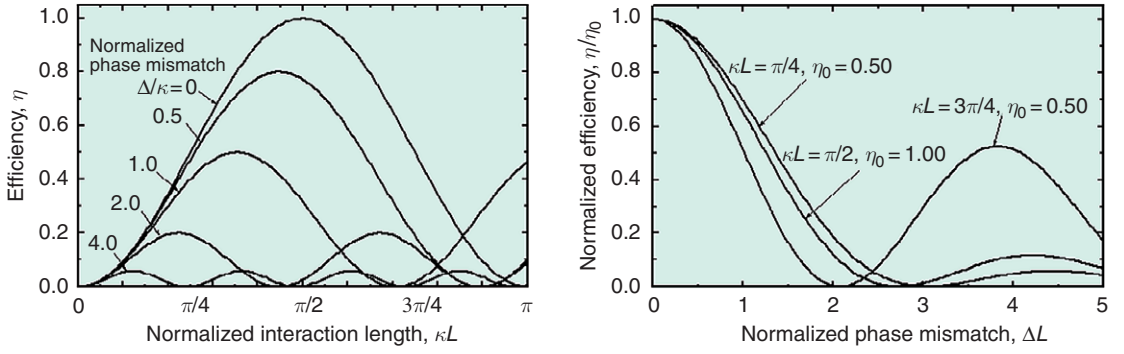


Figure 13 Dependence of the efficiency on the interaction length and the phase mismatch for codirectional coupling.

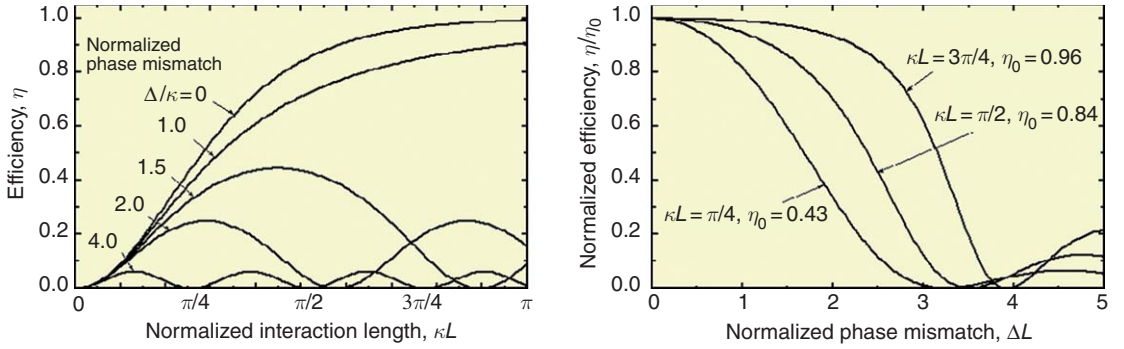


Figure 14 Dependence of the efficiency on the interaction length and the phase mismatch for contradirectional coupling.

exhibits wavelength selectivity much sharper than that of codirectional coupling.

The coupling coefficient κ is an important parameter needed to design the grating structure and predict the performances. The magnitude of κ depends on the order and polarizations of the coupling modes, and coupling order q, as well as the grating parameters. For gratings of index modulation type, for example, the κ is given by $\kappa = \pi\Delta n_q F/\lambda$, where Δn_q is the qth Fourier amplitude of the index modulation and F is an overlap integral between modes and index modulation describing the effect of mode confinement in the waveguide.

3.05.3.3.2 Coplanar coupling

In a planar waveguide (in yz-plane) with a grating (of length L in z-direction) of appropriate orientation, coplanar coupling takes place between guided waves of different propagation directions. This type of coupling is called diffraction or deflection, as it involves a change in propagation direction. Coplanar coupling is classified into two categories using a parameter defined by $Q = K^2 L/\beta$.

In a thin grating of $Q < 1$, several beams appear corresponding to different diffraction orders q, because coupling may take place even without exact phase matching in the z component. This type is called Raman–Nath diffraction. The first-order efficiency $\eta_{\pm 1}$ takes the maximum 0.34 at $2\kappa L = 1.84$. The grating does not exhibit sharp angular and wavelength selectivity. In a thick grating of $Q >> 1$, strong coupling takes place only when the Bragg condition is satisfied between the incident and the diffracted waves. Therefore, a diffracted wave of a single order appears, as illustrated in **Figure 15**. This type of diffraction is called Bragg diffraction. The diffraction efficiencies for coplanar coupling are described by the formulae similar to eqns [21] and [22]. The maximum efficiency for a transmission grating is 100%, and the diffraction efficiency of a reflection grating approaches asymptotically 100% with increasing L. In either case, the efficiency decreases with deviation from the Bragg condition, so that they exhibit the angular and wavelength selectivities.

3.05.3.4 Coupling between Guided Mode and Radiation Modes

3.05.3.4.1 Prism coupling

Prism coupling is a convenient laboratory method useful not only for guided mode excitation but also for measurement of the mode index and waveguide

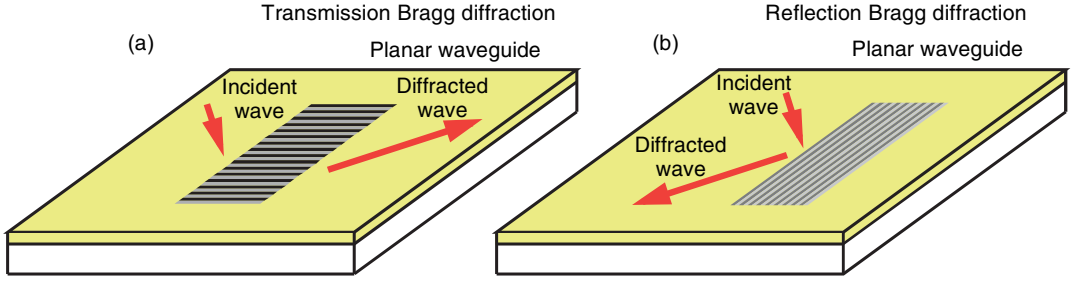


Figure 15 Bragg diffraction of guided wave in a planar waveguide.

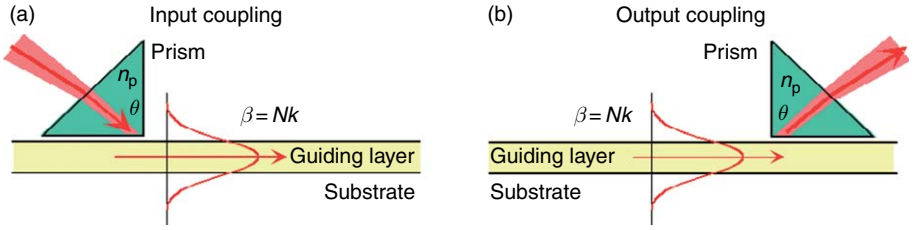


Figure 16 Prism coupling between a guided mode and radiation modes.

performances. A prism of a high refractive index n_p is pressed slightly to the waveguide surface, so that the prism base is aligned parallel to the waveguide surface with a submicron air gap between them as illustrated in Figure 16. A laser beam is weakly focused by a lens onto the edge of the prism base. The assembly is mounted on a rotary stage, and the angle of incidence θ, measured in the prism with respect to the waveguide surface normal, is adjusted so as to satisfy

$$n_p\, k \sin\theta = \beta\left(k = \frac{2\pi}{\lambda}, \beta = Nk\right) \qquad [23]$$

with the propagation constant β and the effective index N of a guided mode of the waveguide. Then the phase matching between the incident beam (radiation mode) and the guided mode is accomplished and the incident beam is coupled into the waveguide. It is necessary to use a prism of $n_p > N$. The mode index N can be determined experimentally by measuring the angle θ under the coupling. Coupling efficiency $>80\%$ is possible. Prism coupling, however, is hardly used in practical IO devices, because of the need of an expensive and rather bulky prism and critical adjustment, poor mechanical stability, and incompatibility with buried waveguides.

3.05.3.4.2 Grating coupling

Output coupling: Figure 17 illustrates coupling between a guided mode and radiation modes with a grating. Coupling takes place between waves satisfying phase matching for the z component. When a guided wave of propagation constant $\beta_0 = Nk$ is incident, the qth harmonics are radiated into air and/or substrate at angles determined by

$$n_c k \sin\theta_q^{(c)} = n_s k \sin\theta_q^{(s)} = \beta_q = Nk + qK \qquad [24]$$

The number of radiation beams equals the number of real values for $\theta_q^{(c)}$ and $\theta_q^{(s)}$. Figure 17(a) shows multibeam coupling where more than three beams are yielded, and Figure 17(b) shows two-beam coupling where only a single beam for the fundamental order ($q = -1$) is yielded in both air and substrate. Another possibility is one-beam coupling where a beam radiates only into the substrate. The amplitude of the guided and the radiation waves decays as $g(z) = \exp(-\alpha_r z)$ due to the power leakage by

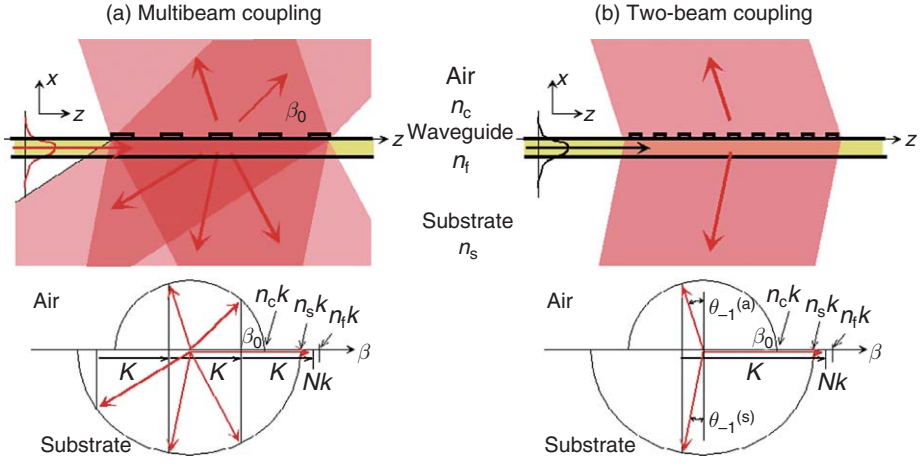


Figure 17 Coupling between a guided mode and radiation modes in a grating coupler.

radiation. Because the attenuation of the guided power corresponds to the power transferred to the radiation modes, the qth-order output coupling efficiency of a grating of length L can be written as

$$\eta_{\text{out}} = P_q^{(i)}\{1 - \exp(-2\alpha_r L)\},$$
$$\alpha_r = \sum_{q,i} \alpha_q^{(i)}, P_q^{(i)} = \frac{\alpha_q^{(i)}}{\alpha_r} \quad [25]$$

where $\alpha_q^{(i)}$ and α_r are radiation decay factors, $P_q^{(i)}$) is a ratio for power distribution to $q-(i)$ radiation beam, and $i =$ a or s. The radiation decay factors $\alpha_q^{(i)}$ are correlated with the coupling coefficient $\kappa_q^{(i)}$ by $\alpha_q^{(i)} = \pi\left|\kappa_q^{(i)}\right|^2$, and $\kappa_q^{(i)}$ can be calculated by substituting the guided and the radiation mode profiles into eqn [20]. For grating couplers of relief type with groove depth h, the decay factor α_r increases in proportion to h^2 in the region where h is small. For larger h, α_r saturates because of the limited penetration of the evanescent tail of the guided mode into the grating layer.

Input coupling: A guided mode can be excited through input coupling of an external beam incident on a grating at an angle satisfying eqn [24]. **Figure 18** compares output and input couplings. A reciprocity theorem analysis shows that the input coupling can be written as analysis shows that the input coupling can be written as

$$\eta_{\text{in}} = P_q^{(i)} I(g,h), I(g,h) = \frac{\left|\int g(z)h(z)\mathrm{d}z\right|^2}{\int |g(z)|^2\mathrm{d}z \int |h(z)|^2\mathrm{d}z} \quad [26]$$

where $h(z)$ is the input beam profile. The overlap integral $I(g, h)$ takes the maximum of unity, when the input profile $h(z)$ is proportional to the output profile $g(z)$. A high efficiency can be achieved under conditions $\alpha_r L >> 1$, $P_q^{(i)} \approx 1$ for a single $q-(i)$ beam, and $h(z) \approx g(z)$. For a Gaussian input beam, the maximum of $I(g, h)$ is 0.801 and the maximum input efficiency is 80.1%.

High-efficiency grating coupler: One-beam coupling is desirable for achieving high input and output efficiencies. Such coupling can be realized in a short-period grating, which allows only fundamental backward coupling into substrate ($\theta_{-1}^{(s)} < 0$). Two-beam couplers as shown in **Figure 17(b)** are more practical, but they have a drawback that the output power is divided into air and substrate. A high directionality into air ($P_{-1}^{(a)} \approx 1$) for eliminating this drawback can be accomplished by inserting a reflection layer on the substrate, by using the Bragg effect in a thick index-modulation grating, or by using the blazing effect in a relief grating having an asymmetric triangular cross section.

3.05.4 Passive Components

This section presents passive waveguide components that perform static functions on guided waves.

3.05.4.1 Modified Waveguide Components

Channel waveguides can be curved or bent to change the direction of guided waves and/or to give a lateral shift. They can also be tapered to change the mode size. If a waveguide is connected abruptly with another waveguide with different parameters, some of the guided wave power is transferred to radiation modes, resulting in mode conversion loss. A method to reduce this loss is to insert an adiabatic tapered region. If the waveguide is curved, the guided wave suffers from bending loss too. The bending loss increases rapidly as the radius of curvature decreases and is larger for smaller difference in index between the core and cladding. Therefore, waveguides of large index difference must be used for devices using largely curved guides. Waveguides can be branched for beam splitting (**Figure 19(a)**) (and combining for waves with

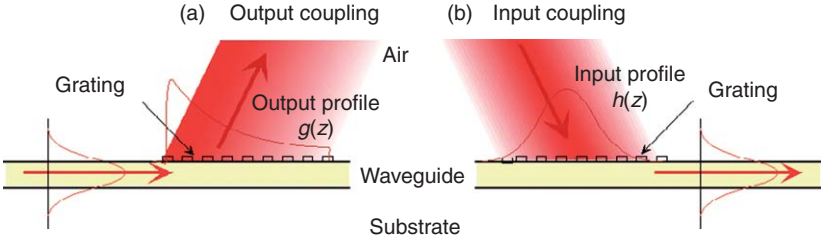


Figure 18 Comparison between grating output coupling and input coupling.

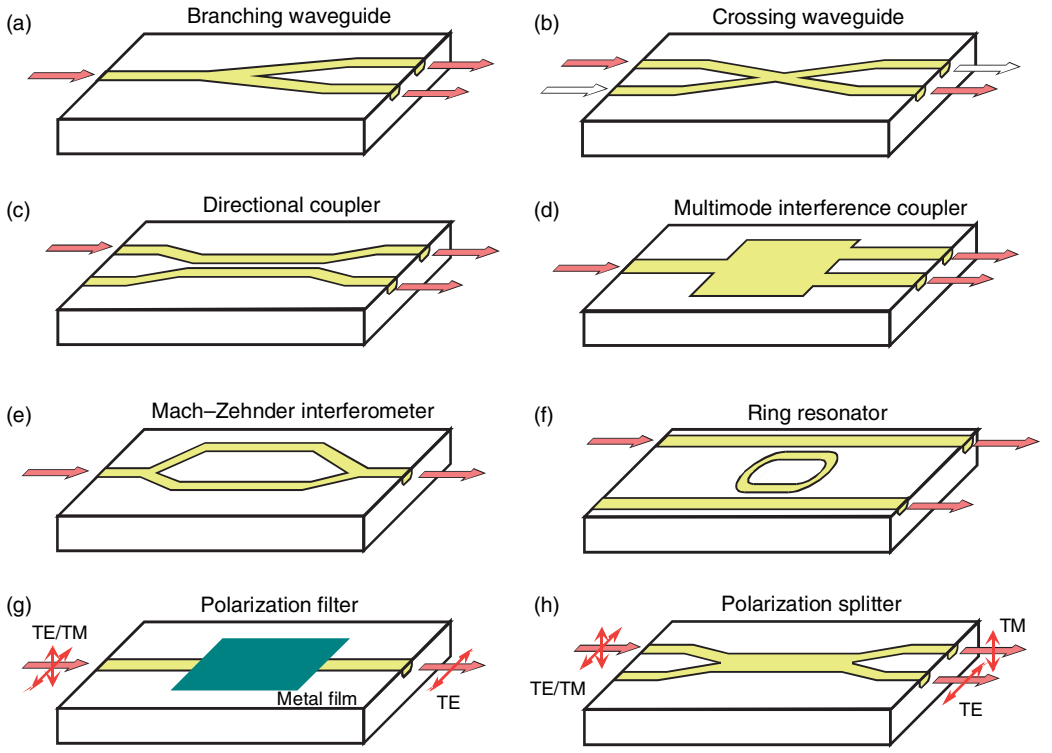


Figure 19 Various passive components using modified channel waveguides. TE, transverse electric; TM, transverse magnetic.

reverse propagation directions) and crossed for intersecting (**Figure 19(b)**). These modifications are made to interface components in IO.

One of the important components is a DC (**Figure 19(c)**), constructed with two-channel waveguides aligned parallel with a small gap for distributed coupling. The transmittances from an input port to the extended and to the crossed output ports are given by $\cos^2(\pi L/2L_c)$ and $\sin^2(\pi L/2L_c)$, respectively, where L is the length of the coupling region and $L_c = \pi/2\kappa$ is the complete coupling length, where κ is the coupling coefficient. For example, a 3-dB (50:50) coupler and a 0-dB (0:100) coupler are realized by setting $L = L_c/2$ and $L = L_c$, respectively, and they are used for beam splitting and beam transfer. Another type of coupler is a multimode interference (MMI) coupler (**Figure 19(d)**), consisting of input and output channels and a multimode section where lateral modes interfere with each other to give split outputs through output channels. Waveguide Mach–Zehnder interferometers (MZIs) (**Figure 19(e)**) are constructed by two channel guides (arms) connected by two branching guides (or two DCs). Asymmetric (different arm lengths) MZIs have applications as a course wavelength filter and sensor elements. Waveguide ring resonators (**Figure 19(f)**), constructed using input/output channel(s) and a ring channel with distributed coupling, perform wavelength filter function. The free spectrum range and the wavelength bandwidth are determined by the round-trip length of the ring and the coupling strength, respectively. Another resonator is waveguide Fabry–Perot (FP) resonator, constructed by integrating two waveguide DBR gratings or by forming multilayer reflectors on both end facets of a straight channel waveguide.

Polarization elements can also be implemented with waveguides. A waveguide polarizer can readily be realized by depositing a metal film on a dielectric waveguide (**Figure 19(g)**). The TM modes (the electric field perpendicular to the metal film) are strongly attenuated because of the coupling to the superficial plasmon modes, whereas the TE modes are transmitted. It should be mentioned that some of the waveguides in anisotropic crystals, such as proton-exchanged waveguide in $LiNbO_3$, can be used as a waveguide polarizer. Polarization beam splitters can be realized, for example, by careful design of DC or intersecting waveguide (**Figure 19(h)**) in anisotropic crystal. Another method to implement polarization elements is to form a narrow and deep groove perpendicular to the axis of a channel waveguide and insert in

the groove a thin-film-type polarizer, half-wave plate, or quarter-wave plate with appropriate orientation.

3.05.4.2 Passive Waveguide Grating Components

Examples of passive waveguide grating components are illustrated in Figure 20. They include input/output coupler, deflector/polarization splitter, focusing grating coupler (FGC), grating lens, mode converter, and reflector/wavelength filter. The grating structures are fabricated by patterning based on photolithography, electron beam lithography, or holographic interference recording, associated with dry etching to form the grating grooves. Other fabrication techniques include holographic interference recording or holographic contact printing using a phase mask in waveguides of a photopolymer or a ultraviolet (UV)-light-sensitive material such as Ge-doped silica, and direct electron beam writing, to produce periodic modulation in the refractive index. Periodic structures for IO include fiber gratings, which are fabricated by UV-light-induced refractive index change of an optical fiber.

Various wavefront conversion functions based on the principle of holography can be incorporated in mode coupling in a planar waveguide by spatially modulating the grating pattern (forming a curved and chirped grating). A useful wavefront conversion is the lens function for focusing, collimating, and imaging. Waveguide grating elements with lens function are illustrated in Figure 20(f)–(h). Suppose that $\Phi_i(y, z)$ and $\Phi_o(y, z)$ are the phase distribution functions, on the yz waveguide grating plane, of the incident wave and the desired output wave, respectively. The desired wavefront conversion is accomplished by giving the input wavefront a phase modulation corresponding to the phase difference $\Delta\Phi = \Phi_o - \Phi_i$. The waveguide grating (waveguide hologram) for such phase modulation consists of grating lines described by

$$\begin{aligned}\Delta\Phi(y, z) &= \Phi_o(y, z) - \Phi_i(y, z) \\ &= 2m\pi (m = \cdots, -2, -1, 0, +1, +2, \cdots)\end{aligned} \quad [27]$$

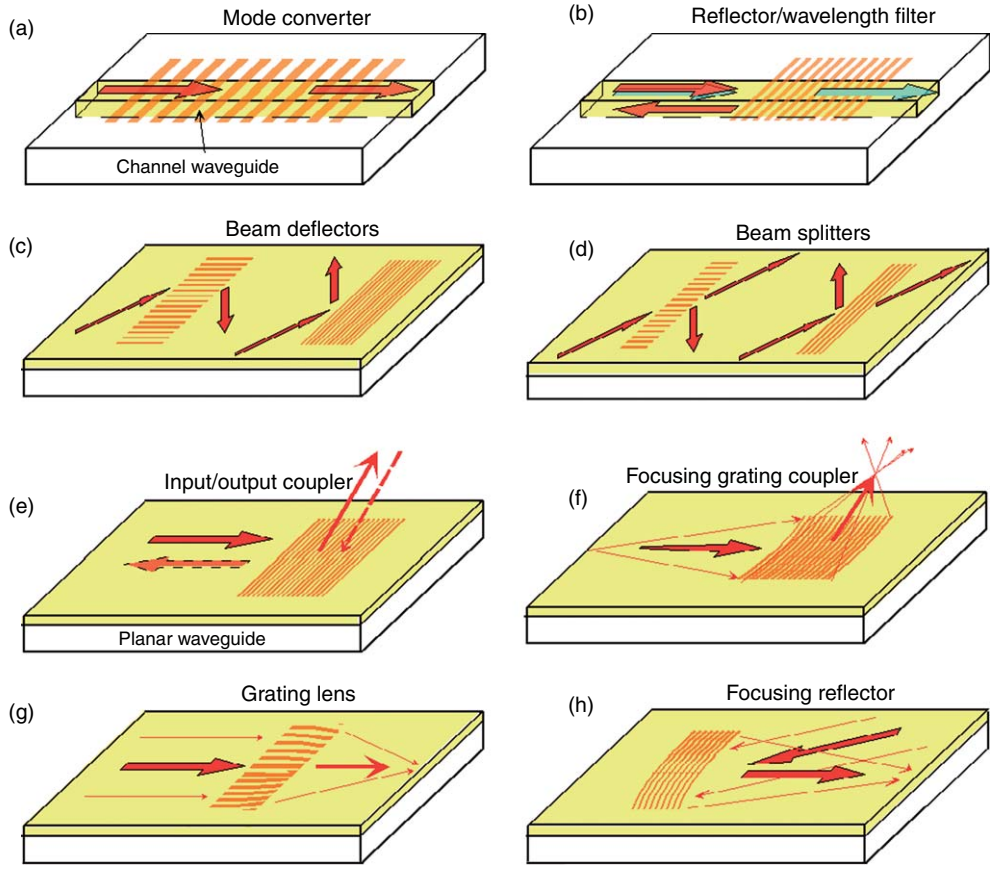


Figure 20 Various passive waveguide grating components.

The grating patterns can be generated and fabricated by computer-controlled electron beam writing technique.

3.05.4.3 Waveguide Lenses

One of the most elementary and important passive element in classic optics is a lens, which offers the functions of beam focusing, imaging, and Fourier transforming. Components that perform the lens functions in a planar waveguide are waveguide lenses. Waveguide lenses can be implemented simply by depositing thin transparent film (or ion exchange) of parabolic convex or concave pattern on a planar waveguide to change the mode index in the pattern region. This type of waveguide lens is called mode index lenses (**Figure 21(a)**), and the working principle is based on refraction. It is difficult to obtain mode index lens of short focal length, as the available index change is small.

One of the important mode index lenses is a Luneburg lens (**Figure 21(b)**), which is formed by depositing a transparent thin film with a rotation-symmetrical graded thickness on a planar waveguide. The structure realizes a graded mode index distribution designed carefully for aberration-free lens function. Another important waveguide lens is the geodesic lens (**Figure 21(c)**), which is constructed by forming a rotation-symmetrical depression on the surface of a substrate and forming a planar waveguide on the deformed surface. The ray of the guided wave changes the direction of propagation on the depression according to Fermat's principle. Although good performances can be obtained by appropriate design using Luneburg and geodesic lenses, their fabrication requires special techniques incompatible with the planar process.

Another class of waveguide lenses is diffraction lenses. It was presented in the previous subsection that waveguide holographic grating components (chirped grating lens, FGCs, etc.) can perform lens functions. A waveguide lens based partly on diffraction and partly on refraction is micro Fresnel lens (**Figure 21(d)**), which results when a convex mode index lens is segmented into zones and rectangular areas useless for lens function are deleted. The advantage of this lens is that they can be fabricated by planar lithographic process. The large wavelength dispersion of the focal length and the chromatic aberration may limit their application.

3.05.4.4 Arrayed Waveguide Gratings

A versatile device developed as wavelength multi/demultiplexers for dense wavelength division multiplexing (DWDM) photonic network systems is arrayed waveguide grating (AWG) (Smit 1988, Takahashi *et al.* 1990). The device consists of a large number of arrayed curved channel waveguides and two planar waveguides connecting them together, as shown in **Figure 22**. The optical wave transmitted through the input channel guide diverges in a sector region in the first planar part and is then divided and coupled in phase into the channel guides through the planar/channel junctions aligned on the curved (circular arc) wavefront of the diverging wave. The lengths of the channels are designed in such a manner that the optical pass difference between adjacent channel guides coincides with an integer multiple of the center wavelength λ_0. Therefore, the optical wave at λ_0 emerges in phase from the end of the channels aligned on an arc where the channels are connected to the second planar part. These wavelets form a wavefront converging into the junction to the

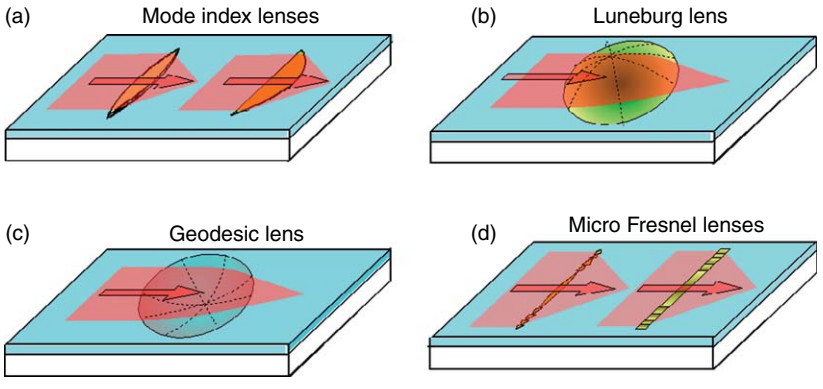


Figure 21 Various waveguide lenses.

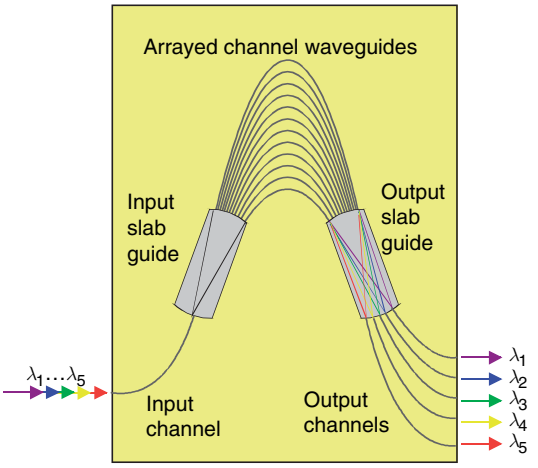


Figure 22 Schematic of an arrayed waveguide grating (AWG).

central output channel. For optical waves with wavelength deviation from λ_0, the pass difference deviates from the integer multiple of the wavelength. This gives rise to phase shifts and wavefront inclination at the planar/channel junction, and therefore, the optical wave is transmitted into another output channel depending on the wavelength deviation.

The AWGs have several advantages such as high resolution, low insertion losses, possibility of multiport input, and high degree of design flexibility. Many AWGs using silica waveguides have been developed and put into practical use. The channel number ranges from 256 channels for one-stage AWGs to 1000 channels for two-stage AWGs (Hibino 2002).

3.05.5 Active and Control Devices

3.05.5.1 Semiconductor Lasers

Many IO devices require a laser element as the coherent light source. Semiconductor LDs are well suited as a light source for IO. In the earlier hybrid IO devices, FP-type LDs were combined with dielectric waveguide components. The FP LDs, however, have problems of the poor controllability of lasing wavelength and the longitudinal mode instability such as mode hopping and multimode lasing.

The problems can be eliminated by using distributed feedback (DFB) LDs or DBR LDs (Carroll *et al.* 1998, Coldren and Corzine 1995, Suhara 2004). The fundamental structures of DFB and DBR LDs are illustrated in **Figure 23(a)** and **23(b)**, respectively. In the DFB LDs, wavelength-selective feedback is accomplished by integrating a grating structure in the active region. In the DBR LDs, grating(s) at one or both sides of the active region provides wavelength-selective feedback. The sharp wavelength selectivity of gratings assures the single longitudinal mode lasing. With appropriate design, DFB and DBR LDs maintain the single longitudinal mode lasing even under high-speed direct modulation. The lasing wavelength can be determined accurately by the grating period. QW structures with narrow gain bandwidth, lower transparency current density, and other advantages are used as an active layer to obtain high laser performances. The DFB and DBR lasers have another advantage that there is no need to form facet mirrors. This is an essential requirement for LDs for monolithic IO devices. Therefore, QW DFB and DBR LDs are now used in most IO devices.

Semiconductor optical amplifiers (SOAs) (Coldren and Corzine 1995, Suhara 2004) are implemented by depositing antireflection (AR) coating on both facets of an FP LD. SOAs can be integrated to boost the output power of an LD, amplify a weak input optical signal, or compensate for the power losses in the passive components. Another important application of SOA is the use as an optical gate switch that controls the transmission of a lightwave by binary control of the injection current.

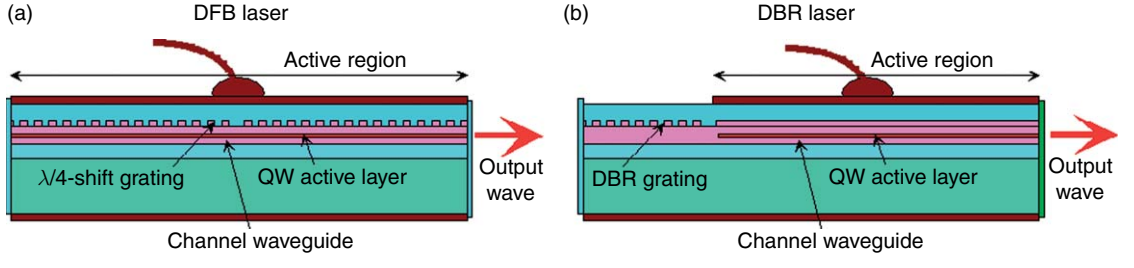


Figure 23 Schematic of fundamental cross-sectional structures of distributed feedback (DFB) and distributed Bragg reflection (DBR) lasers. QW, quantum well.

3.05.5.2 Guided Wave Photodetectors

Some IO devices require photodetectors for optical-to-electrical signal conversion. Ordinary PDs such as Si PIN (positive-intrinsic-negative) PD, Si-avalanche photo diode (Si-APD), and InGaAs PIN PD can be used by butt coupling to the waveguide. Waveguide PDs, in which the optical wave propagates along the junction interface, were also developed for monolithic integration and achieving better performances such as a large bandwidth–efficiency product (Kato 1999). PDs can be fabricated on the surface of a Si substrate for various waveguides. The guided wave coupling is accomplished by forming a taper in the SiO_2 buffer layer so that the evanescent tail overlaps with the Si PD, or using an integrated grating coupler (Suhara and Nishihara 1986). For Group III–V semiconductor waveguides, a structure similar to the active region of LDs, or actually LD itself without the feedback element, with bandgap energy control, can be used as a PD by applying appropriate back-bias voltage (Takeuchi *et al.* 1996), although an appropriate groove structure must be fabricated for electrical isolation (Hirata *et al.* 1995).

3.05.5.3 Electrooptic Devices

An effective means for high-speed dynamic control of optical wave by electric signals is to utilize the EO effects. Dielectric properties of a transparent crystal are represented by a dielectric tensor $[\boldsymbol{\varepsilon}]$, which correlates electric field and displacement vectors with dielectric constant of vacuum ε as $[\mathbf{D}] = \varepsilon[\boldsymbol{\varepsilon}][\mathbf{E}]$. The off-diagonal elements of the 3×3 tensor $[\boldsymbol{\varepsilon}]$ vanish when the coordinate axes are taken along the crystallographic optical principal axes. For optical fields, the diagonal elements are correlated with refractive indexes $\varepsilon_{11} = n_1^2, \varepsilon_{22} = n_2^2$, and $\varepsilon_{33} = n_3^2$,. For widely used uniaxial crystals, $n_1 = n_2 = n_0$ and $n_3 = n_e$, where n_0 and n_e are ordinary and extraordinary indexes, respectively. The crystals exhibit birefringence, i.e., a different refraction for ordinary and for extraordinary rays. The indexes for the ordinary and extraordinary waves propagating in the crystal in an arbitrary direction are determined by using an index ellipsoid that has radii n_o in the x- and y-directions and n_e in the z-direction. When an external electric field $\mathbf{E}^e$ is applied to the crystal through electrodes, the optical properties are slightly changed. The EO effects are described by deformation of the index ellipsoid. The deformation proportional to $\mathbf{E}^e$ is called Pockels effect and that proportional to the square of $\mathbf{E}^e$ is called Kerr effect. For Pockels effect, the deformation is given by $[\Delta(1/\mathbf{n}^2)] = [\mathbf{r}][\mathbf{E}^e]$, where $[\mathbf{r}]$ is the EO (Pockels) tensor consisting of 6×3 elements (some of them vanish because of the crystal symmetry). For example, when an external electric field of magnitude E_z^e is applied along the crystallographic $z(c)$-axis of a uniaxial crystal, the index for the ordinary wave ($\mathbf{E}$ perpendicular to the z-axis) and the index for the extraordinary wave ($\mathbf{E}$ parallel to the z-axis) change by

$$\Delta n_0 = -\frac{n_o^3 r_{13} E_z^e}{2}, \quad \Delta n_e = \frac{n_e^3 r_{33} E_z^e}{2} \qquad [28]$$

respectively.

The representative materials for EO devices are ferroelectric crystals $LiNbO_3$ and $LiTaO_3$. Various EO modulators/switches are implemented by combining channel waveguides and planar electrodes on the crystals as shown in **Figure 24**. The simplest device is phase modulators (**Figure 24(a)**). The refractive index change caused in the waveguide at the electrode gap by the driving signal gives rise to phase shift of the guided wave. The simplest waveguide intensity modulator is cutoff-type modulators (**Figure 24(b)**). An input and an output channel guide are aligned on line with short gap and electrodes are aligned at both sides of the gap. Without applied voltage, optical wave is not transferred through the gap. Voltage application gives rise to index increment that connects the channels, so that the optical wave is transmitted. Waveguide EO devices with better performances are MZI type (**Figure 24(c)**) and DC type (**Figure 24(d)**). In the MZI type, an MZI is constructed by branching waveguides (or 3 dB DCs) and parallel arm channels. Differential phase shifts are given to the guided waves propagating in each arm by the EO index change by the electric signal applied to the electrodes on the arms. The phase modulation is converted into intensity modulation by interference effect. In the DC type, a 0-dB DC is constructed by two identical channel guides aligned parallel with small gap, and electrodes are formed on the coupler. Without applied voltage, the optical wave is transmitted from one guide to another (cross state). With electric signal applied, the differential index change gives rise to the phase mismatch between channels and therefore to reduction in coupling. At a voltage, the light is

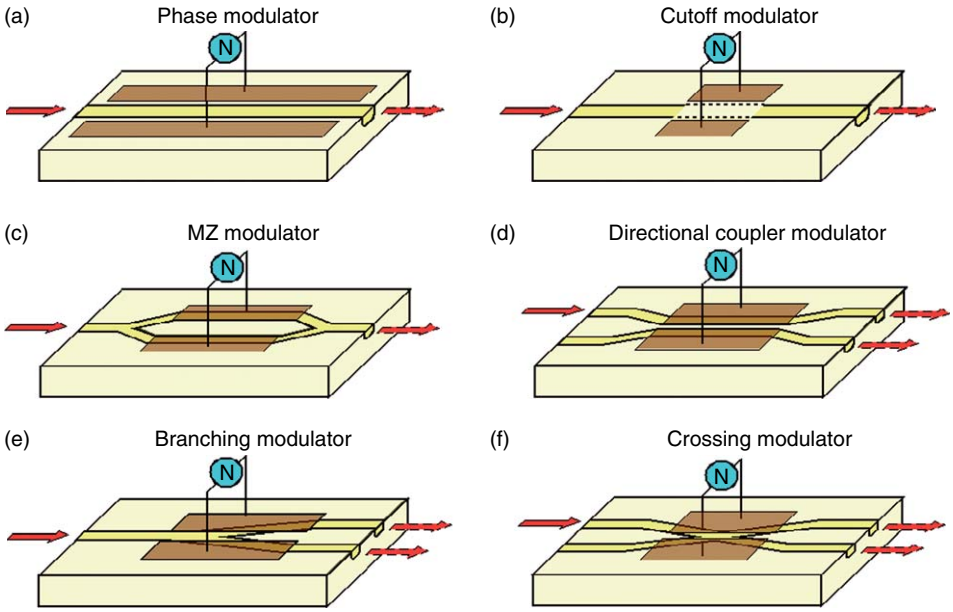


Figure 24 Various electrooptic waveguide optical modulators/switches. MZI, Mach–Zehnder interferometer.

not coupled but transmitted through the input channel (parallel state). A modification of the electrode configuration to overcome the problem that deviation of the coupling length from the complete coupling length due to fabrication errors leads to incomplete cross state is the reversed $\Delta\beta$ configuration (Alferness 1982, Kogelnik and Schmidt 1976, Thylén 1988), where one of the electrodes is split into halves and voltages of opposite polarity are applied to them. Then complete cross and parallel states can be realized by adjusting the voltage. Waveguide EO modulator/switch configurations that may allow shorter device length include those using a branching waveguide (**Figure 24(e)**) and those using a crossing waveguide (**Figure 24(f)**). In the former the branching ratio is controlled, and in the latter the TIR is controlled. Other EO devices include EO grating devices using interdigital electrodes on a planar waveguide and TE–TM mode converters using interdigital electrodes on a channel waveguide.

The EO effect is an ultrafast phenomenon originating from the change of electron states. The time response of EO devices is dominated by the driving circuit parameters and the distributed interaction of optical and RF waves. In EO devices with electrodes acting as a lumped capacitor, the modulation bandwidth and the driving voltage must be traded off, as they are inversely proportional to the electrode length L. An effective method to enhance the bandwidth without the expense of the voltage is to use the traveling wave configuration, where the modulating RF signal propagates in the same direction as the lightwave along the electrodes designed as a transmission line (Izutsu *et al.* 1977, Thylén 1988). Then the bandwidth is limited by the mismatch in the group velocities between the optical and the RF waves. It is possible to achieve a wide bandwidth and a low driving voltage simultaneously. $LiNbO_3$ waveguide phase and intensity modulators of $\sim$5 V driving voltage and up to 40 Gbits s^{-1} bandwidth are commercially available. High-speed EO polymer waveguide modulators have also been developed (Girton *et al.* 1991).

EO devices can be implemented also in Group III–V semiconductor waveguides. The Kerr effect is utilized, as the Pockels effect is small in semiconductors. In principle, EO devices with the structure and working principle similar to that of ferroelectric waveguide EO devices can be implemented, although the planar electrodes must be replaced by those of pn junction or Schottky barrier. It is also possible to implement modulator/switch using the refractive index change associated with the plasma effect caused by carrier injection, although the response is limited by the carrier lifetime (in the order of nanoseconds) and is not very fast. More effective method to implement optical intensity modulators and switches with simple device configuration is to utilize electroabsorption (EA). When an electric field is applied to a semiconductor bulk crystal, the valence and conduction band edges are inclined, and the electron wave functions are

extended slightly into the forbidden band. As a result, the effective bandgap energy is decreased, and light with a photon energy slightly smaller than the original bandgap energy is absorbed. This phenomenon is well known as the Franz–Keldish (FK) effect. In a semiconductor QW structure, the FK effect is enhanced by the electron wave confinement in the QW. In addition, absorption of light with photon energy close to the bandgap energy is significantly affected by the exciton that has a binding energy much larger than that in bulk semiconductor, and the exciton absorption spectrum is sensitively changed by the application of electric field. The total EA effect of the enhanced FK and field-dependent exciton absorption is called quantum-confined Stark effect (QCSE). By appropriate design of the QW optimized for an optical wavelength, large EA can be obtained. The absorption change is associated with the refractive index change, due to the Kramers–Kronig relation. EA modulators can be implemented simply by combining a QW waveguide with PIN-type electrodes. High-performance EA modulators of fast response (>40 GHz bandwidth), low driving voltage (<5 V), and small dimension (<1 mm) have been developed using the QCSE in multiple QW. A drawback is that the good performances are obtained only in a narrow wavelength range in the infrared region.

Nematic liquid crystals (NLCs) exhibit unique EO effect. An NLC cell is a thin layer of NLC sandwiched between two parallel substrates with narrow (several micrometers) gap. The substrate surface is treated for alignment of NLC molecule axis along a direction in the substrate surface. Because the NLC molecules are long, NLC in the cell exhibits large optical anisotropy ($|n_e - n_o|$ as large as ~ 0.1). The NLC molecular axis can be rotated within the plane of the substrate by applying an electric field through electrodes, to cause a large change (maximum is $|n_e - n_o|$)in the index for extraordinary ray with a fixed propagation direction. Waveguide optical switches are constructed, for example, by forming an NLC cell with electrodes on a crossed waveguide. The device configuration is similar to that in **Figure 24(f)**. Switch operation is accomplished by controlling the TIR at the intersection. The device has the advantage of low driving voltage, although the response is not very fast (millisecond or submillisecond range).

3.05.5.4 Thermooptic Devices

An effective method to implement lightwave control devices with waveguides of amorphous materials such as glasses and polymers is to utilize the TO effect. Because the refractive index of optical medium depends on the temperature, local heating of an optical waveguide channel gives rise to change in effective index of refraction, with which lightwave control can be accomplished. Practical TO devices can be implemented by replacing the electrodes of the EO devices shown in **Figure 24** by microheaters consisting of patterned thin metal film for Joule heating. The temperature rise and the index change are determined by the balance between heat generation and sinking, and they are proportional to the electric power fed to the heater. Many polymer materials exhibit a negative TO coefficient (temperature coefficient of index), and the magnitude is much (one order of magnitude) larger than that (positive) of silica. Therefore, polymer waveguide TO devices have the advantage of low driving power. Better stability and faster time response are obtained in TO devices using silica waveguides on an SiO_2/Si substrate. Although the response is much slower than that of EO devices, millisecond (or even submillisecond) response can be obtained by appropriate device design.

3.05.5.5 Acoustooptic Devices

When an optical medium is subjected to mechanical stresses, changes in the refractive index and birefringence are induced (photoelastic effect). The effects caused by stresses associated with an acoustic wave propagating in the medium, called AO effects, provide an effective means of controlling optical waves (Nishihara *et al.* 1989, Suhara and Nishihara 1986, Tsai 1979). Although the response of AO control is not as fast as that of EO control, unique functions such as variable angle optical deflection and optical frequency shifting can be implemented in AO devices.

Acoustic wave is a wave in the form of periodic modulation in the displacement of the medium and is expressed by the strain tensor [**S**] given by the space derivatives of the displacement vector. The stress caused by the strain is described by stress tensor [**T**], which is correlated linearly to [**S**] through Fick's law. From this relation and the equation of motion, a wave equation is derived. The equation shows that the periodic modulation propagates in the medium with a velocity v. The acoustic wavelength Λ and the frequency f are correlated by $\Lambda f = v$. In the vicinity of the surface of a medium, a surface acoustic wave (SAW) can propagate. The

amplitudes of displacement, stress, and strain of SAW are a function of the depth from the surface, and the wave is localized in a surface layer with depth comparable to the SAW wavelength Λ. The amplitudes are proportional to the square root of the SAW power density. The description of AO effects is given as the deformation of the index ellipsoid by the strain [**S**]. The deformation is correlated with [**S**] by $[\Delta(\mathbf{1}/\boldsymbol{n}^2)] = [\mathbf{p}][\mathbf{S}]$, where [**p**] is a fourth-order tensor called photoelastic tensor. The index change is proportional to the stress amplitude and is proportional to the square root of the SAW power density.

Waveguide AO devices make use of the AO effect on an optical guided wave by a SAW. They are classified into coplanar and collinear types. **Figure 25** illustrates a basic structure of coplanar AO device. It consists of a planar optical waveguide and a SAW transducer with interdigital electrodes on a substrate of ferroelectric crystal such as $LiNbO_3$. By driving the interdigital transducer (IDT) with an RF signal of frequency f, a SAW of the same frequency and a wavelength $\Lambda = v/f$ is excited through the piezoelectric effect. The SAW induces a traveling periodic modulation in the refractive index by the AO effect, and this temporal traveling grating gives rise to the coplanar diffraction of the optical guided wave. The Q parameter is defined as $Q = K^2L/\beta = 2\pi\lambda L/n_e\Lambda^2$, where λ and n_e are the wavelength and the mode index of the optical guided wave, respectively, and L is the SAW aperture. For cases where L is small or Λ is large so that $Q < 1$, the Raman–Nath diffraction takes place. The diffraction efficiency, however, cannot be very high. Devices of $Q > 10$ (large L or small Λ) are more important, as the Bragg diffraction characterized by a diffracted beam of single order with high efficiency can be obtained. Such devices are called waveguide AO Bragg cell. The propagation direction of the incident guided wave must be adjusted so as to satisfy the Bragg condition described by the closed triangle formed by the wave vectors of the incident and the diffracted optical waves and the SAW wave vector (the grating vector). The Bragg angle is given by $\theta_B = \sin^{-1}(K/2\beta) = \sin^{-1}(\lambda/2n_e\Lambda) \approx \lambda/2n_e\Lambda$. The deflection angle is given by $2\theta_B \approx \lambda/n_e\Lambda = \lambda f/n_e v$ and is approximately proportional to the frequency f of the input RF signal.

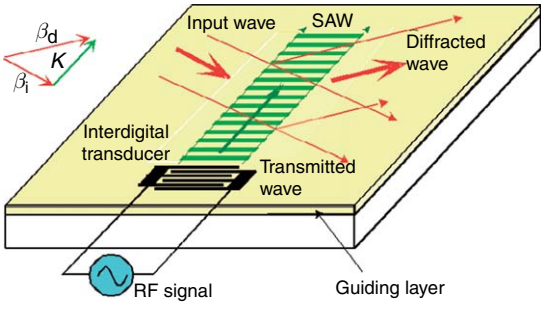


Figure 25 Waveguide acoustooptic Bragg cell. RF, radio frequency; SAW, surface acoustic wave.

The diffraction efficiency can be written as

$$\eta = \sin^2\{c\sqrt{P_s L}\} \qquad [29]$$

where P_s is the SAW power that is proportional to the RF power fed into the IDT, and c is a constant that depends on the material constants, optical wavelength, diffraction angle, and the overlap between the optical and the SAW profiles. Nearly 100% diffraction efficiency is obtained by giving appropriate RP power. In the low-efficiency region, the diffracted beam power increases approximately in proportion to the input RF power. The simplest application of the waveguide Bragg cell is its use as an AO modulator. The SAW IDT is driven by an RF signal of a constant carrier frequency with power modulated by a signal to modulate the diffracted beam intensity. Because the AO grating is traveling at a velocity v, the frequency of the diffracted wave is shifted by $\pm f$ due to the Doppler effect. This means that the AO Bragg cell can be used as an optical frequency shifter. More important application is its use as an AO deflector. The RF f is swept to scan the beam deflection angle. Because the high diffraction efficiency is obtained only at and near the Bragg condition, and the RF change gives rise to deviation from the Bragg condition, AO Bragg cells using a simple standard IDT as shown in **Figure 25** perform well only in a narrow RF region. In order to obtain a wider frequency bandwidth and deflection angle, special IDT design is required to accomplish wide SAW excitation bandwidth as well as RF impedance matching and maintain the Bragg condition. An effective method is the frequency-controlled SAW beam steering using multiple tilted array IDT, curved-finger IDT, phased array IDT, and tilted-finger chirped IDT (Tsai 1979). Bandwidth as large as 1 GHz has been obtained in $LiNbO_3$ waveguide AO Bragg cell. The temporal signal modulating the RF carrier is converted into spatial distribution by the SAW traveling effect. Therefore, a wideband Bragg cell can also be used as a spatial light modulator (SLM).

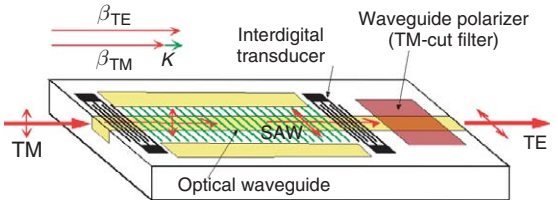


Figure 26 Waveguide acoustooptic tunable wavelength filter. SAW, surface acoustic wave; TE, transverse electric; TM, transverse magnetic.

Figure 26 illustrates a typical waveguide AO device of collinear type. It consists of a channel optical waveguide and a SAW IDT. A SAW waveguide structure for confinement in a narrow strip region may be formed to reduce the driving power. Collinear AO interaction in an anisotropic waveguide may give rise to coupling between TE and TM modes through the off-diagonal dielectric tensor element induced by AO effect. The phase-matching condition is given by

$$\begin{aligned} |\beta_{TE} - \beta_{TM}| &= \frac{2\pi}{\lambda}|N_{TE} - N_{TM}| \\ &= K = \frac{2\pi}{\Lambda} = \frac{2\pi f}{v} \end{aligned} \quad [30]$$

where β and N denote the propagation constant and the mode index, respectively. Efficient mode conversion is obtained only when the phase-matching condition is satisfied, and the conversion efficiency for interaction length L is written as

$$\eta = \sin^2\{\gamma\sqrt{P_s L}\} \quad [31]$$

An important application of AO mode converters is their use as an optical wavelength-tunable filter. Mode conversion takes place only when the phase-matching condition is satisfied at least approximately. When SAW frequency f is fixed at the value satisfying the phase-matching condition for an optical wavelength λ, a deviation in wavelength $\Delta\lambda$ results in a deviation from the phase-matching condition. The wavelength selectivity 3 dB full width is given approximately by $2\Delta\lambda/\lambda \approx 0.80\Lambda/L$. When a wavelength-selective mode converter is combined with a polarizer (or polarization beam splitter), the device works as a wavelength filter. The center wavelength can be tuned by changing the SWA frequency f. The tuning range is limited mainly by the SAW IDT bandwidth.

Although AO devices using piezoelectric (ferroelectric) crystal waveguides are described above, it is possible to construct AO devices with various material combinations, as all optical materials exhibit AO effect. They include piezoelectric waveguide on piezoelectric substrate (e.g., Ti:$LiNbO_3$), nonpiezoelectric film waveguide on piezoelectric substrate (e.g., As_2S_3 on $LiNbO_3$), piezoelectric film waveguide on nonpiezoelectric substrate (e.g., ZnO on SiO_2), and nonpiezoelectric waveguide/substrate with piezoelectric film for IDT (e.g., As_2S_3 on SiO_2/Si with ZnO IDT, and GaAs/AlGaAs/GaAs and ZnO IDT).

3.05.5.6 Nonlinear Optic Devices

NLO effects provides various unique optical functions such as optical wavelength conversion and ultrafast all-optical switching (Suhara and Fujimura 2003). The earlier R&D work on waveguide NLO devices was motivated by the need for implementation of compact short-wavelength coherent light source for optical memories. Waveguide NLO devices have also found important applications as an all-optical signal wavelength converter for WDM photonic network systems. Application of waveguide NLO devices includes ultrafast all-optical signal processing for high-speed photonic communications systems.

Figure 27(a) illustrates a fundamental waveguide quasiphase-matching (QPM)–NLO device structure. An important requirement for efficient NLO interaction for second harmonic generation (SHG), sum frequency generation (SFG), difference frequency

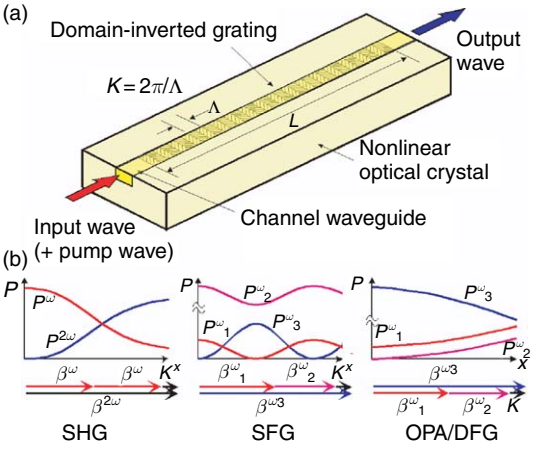


Figure 27 (a) Fundamental device configuration of waveguide quasiphase-matched (QPM) nonlinear optic (NLO) devices. (b) Optical power distributions and phase-matching diagrams for second harmonic generation (SHG), sum frequency generation (SFG), and optical parametric amplification/difference frequency generation (OPA/DFG).

generation (DFG), and optical parametric amplification (OPA) is phase matching. QPM is a technique that uses a structure with sign inversion of period Λ in the NLO coefficient along the propagation axis to compensate for the difference between the wave vectors β of the NLO polarization and the wave to be generated by the grating vector $\mathbf{K}(|\mathbf{K}| = 2\pi/\Lambda)$. The QPM condition can be described by a wave vector diagram as shown in **Figure 27(b)**, and the interaction is described by coupled mode equations. Spatial variations of the interacting wave powers are illustrated in **Figure 27(b)**.

The outstanding advantage of QPM is that the phase matching can be accomplished for any combinations of arbitrary wavelengths in the transmission region of the NLO crystal only by appropriate choice of the period, and therefore, very wide wavelength region can be covered by a single NLO crystal. Another advantage is that the maximum NLO tensor element can be exploited. By combining a QPM structure and a channel waveguide structure for confining optical waves to enhance the power densities, highly efficient and compact NLO devices of various types can be implemented. The QPM structures can be formed by periodic inversion of ferroelectric domain in NLO crystals such as $LiNbO_3$ and $LiTaO_3$. The period range is 2–20 μm for most applications. A high-voltage pulse is applied across the crystal using periodic electrodes to give electric charge corresponding to the value required to invert the spontaneous polarization.

For SHG ($\omega \rightarrow 2\omega$) under the QPM condition, the harmonic output power $P^{2\omega}$ increases with the input fundamental power P_0^{ω} approximately as $P^{2\omega} = \left(P_0^{\omega}\right)^2 \kappa_{\mathrm{SHG}}^2 L^2$, where L is the interaction length and κ_{SHG} the coupling coefficient. For SFG (ω_1, $\omega_2 \rightarrow \omega_3 = \omega_1 + \omega_2$) under the QPM condition, the output power P^{ω_3} increases approximately in proportion to $P_0^{\omega_1} P_0^{\omega_2} \kappa_{\mathrm{SHG}}^2 L^2$ in the range of small input power, while for large $P_0^{\omega_1} P_0^{\omega_2} \kappa_{\mathrm{SHG}}^2 L^2$ the interaction turns from SFG to DFG on the way of interaction region so that the output power P^{ω_3} decreases. For DFG/OPA ($\omega_1, \omega_3 \rightarrow \omega_2 = \omega_3 - \omega_1$), the difference frequency (ω_2) wave (idler wave) is generated and the input signal (ω_1) wave may be amplified. In either interaction, the wavelength bandwidth is inversely proportional to L. The SFG and OPA/DFG interactions follow the Manlay–Rowe relation, which correlates the number of photons (e.g., for SFG, the number of the generated ω_3 photons equals to the number of annihilated ω_1 photons and the number of annihilated ω_2 photons).

Many waveguide QPM SHG devices were developed for generating SH waves in wavelength region covering the visible and near-UV region. For $LiNbO_3$ waveguide QPM blue SHG generation devices, normalized conversion efficiency higher than 1500% W^{-1} has been achieved, and blue SHG laser modules have been implemented by hybrid integration with a pump semiconductor laser. Waveguide QPM DFG devices have been extensively studied as an all-optical wavelength converter for WDM photonic network, and normalized conversion efficiency higher than 1500% W^{-1} has been achieved. Several ultrafast all-optical signal processing devices using $LiNbO_3$ QPM waveguide were also reported. They include SFG device for picosecond optical sampling and picosecond SFG interferometer optical gate switches.

3.05.6 IO Devices and Applications

IO devices are implemented by integrating many components of same function, or by integrating components of different functions, to realize a higher function for a specific application. Integration schemes include hybrid integration using dielectric or ferroelectric waveguide device and butt-coupled semiconductor lasers and monolithic integration using semiconductor lasers. In the following, examples of IO devices for applications in several areas are discussed.

3.05.6.1 Signal Processing

The first application of the integrated optics concept was as an integrated optical correlator incorporating SLMs and lenses in a planar waveguide (Shubert and Harris 1968). Practical applications of RF signal processing devices were developed by using a waveguide AO Bragg cell as the SLM. The time-to-space conversion function of waveguide AO Bragg cells is utilized in combination with various types of waveguide lenses for Fourier transforming and/or imaging to develop various RF signal processing devices. Developed devices include convolvers and correlators of space-integrating type and time-integrating type (Tsai 1979).

IO RF spectrum analyzers (IOSAs), in particular, were extensively studied for their practical application to high-speed real-time radar signal analysis. Several types of IOSAs using wideband AO Bragg cells with Luneburg, geodesic (Mergerian *et al.* 1980), grating, and Fresnel waveguide lenses for Fourier

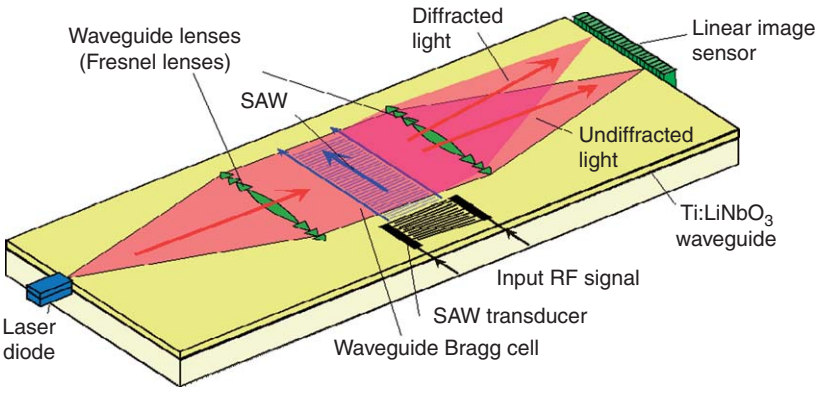


Figure 28 An integrated optic radio frequency (RF) spectrum analyzer using a waveguide acoustooptic Bragg cell and waveguide Fresnel lenses. SAW, surface acoustic wave. (Source: Suhara T, Fujiwara T, Nishihara H 1986 Proton-exchanged Fresnel waveguide lenses in Ti:$LiNbO_3$ waveguides. *Appl. Opt.* **25**, 3379–83.)

transforming were developed. **Figure 28** illustrates an IOSA using a waveguide Fresnel lens (Suhara *et al.* 1986). The RF signal to be analyzed is amplified and fed into the wideband IDT of the Bragg cell to deflect and spatially modulate the incident optical guided beam. The diffracted guided beam is then Fourier transformed by the waveguide lens and the intensity distribution on the focal line is converted to electrical signal by a linear image sensor, which gives the output spectrum signal. Typical parameters are 4 MHz resolution, 1 GHz bandwidth, and 1 μs response (analysis) time. Devices evolved from the IOSA include IO scanner for optical printer heads.

IO devices for signal processing were implemented by using EO control elements also. Many devices for high-speed electrical signal processing, such as analog-to-digital converters, analog signal multipliers, matrix–vector multipliers, and digital processors for correlation, parity generation, and error detection/correction, were implemented by integrating EO grating spatial modulators, or by integrating EO MZIs with modified electrode configurations (Taylor 1987).

3.05.6.2 Optical Communications

Optical communications have been the most important area of application of IO. Various IO devices were studied and developed. Many of them have been commercialized and put into practical use. An example of simple IO devices is $1 \times N$ couplers and $N \times N$ star couplers, which are required in multiple-access optical networks to evenly distribute optical signals to many receivers and to interconnect the transmitters and receivers. The $1 \times N$ couplers are constructed by cascading Y-branching waveguides, and the $N \times N$ star couplers are constructed by inserting a slab mixing waveguide between fan-shaped input and output channel guide arrays.

Optical switches for dynamic control of optical signal path are one of the key devices required for constructing photonic network systems, including recent advanced WDM systems using optical add/drop multiplexers (OADMs) and optical cross connects (OXCs). High-speed EO switch matrixes, such as a nonblocking 8×8 crossbar switch matrix consisting of 64 EO DCs, were implemented using $LiNbO_3$ waveguides (Thylén 1988). Several types of large-scale IO switches, such as multiple cascaded 2×2 switches, $1 \times N$ switches, and $N \times N$ switch matrixes, have been developed by using silica waveguide TO MZI switches as the fundamental building unit. The millisecond response of TO switching is appropriate for many applications in OXC and OADM systems. A high extinction ratio (>60 dB) can be achieved by cascading the MZI units. Nonblocking $N \times N$ switch matrix as large as 16×16 has been implemented (Goh *et al.* 2001) using 512 MZIs. The switch element and the waveguide layout configuration are shown in **Figure 29**. Polymer $1 \times N$ switches were also implemented by integrating Y-branching waveguide TO switches (Ooba *et al.* 2000). It should also be mentioned that many optical MEMS, free-space and fiber optic types, including optical switches, variable attenuators, tunable filters, add/drop multiplexers, gain equalizers, dispersion compensators, and polarization controllers, were developed for application to optical networks (Giles *et al.* 1999, Lin and Goldstein 2002, Motamedi 1994).

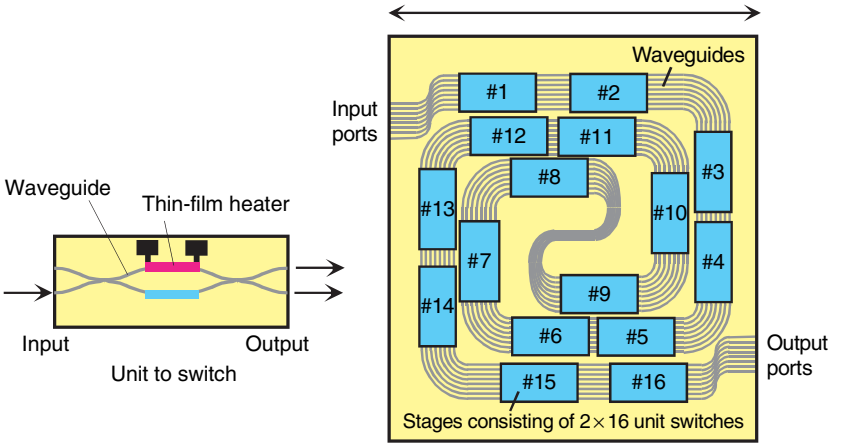


Figure 29 The unit thermooptic Mach–Zehnder interferometer switch element and the waveguide layout configuration of a nonblocking 16 × 16 switch matrix using 512 switches. (Source: Goh T, Yasu M, Hattori K, Himeno A, Okuno M, Ohmori Y 2001 Low loss and high extinction ratio strictly nonblocking 16×16 thermooptic matrix switch on 6-in wafer using silica-based planar lightwave circuit technology. *J. Lightwave Technol.* **19**, 371–9.)

WDM, proposed as an effective means of drastically enhancing the capacity of data transmission through an optical fiber, requires wavelength multiplexers and demultiplexers at the transmitter and the receiver terminals. Development of high-performance compact multi/demultiplexers has been a major subject of R&D of IO devices for optical communications. The earlier work was implementation of wavelength multi/demultiplexers using grating components as a dispersion element (Nishihara *et al.* 1989, Suhara *et al.* 1986). Waveguide demultiplexers using a curved grating, a grating and a waveguide lens, and a micrograting array were studied. **Figure 30** illustrates a demultiplexer constructed by a patterned ion-exchanged waveguide and a chirped grating butt-coupled to the waveguide (Suhara *et al.* 1982). **Figure 31** illustrates a practical IO device developed for WDM access network systems using the 1.3- and 1.5-μm bands for bidirectional signals and broadband signals (Himeno *et al.* 1998, Uchida *et al.* 1997). A silica Y-branching waveguide, a thin-film wavelength filter, an LD, and waveguide PDs are hybrid integrated on a Si platform to construct a transceiver module. The semiconductor devices were flip-chip bonded by the passive alignment technique.

Higher performances and advanced functions were accomplished by using the AWG as a key component. Development of the AWG includes modification for use as an $N \times N$ wavelength-selective star coupler, cancellation of the polarization dependence by inserting a half-wave plate, athermalization by inserting polymer section to compensate for the temperature coefficient of mode index,

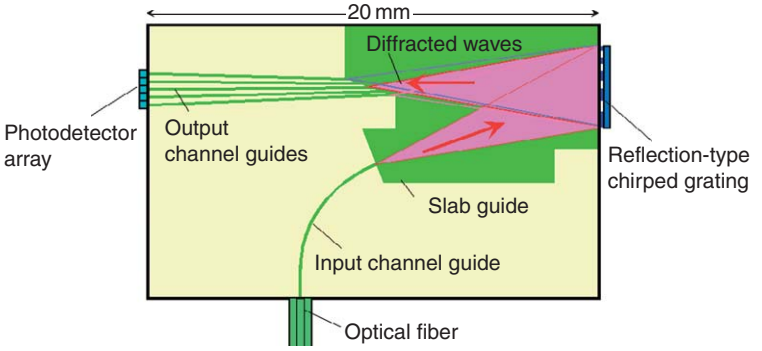


Figure 30 Wavelength demultiplexer using a reflection-type chirped grating and an ion-exchanged glass waveguide. (Source: Suhara T, Viljanen J, Leppihalme M 1982 Integrated optic wavelength multi- and demultiplexers using a chirped grating and an ion-exchanged waveguide. *Appl. Opt.* **21**, 2195–8.)

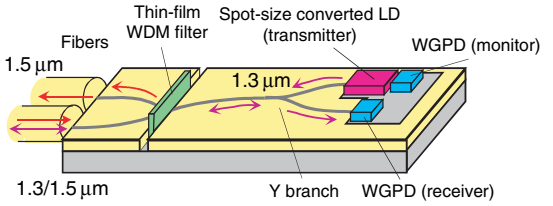


Figure 31 An integrated optic transceiver for wavelength division multiplexing (WDM) access network systems using the 1.3- and 1.5-μm bands for bidirectional signals and broadband signals. WGPD, waveguide photodiode. (Sources: Himeno A, Kato K, Miya T 1998 Silica-based planar lightwave circuits. *IEEE J. Sel. Top. Quantum Electron.* **4**, 913–24; Uchida N, Yamada Y, Hibino Y, Suzuki Y, Ishihara N 1997 Low-cost hybrid WDM module consisting of a spot-size converted laser diode and a waveguide photodiode on a PLC platform for access network systems. *IEICE Trans. Electron.* **E 80-C**, 88–97.)

tailoring the wavelength filter characteristics to obtain a flat passband, and reduction of the insertion loss by the use of silica waveguide and careful design (Himeno *et al.* 1998, Sugita *et al.* 2000). A two-stage AWG multi/demultiplexer of 1000 channels has been developed (Hibino 2002). AWGs are used to implement a TO wavelength-tunable filter and are also used as a building block to construct devices of more complicated functions such as reconfigurable add/drop multiplexers, channel gain equalizers, and dispersion compensators. Applications of AWG devices are not limited to optical communications but extended to optical sensing and signal processing. AWGs were also fabricated with semiconductor waveguides and were integrated with lasers and PDs to implement monolithic one-chip WDM transmitters and receivers (Yoshikuni 2002).

Various IO devices for optical communications have also been implemented by monolithic integration of semiconductor lasers with other waveguide components and devices. A typical example is a DFB laser integrated with an intensity modulator. Although optical communications signal can be produced by direct modulation of a DFB laser, the associated carrier density modulation gives rise to the lasing frequency chirping, and the resultant spectral broadening imposes limitation on the transmission capacity. An external modulator is required to solve this problem. A compact and high-speed transmitter is implemented by monolithic integration of the DFB laser and the modulator. **Figure 32** illustrates a DFB laser integrated with an EA modulator (Suzuki *et al.* 1992). DFB-EA transmitters of 40 Gbits s^{-1} data rate are commercially available.

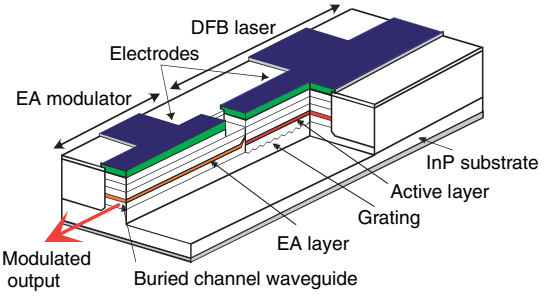


Figure 32 A distributed feedback (DFB) laser monolithically integrated with an electroabsorption (EA) external modulator.

The need of multiwavelength laser light sources with controlled wavelengths for use in WDM photonic network systems motivated the development of wavelength-tunable lasers and wavelength-selectable lasers. Various integrated tunable semiconductor lasers, such as two-section and three-section DBR lasers (Koch and Koren 1990), and advanced widely tunable lasers, such as grating-coupled sampled reflector (GCSR) laser (Oberg *et al.* 1993), superstructure grating DBR (SSDBR) lasers (Tohmori *et al.* 1993), and sampled grating DBR (SGDBR) lasers (Mason *et al.* 2000), which evolved from the multisection DBR lasers, were developed. **Figure 33** shows the cross sections of widely tunable DBR lasers. **Figure 34** illustrates an example of integrated wavelength-selectable laser consisting of eight microarray DFB lasers with slightly different lasing wavelengths, an MMI coupler, and an SOA for compensating the loss in the coupler (Hatakeyama *et al.* 2003). DFB-EA transmitter chips with assorted wavelengths for WDM systems were also developed (Kudo *et al.* 1998).

Other IO devices for optical communications include EO wavelength-tunable filters (Warzanskyj *et al.* 1988), IO heterodyne receivers for coherent transmission (Hernandez-Gil *et al.* 1989, Stallard *et al.* 1986), EO polarization controller (Heismann and Whalen 1992), and AO wavelength-selective switches (Wehrmann *et al.* 1996). **Figure 35** shows the configuration of an AO add/drop multiplexer for dynamic routing of the WDM signals (Nakazawa 2001).

3.05.6.3 Optical Interconnection

The microelectronic processing chips in present large-scale computer systems operate above gigahertz clock speed in parallel. However, severe bottlenecks of data transmission in the electrical interconnection between chips and between boards restrict the high-speed

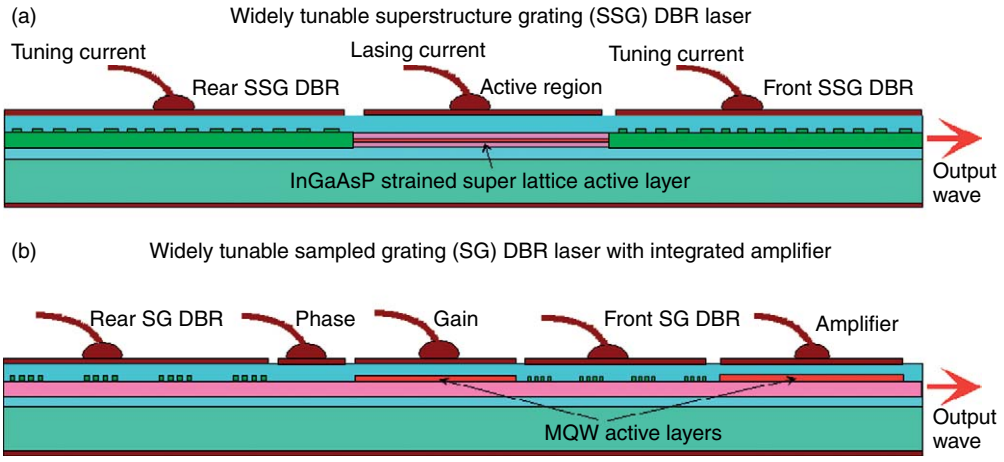


Figure 33 Cross-sectional structures of widely tunable distributed Bragg reflector (DBR) lasers. (a) Superstructure grating DBR laser. (Source: Tohmori Y, Yoshikuni Y, Ishii H, Kano F, Tamamura T, Kondo Y, Yamamoto M 1993 Broad-range wavelength-tunable superstructure grating (SSG) DBR lasers. *IEEE J. Quantum Electron.* **29**, 1817–23.) (b) Sampled grating DBR laser with integrated amplifier. MQW, multiple quantum well. (Source: Mason B, Barton J, Fish G A, Coldren L A, Denbaars S P 2000 Design of sampled grating DBR lasers with integrated semiconductor optical amplifiers. *IEEE Photon. Technol. Lett.* **12**, 762–4.)

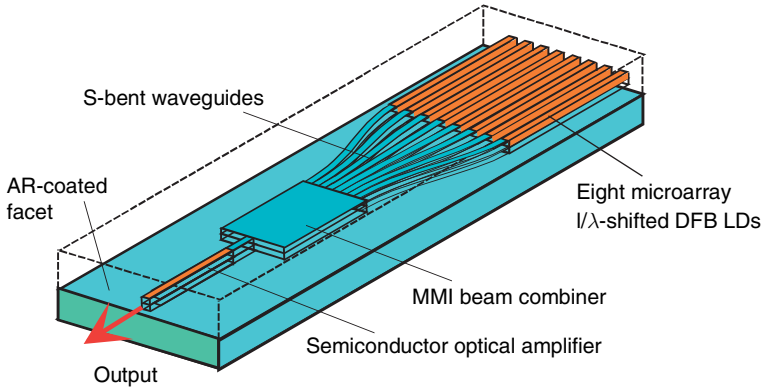


Figure 34 Integrated wavelength-selectable semiconductor laser consisting of eight microarray distributed feedback (DFB) lasers, a multimode interference coupler, and an optical amplifier. AR, antireflection; LDs, laser diodes; MMI, multimode interference. (Source: Hatakeyama H, Naniwae K, Kudo K, Suzuki N, Sudo S, Ae S, Muroya Y, Yashiki K, Satoh K, Morimoto T, Mori K, Sasaki T 2003 Wavelength-selectable microarray light sources for S-, C-, and L-band WDM systems. *IEEE Photon. Technol. Lett.* **15**, 903–5.)

processing of the systems. Optical interconnection, including high-speed parallel data transfer by laser light, as well as optical clock and data signal distribution to a large number of chips free from signal skew, is considered a promising solution to overcome this problem (Goodman *et al.* 1984). For frame-to-frame and board-to-board interconnection, optical printed circuit board (OPCB) consisting of polymer waveguides or optical fiber ribbons was developed and is widely used in combination with optical connectors and optical transmitter/receiver modules (Nordin *et al.* 1995). To achieve optical interconnection between chips in computer systems, appropriate optical components and packaging strategies for these components are required. To meet the requirements, a large number of development works are being continued (Goodman and Hutcheson 1990, Miller 2000). One of the key issues is to implement appropriate coupling elements to link the light from the laser sources (vertical cavity surface-emitting lasers) and the waveguides in the OPCB or from the waveguides to detectors (Mikawa *et al.* 2003, Ohmori *et al.* 2005). The proposed

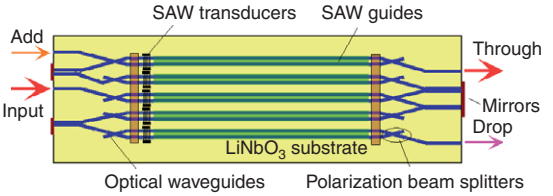


Figure 35 A multistage acoustooptic tunable filter for optical add/drop multiplexing. SAW, surface acoustic wave. (Source: Nakazawa T 2001 Integrated acousto-optic tunable filter for optical add/drop multiplexers. *Proc. Eur. Conf. Integrated Optics*, Paderborn, Germany, pp. 27–30.)

devices include the use of 45° ended waveguides and microlenses, 45° mirror and microlenses, 90° bent fiber connectors, butt-coupling devices, grating couplers, and waveguide holograms (Lin *et al.* 1990).

3.05.6.4 Optical Memories

The first work on the application of waveguide technique to optical memories was proposed and demonstrated using waveguide holograms (Suhara *et al.* 1976). The waveguide holograms stores pictorial information in a photorefractive recording layer on a planar waveguide, and the information is read out by a guided wave. They offer the possibility of hologram integration and have a very large field of view. Integrated waveguide hologram memories using EO switches and liquid-crystal switch matrixes were demonstrated (Suhara *et al.* 1981). Multilayer waveguide hologram memories are also developed (Imai *et al.* 2003).

An important device for application in optical memories is IO disk pickup (IODPU). The prototype first proposed and demonstrated is shown in **Figure 36** (Suhara *et al.* 1985, Ura *et al.* 1986). The device is constructed by an FGC, a twin grating focusing beam splitter (TGFBS), and an array of four PDs all integrated in a planar glass waveguide on an SiO_2/Si substrate. The FGC focuses the light from a butt-coupled LD into a diffraction-limited small spot on the surface of a disk. The light reflected by the disk is collected back into the waveguide by the FGC and is processed and deflected by the TGFBS to the PDs. Readout and focusing/tracking error signals are obtained by electric processing of the four photocurrents. IO differential detection system for MO disk pickup was also developed (Suhara *et al.* 1990). Readout of standard ISO CD disk by an IODPU was demonstrated by using a He–Ne laser (Kawakubo *et al.* 1994). A monolithic IO disk pickup head consisting of a DFB LD, grating components, and PDs in a semiconductor QW waveguide was also proposed and studied (Suhara *et al.* 1996).

3.05.6.5 Optical Sensors

Some of the characteristics of waveguide components are sensitive under ambient conditions and this can

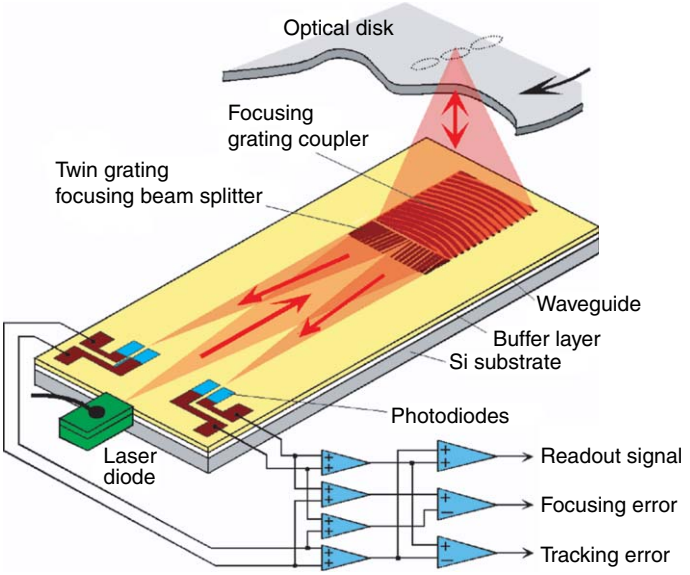


Figure 36 An integrated optic disk pickup device. (Sources: Suhara T, Ura S, Nishihara H, Koyama J 1985 An integrated optic disc pickup device. *Int. Conf. Integrated Optics Optical Fiber Communications*, Venice, Italy; Ura S, Suhara T, Nishihara H, Koyama J 1986 An integrated-optic disk pickup device. *J. Lightwave Technol.* **4**, 913–7.)

be made use of when developing sensor devices (Taylor 1987). The simplest example is the use of a waveguide interferometer as a temperature sensor head. A number of asymmetric waveguide MZIs with different arm lengths are integrated, and a laser light of constant power is fed into the input port. Because the mode index and therefore the phase difference between the waves transmitted through each arm depend on the ambient temperature, the optical power detected at the output port varies sinusoidally with temperature. The use of several sets of asymmetric interferometers with different arm length differences enables temperature measurement free from sensitivity null and ambiguity due to the periodic response of a single interferometer. The device construction can also be applied to implement sensors for other physical quantities, by incorporating some element to convert the change in the quantity to be sensed into change in the mode index. The surface of the interferometer waveguides is covered by a cladding layer with a sensing window on one of the interferometer arms. Then the device can sense the change in the refractive index of liquid in contact with the waveguide within the window area and thus can be used as chemical sensors and biomedical sensors. Waveguide sensor heads based on the similar concept include humidity sensors with polymer and metal film cladding and pressure sensors with interferometer arm formed in a diaphragm structure. Similar sensor devices can be implemented also with other interferometer configurations, such as Michelson interferometer, and also with ring resonators. Waveguide grating components such as input coupler can also be used as a sensor element, as the coupling efficiency for a fixed optical wavelength and the incident angle depend on the mode. A compact IO immunosensor using chirped grating couplers was developed (Dübendorfer and Kunz 1998).

Waveguide EO modulators have application as a sensor head for RF signal. An RF antenna is directly connected to the electrodes of an EO modulator, and the modulated optical wave is transmitted through an optical fiber to a remote receiver.

One of the IO sensing devices is the IODPU presented in Section 3.05.6.4. Another sensor that evolved from the IODPU is an IO head of confocal scanning laser microscope that was proposed and demonstrated (Sheard *et al.* 1993). IO interferometer position/displacement sensor (IOPS) was constructed based on the grating elements (Suhara *et al.* 1989). **Figure 37** illustrates the IOPS demonstrated as the first fully integrated monolithic one-chip optical sensor device (Suhara *et al.* 1995). A monolithic double Michelson interferometer

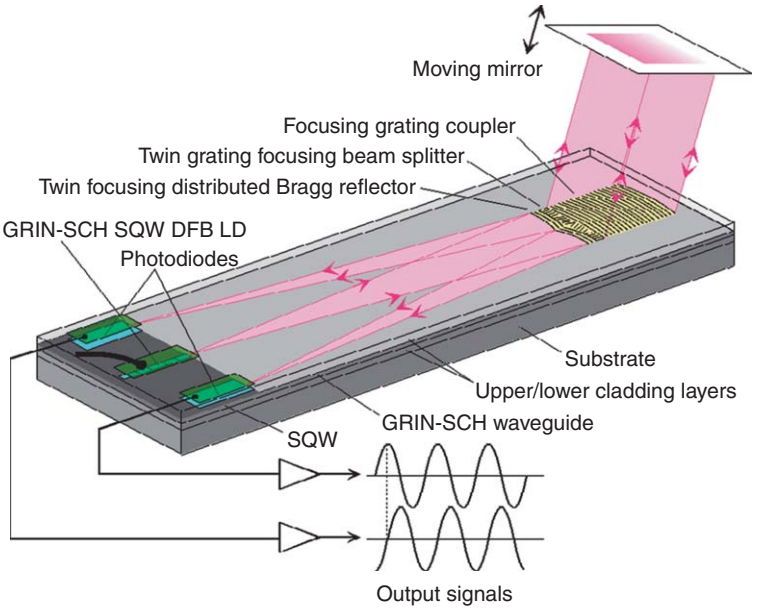


Figure 37 A monolithic IO interferometer position/displacement sensor. DFB, distributed feedback; GRIN, Graded-Index; LD, laser diode; SQW, single quantum well; SCH, separately confinement hetero. (Source: Suhara T, Taniguchi T, Uemukai U, Nishihara H, Hirata T, Iio S, Suehiro M 1995 Monolithic integrated-optic position/displacement sensor using waveguide ratings and QW-DFB laser. *IEEE Photon. Technol. Lett.* **7**, 1195–7.)

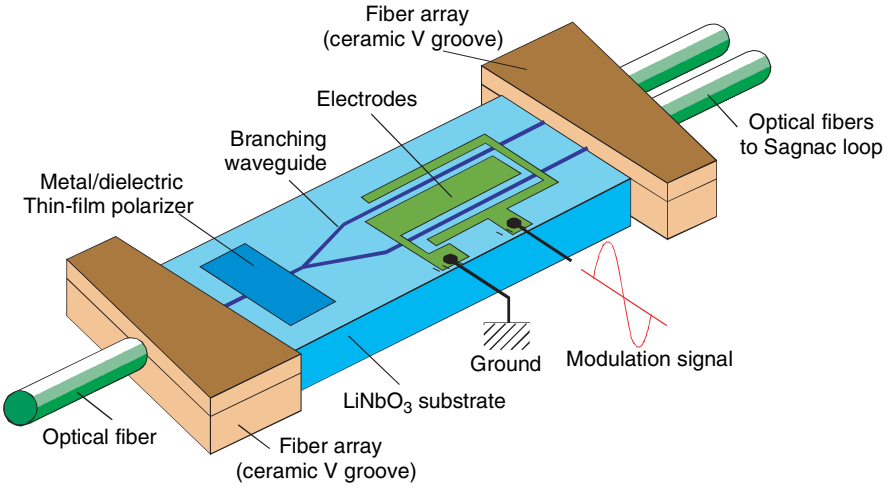


Figure 38 An integrated optic electrooptic push–pull phase modulator, a polarizer and a Y-branching waveguide for optical fiber gyroscope. (Source: Kozuka Y, Toyoda S, Ichigi T, Kondo A, Asano M, Osugi Y, Kondo J, Hamajima A, Ejiri T, Enomoto A 1999 High-reliable $LiNbO_3$ integrated optical device for consumer fiber optic gyroscopes. *Proc. Eur. Conf. Integrated Optics*, pp. 269–74.)

displacement sensor was also demonstrated (Hofstetter *et al.* 1996).

An IO device for fiber laser Doppler velocimeter (LDV) was constructed by integrating an EO frequency shifter, a TE–TM mode converter, and a mode splitter (Toda *et al.* 1987). The device produces a frequency-shifted reference wave and combines it with a Doppler-shifted signal wave transmitted through an optical fiber sensing probe for heterodyne detection.

Advanced fiber sensors include the fiber optic laser gyroscope using the Sagnac interferometer configuration for rotation sensing. To achieve high sensitivity and high stability, several components such as couplers and a phase modulator or a frequency shifter are required. IO devices for constructing such a gyroscope in combination with a fiber coil were developed. As the frequency shifter, an EO phase modulator driven by a sawtooth wave, called serrodyne frequency shifter, and AO Bragg cell integrated with waveguide lenses were developed (Taylor 1987). One of the commercialized device is constructed with an EO push–pull phase modulator, a polarizer, and a Y-branching waveguide in $LiNbO_3$ substrate as shown in **Figure 38** (Kozuka *et al.* 1999). Considerable effort was devoted also to the implementation of IO version of Sagnac interferometer rotation sensor. The first device using a silica waveguide coil fully integrated on a Si substrate demonstrated a typical sensitivity of $1°s^{-1}$ and a full linearity in $\pm 900° s^{-1}$ (Mottier and Pouteau 1997). IO gyro device using silica waveguide and TO phase shifter was also demonstrated (Suzuki *et al.* 2000).

3.05.7 Conclusions

An outline of the IO technology was presented – the concept, theories, materials, design and fabrication techniques, and applications. The IO technology has now reached a high degree of maturity and is one of the most important elements supporting information technology. Many IO devices for optical communications and optical memories including those with advanced functions are or will be practically and widely used. It is clearly expected that further development will be made to improve the performances and production efficiency and to extend the applications. Future research and development works are also required to extend the applications in the areas such as optical interconnection for computer systems, integrated waveguide nonlinear optics, biomedical sensing, and quantum information processing.

References

Alferness R C 1982 Waveguide electrooptic modulators. *IEEE Trans. Microwave Theory Technol.* **MTT-30**, 1121–37

Carroll J, Whiteaway J, Plumb D 1998 *Distributed Feedback Semiconductor Lasers*. IEE, London and SPIE, Washington

Coldren L A, Corzine S W 1995 *Diode Lasers and Photonic Integrated Circuits*. John Wiley & Sons, New York

Dübendorfer J, Kunz R E 1998 Compact integrated optical immunosensor using replicated chirped grating coupler sensor chips. *Appl. Opt.* **37**, 1890–4

Gabriel K J 1998 Microelectromechanical systems. Special issue on "Integrated sensors, microactuators, and microsystems (MEMS)." *Proc. IEEE* **86**, 1531–787

Giles C R, Aksyuk V, Barber B, Ruel R, Sultz L, Bishop D 1999 A silicon MEMS optical switch attenuator and its use in lightwave subsystems. *IEEE J. Sel. Top. Quantum Electron.* **5**, 18–25

Girton D G, Kwiatowski S L, Lipscomb G F, Lytel R S 1991 20 GHz electro-optic polymer Mach–Zehnder modulator. *Appl. Phys. Lett.* **58**, 1730–2

Goh T, Yasu M, Hattori K, Himeno A, Okuno M, Ohmori Y 2001 Low loss and high extinction ratio strictly nonblocking 16×16 thermooptic matrix switch on 6-in wafer using silica-based planar lightwave circuit technology. *J. Lightwave Technol.* **19**, 371–9

Goodman J W, Hutcheson L D 1990 Optical interconnects: Introduction by the feature editors. *Appl. Opt.* **29**, 1067

Goodman J W, Leonberger F I, Kung S Y, Athale R A 1984 Optical interconnections for VLSI systems. *Proc. IEEE* **72**, 850–66

Hatakeyama H, Naniwae K, Kudo K, Suzuki N, Sudo S, Ae S, Muroya Y, Yashiki K, Satoh K, Morimoto T, Mori K, Sasaki T 2003 Wavelength-selectable microarray light sources for S-, C-, and L-band WDM systems. *IEEE Photon. Technol. Lett.* **15**, 903–5

Heismann F, Whalen M S 1992 Fast automatic polarization control system. *IEEE Photon. Technol. Lett.* **4**, 503–5

Hernandez-Gil F, Koch T L, Koren U, Gnall R A, Bums C A 1989 A tunable MQW-DBR laser with a monolithically integrated InGaAsP/InP directional coupler switch. *Electron. Lett.* **25**, 1271–2

Hibino Y 2002 Recent advances in high-density and large-scale AWG multi/demultiplexers with higher index-contrast silica-based PLCs. *IEEE J. Sel. Top. Quantum Electron.* **8**, 1090–101

Himeno A, Kato K, Miya T 1998 Silica-based planar lightwave circuits. *IEEE J. Sel. Top. Quantum Electron.* **4**, 913–24

Hirata T, Iio S, Suehiro M, Hihara M, Suhara T, Uemukai M, Taniguchi T, Nishihara H 1995 Integration of laser diodes and photodiodes in photonic integrated circuit for measurement. *Jpn. J. Appl. Phys.* **34**, L1288–90

Hofstetter D, Zappe H, Dändliker R 1996 A monolithically integrated double Michelson interferometer for optical displacement measurement with direction determination. *IEEE Photon. Technol. Lett.* **8**, 1370–2

Imai T, Kurihara T, Yagi S, Kurokawa Y, Endo M, Tanabe T 2003 Orthogonal aperture multiplexing for multilayered waveguide holographic read-only memories. *Appl. Opt.* **42**, 7085–92

Izutsu M, Yamane Y, Sueta T 1977 Broad-band traveling-wave modulator using $LiNbO_3$ optical waveguide. *IEEE J. Quantum Electron.* **QE-13**, 287–90

Jalali B 1997 Silicon-on-insulator photonic integrated circuit (SOI-PIC) technology. *Proc. SPIE* **2997**, 60–71

Kato K 1999 Ultrawide-band/high-frequency photodetectors. *IEEE Trans. Microwave Theory Technol.* **47**, 1265–81

Kawakubo I, Funazaki J, Shirane K, Yoshizawa A 1994 Integrated optical-disk pickup that uses a focusing grating coupler with a high numerical aperture. *Appl. Opt.* **33**, 6855–9

Koch T L, Koren U 1990 Semiconductor lasers for coherent optical fiber communications. *J. Lightwave Technol.* **LT-8**, 274–93

Kogelnik H, Schmidt R V 1976 Switched directional couplers with alternating $\Delta\beta$. *IEEE J. Quantum Electron.* **QE-12**, 396–401

Koshiba M, Saitoh H, Eguchi M, Hirayama K 1992 Simple scalar finite element approach to optical waveguides. *IEE Proc. J.* **139**, 166–71

Kozuka Y, Toyoda S, Ichigi T, Kondo A, Asano M, Osugi Y, Kondo J, Hamajima A, Ejiri T and Enomoto A 1999 High-reliable $LiNbO_3$ integrated optical device for consumer fiber optic gyroscopes. *Proc. Eur. Conf. Integrated Optics*, Torino, Italy, pp. 269–74.

Kudo K, Sasaki T, Yamaguchi M 1997 Migration-controlled narrow-stripe selective MOVPE for high-quality InGaAsP/InGaAsP MQWs. *J. Crystal Growth* **170**, 634–8

Kudo K, Ishizaka M, Sasaki T, Yamazaki H, Yamaguchi M 1998 1.52–1.59-μm range different-wavelength modulator-integrated DFB-LDs fabricated on a single wafer. *IEEE Photon. Technol. Lett.* **10**, 929–31

Li E H 2000 *Semiconductor Quantum Wells Intermixing*. Gordon and Breach Science Publishers, Singapore

Lifante G 2003 *Integrated Photonics*. John Wiley & Sons, Chichester

Lin L Y, Goldstein E L 2002 Opportunities and challenges for MEMS in lightwave communications. *IEEE J. Sel. Top. Quantum Electron.* **8**, 163–72

Lin F, Strzekecju E M, Jannson T 1990 Optical multiplanar VLSI interconnects based on multiplexed waveguide holograms. *Appl. Opt.* **29**, 1126–33

Marcatili E A J 1969 Dielectric rectangular waveguide and directional coupler for integrated optics. *Bell Syst. Tech. J.* **48**, 2071–102

Marcuse D 1974 *Theory of Dielectric Optical Waveguides*. Academic Press, New York

Mason B, Barton J, Fish G A, Coldren L A, Denbaars S P 2000 Design of sampled grating DBR lasers with integrated semiconductor optical amplifiers. *IEEE Photon. Technol. Lett.* **12**, 762–4

Mergerian D, Malarkey E C, Pautienus R P, Bradley J C, Marx G E, Hutcheson L D, Kellner A L 1980 Operational integrated optical RF spectrum analyzer. *Appl. Opt.* **19**, 3033–4

Mikawa T, Kinoshita M, Hiruma K, Ishitsuka T, Okabe M, Hiramatsu S, Furuyama H, Matsui T, Kumai K, Ibaragi O, Bonkohara M 2003 Implementation of active interposer for high-speed and low-cost chip level optical interconnects. *IEEE J. Sel. Top. Quantum Electron.* **9**, 452–9

Miller S E 1969 Integrated optics: An introduction. *Bell Syst. Technol. J.* **48**, 2059–68

Miller D A B 2000 Rationale and challenges for optical interconnects to electronic chips. *Proc. IEEE* **88**, 728–49

Motamedi M E 1994 Micro-opto-electro-mechanical systems. *Opt. Eng.* **33**, 3505–17

Mottier P, Pouteau P 1997 Solid state optical gyrometer integrated on silicon. *Electron. Lett.* **33**, 1975–7

Murphy E J (ed). 1999 *Integrated Optical Circuits and Components*, Marcel Dekker, New York

Nakazawa T 2001 Integrated acousto-optic tunable filter for optical add/drop multiplexers. *Proc. Eur. Conf. Integrated Optics*, Paderborn, Germany, pp. 27–30

Nishihara H, Haruna M, Suhara T 1989 *Optical Integrated Circuits*. McGraw-Hill, New York

Nordin R A, Holland W R, Shahid M A 1995 Advanced optical interconnection technology in switching equipment. *J. Lightwave Technol.* **13**, 987–94

Oberg M, Nilsson S, Streubel K, Wallin J, Backbom L, Klinga T 1993 74 nm wavelength tuning range of an InGaAsP/InP vertical grating assisted codirectional coupler laser with rear sampled grating reflector. *IEEE Photon. Tech. Lett.* **5**, 735–7

Ohmori J, Imaoka Y, Ura S, Kintaka K, Satoh R, Nishihara H 2005 Integrated-optic add/drop multiplexing of free-space waves for chip-to-chip optical interconnecting board. *Jpn. J. Appl. Phys.* **44**, 7987–92

Ooba N, Toyoda S, Kurihara T 2000 Low crosstalk and low loss polymeric digital optical switch. *Jpn. J. Appl. Phys.* **39**, 2369–71

Sheard S, Suhara T, Nishihara H 1993 Integrated-optic implementation of a confocal scanning optical microscope. *J. Lightwave Technol.* **11**, 1400–3

Shubert R, Harris J H 1968 Optical surface waves on thin films and their application to integrated data processors. *IEEE Trans. Microwave Theory Technol.* **MTT-16**, 1048–54

Smit M K 1988 New focusing and dispersive planar component based on an optical phased array. *Electron Lett.* **24**, 385–6
Snyder A W, Love J D 1983 *Optical Waveguide Theory*. Chapman & Hall, London
Stallard W A, Bearumont A R, Booth R C, 1986 Integrated optics devices for coherent transmission. *J. Lightwave Technol.* **LT-4**, 852–7
Stern M 1988 Semivectorial polarized finite difference method for optical waveguides with arbitrary index profiles. *IEE Proc. J.* **135**, 333–8
Sugita A, Kaneko A, Okamoto K, Itoh M, Himeno A, Ohmori Y 2000 Very low insertion loss arrayed-waveguide grating with vertically tapered waveguides. *IEEE Photon. Technol. Lett.* **12**, 1180–2
Suhara T 2004 *Semiconductor Laser Fundamentals*. Marcel Dekker, New York
Suhara T, Fujimura M 2003 *Waveguide Nonlinear-optic Devices*. Springer-Verlag, Berlin
Suhara T, Nishihara H 1986 Integrated optics components and devices using periodic structures. *IEEE J. Quantum Electron.* **QE-22**, 845–67
Suhara T, Nishihara H, Koyama J 1976 Waveguide holograms: A new approach to hologram integration. *Opt. Commun.* **19**, 353–8
Suhara T, Nishihara H, Koyama J 1981 Integrated hologram memory device using optical guide and liquid-crystal switch matrix. *Tech. Digest. Int. Conf. Integrated Optics Optical Fiber Communications*, San Francisco, CA, USA, WK-5
Suhara T, Viljanen J, Leppihalme M 1982 Integrated optic wavelength multi- and demultiplexers using a chirped grating and an ion-exchanged waveguide. *Appl. Opt.* **21**, 2195–8
Suhara T, Ura S, Nishihara H, Koyama J 1985 An integrated optic disc pickup device. *Int. Conf. Integrated Optics Optical Fiber Communications*, Venice, Italy
Suhara T, Fujiwara T, Nishihara H 1986 Proton-exchanged Fresnel waveguide lenses in Ti:$LiNbO_3$ waveguides. *Appl. Opt.* **25**, 3379–83
Suhara T, Ura S, Nishihara H 1989 Integrated-optic interferometer position sensor. *J. Lightwave Technol.* **7**, 270–3
Suhara T, Ishimaru H, Ura S, Nishihara H 1990 Integration of detection optics for magnetooptic disk pickup. *Trans. IEICE* **E73**, 110–5
Suhara T, Taniguchi T, Uemukai U, Nishihara H, Hirata T, Iio S, Suehiro M 1995 Monolithic integrated-optic position/displacement sensor using waveguide ratings and QW-DFB laser. *IEEE Photon. Technol. Lett.* **7**, 1195–7
Suhara T, Taniguchi T, Uemukai M, Yoshimoto A, Nishihara H, Hirata T, Iio S, Suehiro M 1996 Monolithic integration of waveguide gratings and QW-DFB laser for implementation of subminiature disc pickup head. *Jpn. J. Appl. Phys.* **35**, 369–71
Suzuki M, Tanaka H, Taga H, Yamamoto S, Matsushima Y 1992 $\lambda/4$-shifted DFB laser/electroabsorption modulator integrated light source for multigigabit transmission. *J. Lightwave Technol.* **10**, 90–5
Suzuki K, Takiguchi K, Hotate K 2000 Monolithically integrated resonator microoptic gyro on silica planar lightwave circuit. *J. Lightwave Technol.* **18**, 66–72
Takahashi H, Suzuki H, Katoh K, Nishi I 1990 Arrayed-waveguide grating for wavelength division multi/demultiplexer with nanometer resolution. *Electron Lett.* **26**, 87–8
Takeuchi T, Sasaki T, Hayashi M, Hamamoto K, Makita K, Taguchi K, Komatsu K 1996 A transceiver PIC for bidirectional optical communication fabricated by bandgap energy controlled selective MOVPE. *IEEE Photon. Technol. Lett.* **8**, 361–3
Tamir T (ed.) (1975) *Integrated Optics* , Springer-Verlag, Berlin
Tamir T (ed.) (1988) *Guided-wave Optoelectronics* , Springer-Verlag, Berlin
Taylor H F 1987 Application of guided-wave optics in signal processing and sensing. *Proc. IEEE* **75**, 1524–35
Thylén L 1988 Integrated optics in $LiNbO_3$: Recent developments in devices for telecommunications. *J. Lightwave Technol.* **6**, 847–61
Toda H, Haruna M, Nishihara H 1987 Optical integrated circuit for a fiber laser Doppler velocimeter. *J. Lightwave Technol.* **5**, 901–5
Tohmori Y, Yoshikuni Y, Ishii H, Kano F, Tamamura T, Kondo Y, Yamamoto M 1993 Broad-range wavelength-tunable superstructure grating (SSG) DBR lasers. *IEEE J. Quantum Electron.* **29**, 1817–23
Tsai C S 1979 Guided-wave acoustooptic Bragg modulators for wideband integrated optic communications and signal processing. *IEEE Trans. Circuits Syst.* **CAS-26**, 1072–89
Uchida N, Yamada Y, Hibino Y, Suzuki Y, Ishihara N 1997 Low-cost hybrid WDM module consisting of a spot-size converted laser diode and a waveguide photodiode on a PLC platform for access network systems. *IEICE Trans. Electron.* **E80-C**, 88–97
Ura S, Suhara T, Nishihara H, Koyama J 1986 An integrated-optic disk pickup device. *J. Lightwave Technol.* **4**, 913–7
Warzanskyj W, Heismann F, Alferness R C 1988 Polarization-independent electro-optically tunable narrow-band wavelength filter. *Appl. Phys. Lett.* **53**, 13–15
Wehrmann F, Harizi C, Herrmann H, Rust U, Sohler W, Westenhofer 1996 Integrated optical, wavelength selective, acoustically tunable 2×2 switches (add-drop multiplexers) in $LiNbO_3$. *IEEE J. Sel. Top. Quantum Electron.* **2**, 263–9
Yariv A 1973 Coupled-mode theory for guided-wave optics. *IEEE J. Quantum Electron.* **QE-9**, 919–33
Yoshikuni Y 2002 Semiconductor arrayed waveguide gratings for photonic integrated devices. *IEEE J. Sel. Top. Quantum Electron.* **8**, 1102–14

Biography

Toshiaki Suhara was born in Wakayama, Japan, in 1950. He received the BS, MS, and Ph.D. degrees from Osaka University in 1973, 1975, and 1978, respectively. He joined the Department of Electronics, Faculty of Engineering, Osaka University, in 1978, and has been a professor since 2002. He was at the Technical Research Centre of Finland as a guest research scientist in 1980, and in Glasgow University, UK, as a guest research fellow in 1986–1987, and in Chalmers University of Technology, Sweden, as a guest professor in 1997. He is the author of *Optical Integrated Circuits* (Japanese, English, and Chinese versions), *Semiconductor Laser Fundamentals* (Japanese, Chinese, and English versions), *Waveguide Nonlinear-optic Devices*, *Quantum Electronics*, and *Optical-wave Engineering*. His research interests include integrated optics, integrated semiconductor lasers, nonlinear optics, quantum optics, and optical memories. Professor Suhara is a senior member of IEEE, and a member of OSA, SPIE, IEICE of Japan, and the Japan Society of Applied Physics. He received the Paper Awards in 1977, 1986, 1994, and 2001, and the Young Engineer Award in 1980 from IEICE, and the Paper Award in 1990 from the Laser Society of Japan. In 1995, he was awarded the IBM Japan Science Prize.

3.06 Artificial Retinas

Wilfried Mokwa, Institute of Materials in Electrical Engineering 1, RWTH Aachen University, Aachen, Germany

Glossary

AIROF Anodic/Activated Iridium Oxide
CDC Charge Delivery Capacity
CMOS Complementary Metal Oxide Semiconductor
DSP Digital Signal Processor
FDA Food and Drug Administration
FES Functional Electrostimulation
MEMS Microelectromechanical Systems
RF-BPN Receptive Field–Adaptive Biology–Inspired Pulse Processing Neural Network
RF Radio Frequency
RP Retinitis Pigmentosa
SIROF Sputtered Iridium Oxide Film
USP United States Pharmacopeia

3.06.1 Introduction

The most frequent cause of blindness is the degeneration of retinal cells. In developed countries macula degeneration is the main cause of blindness and strong visual impairment. About 25–30 million people are suffering from this disease. During the course of this disease the tissue of the macula is first destroyed. This leads to a complete loss of vision in the center. Reading is no longer possible for these patients. Another group of people suffer from retinitis pigmentosa (RP), a genetic disorder. The worldwide number amounts to 1–3 million affected persons. This disease often starts with night blindness.

Contrast and color vision as well as visual acuity diminish. Peripheral vision diminishes first. The field of vision becomes narrower and is completely lost in later stages. These symptoms are because of the degeneration of the light-sensitive rods and cones of the retina. This process can continue for decades depending on the individual concerned. **Figure 1** shows a photograph of a normal and a degenerated retina.

Bypassing the nonworking retina or the optic nerve by a technical implant might restore some vision to these patients. Brindley and Lewin (1968) were the first who tried to develop a visual prosthesis that stimulates the visual cortex directly for blind people.

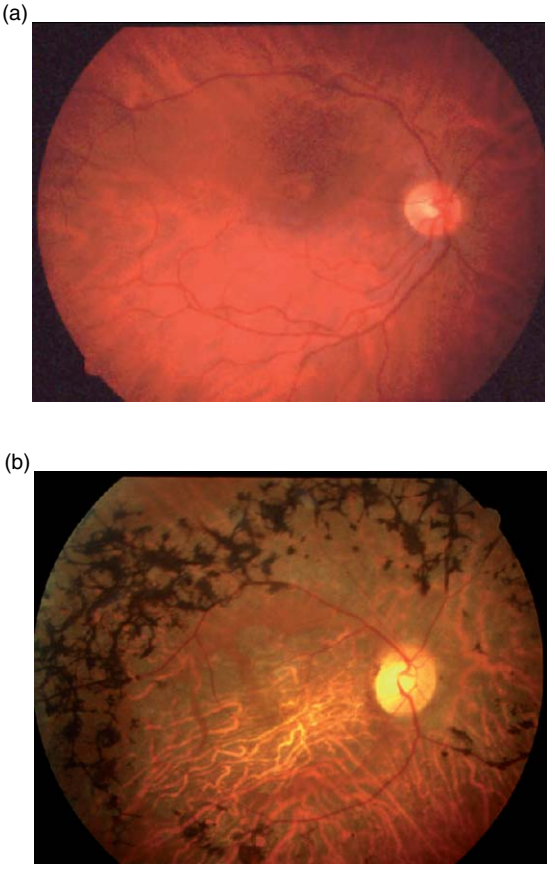


Figure 1 (a) Photograph of a normal human retina. The yellow spot in the right part is the optic disk where the optic nerve leaves the eye. It has a diameter of about 3 mm. (b) Photograph of a degenerated retina of a retinitis pigmentosa (RP) patient. The degenerated regions at the outer border look black. The field of vision is drastically reduced to the central part of the retina. (Courtesy P Walter, Department of Ophthalmology, University Hospital, RWTH Aachen University, Aachen, Germany.)

3.06.2 Physiology of the Human Eye

A cross section of the human eye is shown in **Figure 2** . The outer tissue consists of three different layers: the sclerotic coat, the choroidic coat, and the retina. The retina is the inner layer of the eye and contains the light receptors: the rods and the cones (**Figure 3**). The number of rods is about 60–125 million. Rods are extremely sensitive to light but do not provide a sharp image. They allow us to see in the dark and are responsible for peripheral vision. The number of cones is about 3–6.5 million. They need bright light to be functional. Cones are responsible for seeing colors and provide us with sharp images. An area with a diameter of about 5 mm around the central axis of the retina is called the macula, the region for sharp vision. In the center of the macula, there is a small pit with a diameter of about 1.5 mm, the fovea, the center responsible for sharpest vision. Here the cones are present at a very high density (10%). It is important to emphasize that the rods and cones are beneath a layer of neural cells. Light enters the eye through the cornea, the pupil, and the lens; passes the vitreous body and the neural cells; and is absorbed by the cones and rods. Here the light is converted into nerve pulses that are transferred via different neural cells (bipolar and ganglion cells) to the brain through the optic nerve. Within the visual cortex, a region of the brain that is responsible for visual perception, the incoming pulses are interpreted and converted to a picture. In the eyes of vertebrates, signal processing starts in the retina. There is a complex array of interneurons (amacrine and horizontal cells) that form synapses with the bipolar and ganglion cells by forming a neural network that modifies the activity (**Figure 3**). By this a preprocessing of the signals is achieved, leading, for example, to an increase in contrast. More detailed information on the physiology of the eye can be found in Kaufman and Alm (2003).

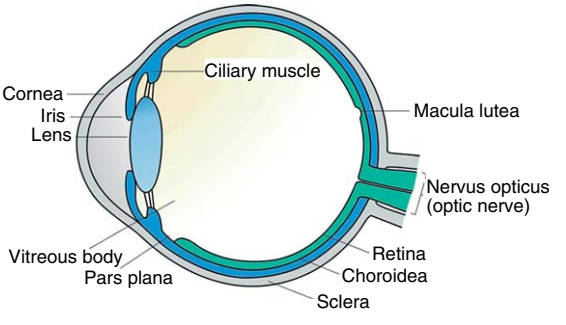


Figure 2 Schematic cross section of the human eye. (Courtesy IIP Technologies, Bonn, Germany.)

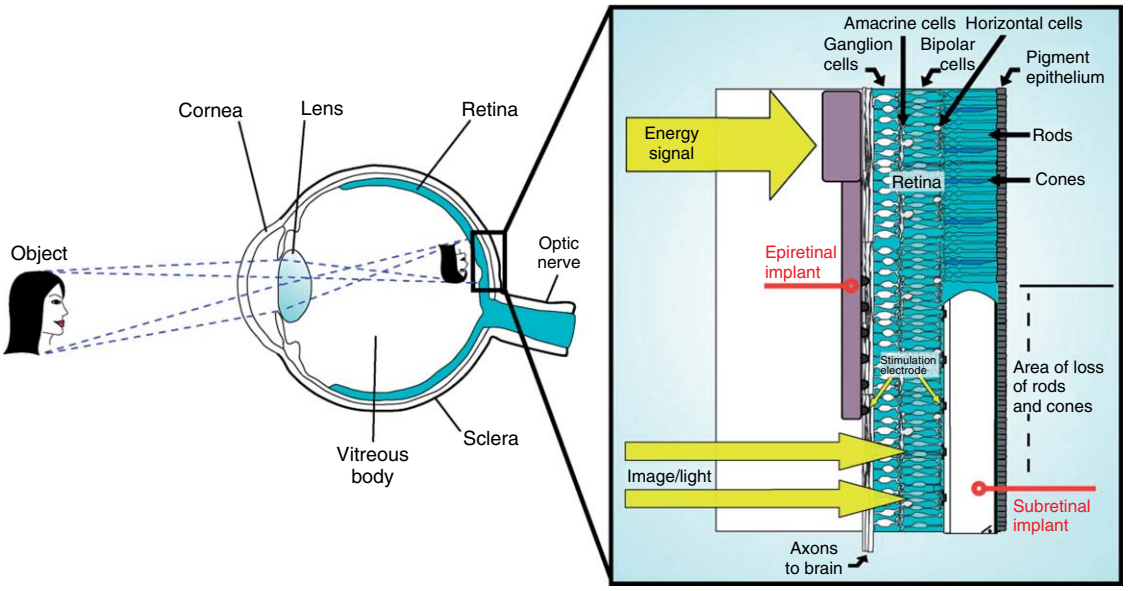


Figure 3 Cross section of the human eye illustrating the epiretinal and the subretinal approach. (Source: Reprinted with permission from Zrenner E 2002 Will retinal implants restore vision? *Science* **295**, 1022–5. Copyright 2002 AAAS.)

3.06.3 Possible Approaches to Restore Vision by Electrical Stimulation of Nerve Cells

3.06.3.1 Retinal Implants

3.06.3.1.1 System concept

Studies have found that about 30% of retinal ganglion cells of RP patients are still functional even after several years of blindness. It has been shown that electrical stimulation of ganglion cells inside the eye at the inner surface of the retina results in visual sensations (Benjamin *et al.* 1994, Santos *et al.* 1997, Stone *et al.* 1992). Therefore, bypassing the degenerated photoreceptors by electrically stimulating the retinal ganglion cells seems possible (Humayun *et al.* 1996, 1999, Walter Heimann 2000, Wyatt *et al.* 1994). By this, action potentials are evoked in retinal ganglion cells, causing a visual sensation.

Most approaches presented up to now for a retinal implant use, besides the implant, an extraocular part that records visual images. The images are transformed by a digital signal processor (DSP) into corresponding signals for stimulation of the nerve cells. These signals are transmitted to the receiver unit of an implant. Integrated circuitry of this unit decodes the signals and transfers the data to a stimulation circuitry that selects stimulation electrodes of the implant and generates current pulses to the electrodes (Schwarz *et al.* 2000). By this, action potentials in the nerve cells of the retina, in the optic nerve, or in the nerve cells of the visual cortex are evoked, causing a visual sensation.

3.06.3.1.2 Subretinal stimulation

Several groups are working on the concept of subretinal stimulation (e.g., Chow 1993, Peachey and Chow 1999, Sachs and Gabel 2004, Schwahn *et al.* 2001, Zrenner 2002, Zrenner *et al.* 1999).

In the subretinal approach the degenerated photoreceptors are replaced by a chip that has a microphotodiode array together with a stimulation electrode array implanted in it (**Figures 3** and **4**). By using the natural optics of the eye an image is mapped onto the photodiode array. The photodiodes respond to the incoming light. Depending on the output of the photodiodes, on-chip electronics generate the necessary stimulation currents that activate healthy bipolar and ganglion cell layers of the retina. Positioning and fixation of the implant are relatively easy. No external camera or image processing is needed. Eye movement can be used to localize objects. A basic requirement for this approach to work is that the optics of the eye should still be functional. Because of the necessary on-chip electronics, energy is required for operation and it has to be supplied externally by means of electromagnetic or optic coupling.

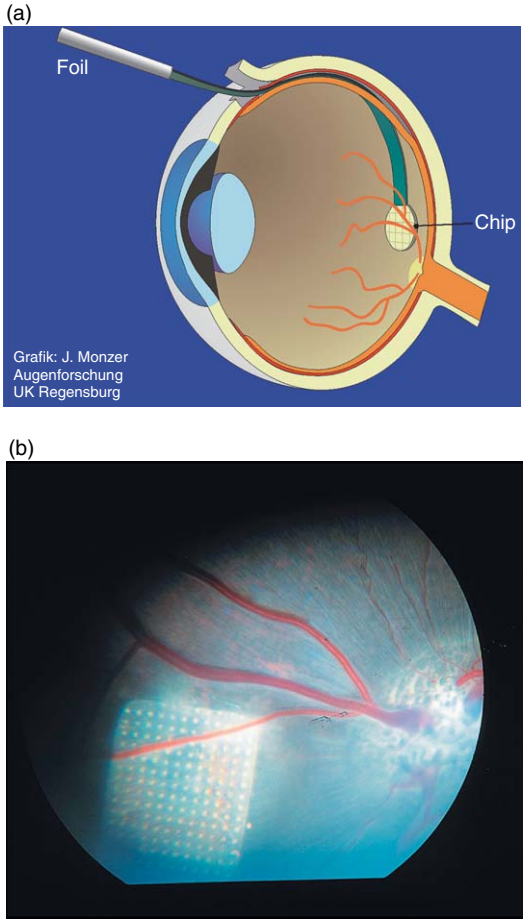


Figure 4 (a) Schematic view of the subretinal approach. (b) View of a subretinal implant through the ganglion cell layer. The implant has an area of 3 mm × 3 mm and a thickness of 50 μm. One thousand five hundred pixels are on this implant with a pixel size of 70 μm × 70 μm. It can be clearly seen from the blood vessel over the implant that the implant is located subretinally. (Courtesy Retinal Implant AG, Reutlingen, Germany.)

3.06.3.1.3 Epiretinal stimulation

In contrast to subretinal stimulation, in epiretinal implants the stimulation electrodes are placed on the inner side of the retina (compare **Figure 3**). Therefore, retinal ganglion cells are stimulated where some preprocessing of the signals has already taken place. This preprocessing has to be considered when transforming the visual information recorded by a camera into data stimulation patterns. Eckmiller (1995, 1996, 1997) was the first to introduce the concept of a retina encoder simulating the neural network of the retina.

The concept of an epiretinal system can be illustrated through **Figure 5** (Schwarz *et al.* 2000). The system consists of an extraocular and an intraocular part. The components of the extraocular part are a complementary metal oxide semiconductor (CMOS) image sensor for capturing visual images, an artificial neural net that imitates the functions of different ganglion layers of the retina (retina encoder), and a transmitter. The visual images are transformed by the neural net into control signals for the stimulation electrodes (Mokwa 2004). These signals are finally transmitted into the interior of the eye together with the energy needed to supply the intraocular part. For the energy transfer a radio frequency (RF) link is used (Mokwa 2004) and for data transfer an RF link (Mokwa 2004) or an optic link is used (Laube *et al.* 2004a). The intraocular part of the system consists of a receiver circuitry where energy and data signals are separated, the clock signal is extracted, and the data signals are decoded. These data are transferred via a microcable to the stimulation unit. An integrated circuitry decodes the information from the serial data stream. From this information the requested electrode is selected and current pulses are generated. **Figure 6** shows an encapsulated implant that was developed by the German EPIRET team and was successfully tested in animals (Laube *et al.* 2004a, b, Mokwa 2004, Walter *et al.* 2005).

Second Sight and partners introduced a similar system based on a cochlea implant (Humayun *et al.* 2003). The whole electronics are mounted on the outer side of the sclera. From there a cable is laid through the sclera into the eye. The cable connects a 4 × 4 stimulation electrode array that is fit into the retina (**Figure 7**).

3.06.3.2 Stimulation of the Optic Nerve

In the optic nerve about 1 million nerve fibers connect the retina with the visual cortex. It is possible to reach the optic nerve through surgery. The idea of realizing a visual prosthesis by stimulating the optic nerve was conceived by Thomas Mortimer and Claude Veraart in the early 1990s (Veraart *et al.* 1998). **Figure 8** shows a schematic view of the system.

For stimulation of the fibers a cuff electrode is wound around the optic nerve. This cuff is connected to implanted stimulation electronics. Information is exchanged with the implanted system via telemetry. The team of Claude Veraart developed a spiral cuff that was made from rubber and flexible silicone (**Figure 9**). The first version was equipped with four platinum contacts and a later version with eight. The electrode diameter is about 3 mm;

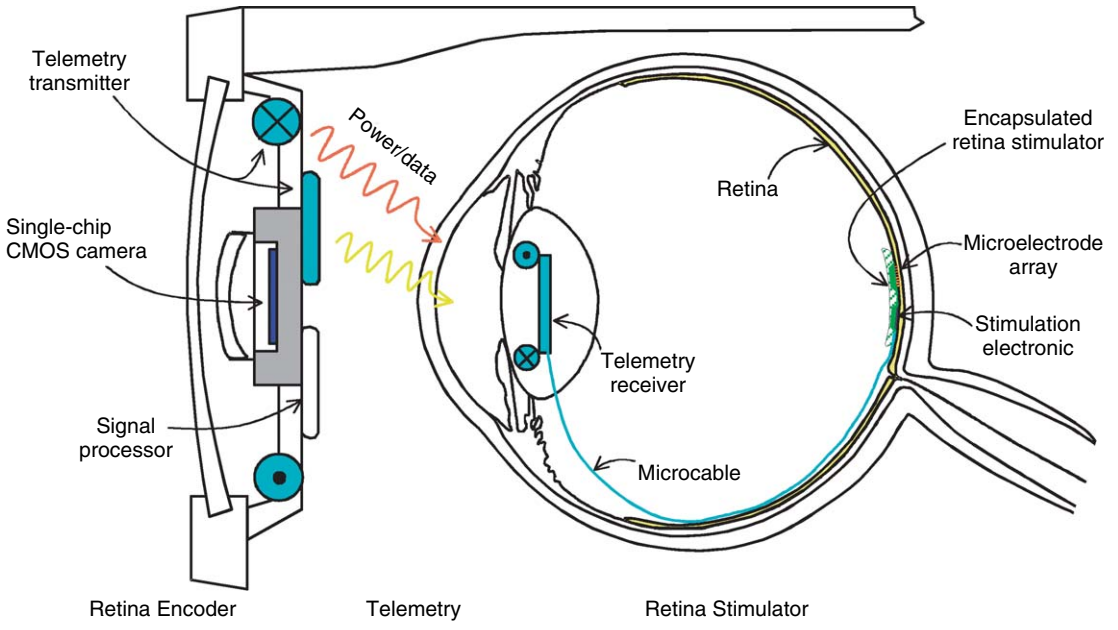


Figure 5 System concept of a visual prosthesis for epiretinal stimulation.

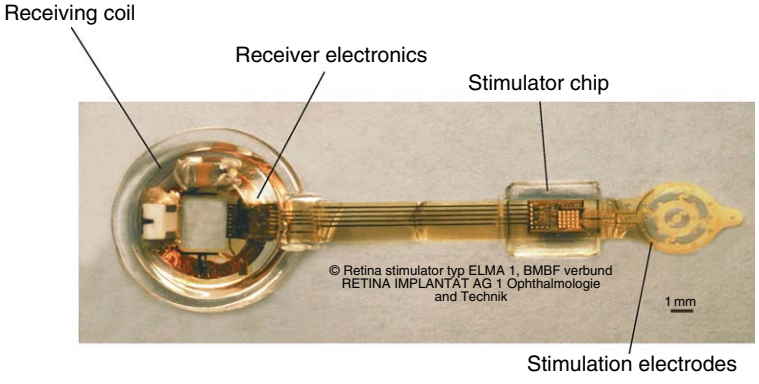


Figure 6 Photograph of an encapsulated retinal implant. The left part is formed like an artificial lens. This lens contains the receiver coil and the chip. A microcable connects the receiver electronics with the stimulator chip. At the right, the stimulating electrodes are visible. The whole implant is encapsulated with silicone.

stimulator dimensions are 25 mm× 30 mm with a 6-mm thickness. To avoid any lesion of the optic nerve the geometry of the cuff allows to adapt continuously to the diameter of the optic nerve while at the same time guaranteeing close contact between the electrode and the nerve. The first implantation of the system was performed in February 1998 by Prof. C. Raftopoulos of the Neurosurgery Unit, St-Luc's University Hospital, UCL, Brussels, on a long-term RP patient (Veraart *et al.* 1998). A spiral cuff was implanted intracranially around the right optic nerve. The cable of the cuff was connected to the implanted connector and was brought through the skin at the clavicle level where it ended in an external connector. By this, optic nerve stimulation could be achieved by using external stimulators. After a brief period of convalescence, the optic nerve fibers were electrically stimulated. The patient reported on sensations (phosphenes) generated by the stimulation of the optic nerve that were exclusively visual. These phosphenes can be obtained at low thresholds. The characteristics of the phosphenes depend on the conditions of electrical stimulation like pulse shape and frequency (Delbecke *et al.* 2003).

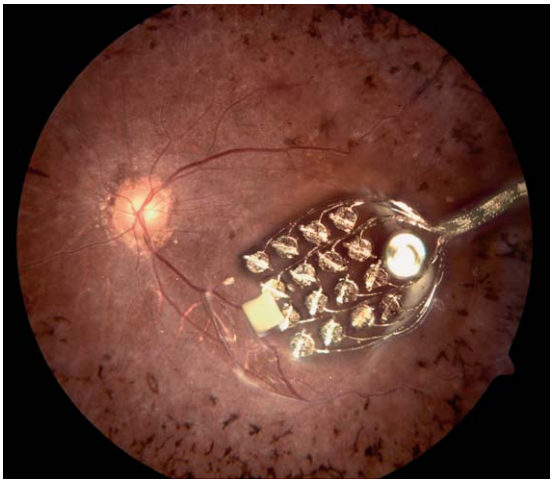


Figure 7 Photograph of a passive epiretinal stimulation array placed on the retina. The array size is about 6 mm × 6 mm and contains 16 stimulation electrodes on a flexible foil. The array is fixed on the retina by a retinal tag placed just to the right of the stimulation electrodes. The stimulation electrodes are connected to electronics outside the eye by a cable that penetrates the sclera (Humayun *et al.* 2003). (Photograph by Jessica Dougall, Copyright 2005, Doheny Eye Institute and Second Sight Medical Products, Inc. All rights reserved, USA.)

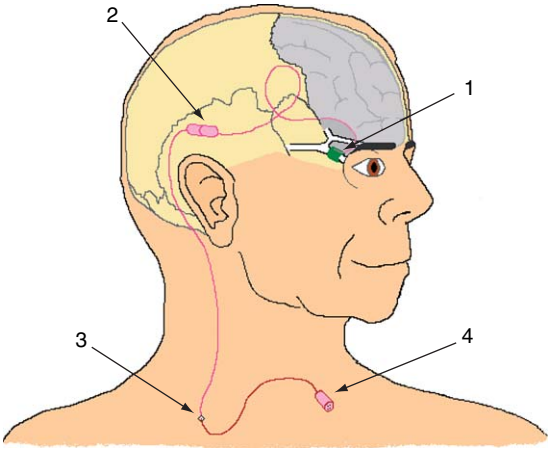


Figure 8 Schematic view of the optic nerve implant system. 1, Cuff electrode; 2, connector; 3, cable; 4, neurostimulator. (Courtesy C Varaart, Université Catholique de Louvain, Belgium.)

Later in 2000, a neurostimulator and an antenna were implanted below the skin and connected to the previously implanted spiral cuff electrode. The X-ray photograph in **Figure 10** shows the different parts of the implanted system (Veraart *et al.* 2003).

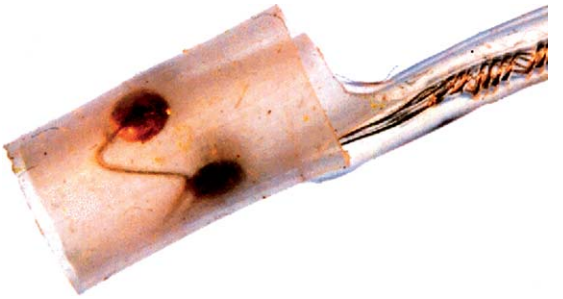


Figure 9 Photograph of the spiral cuff equipped with four electrodes. The electrode diameter is about 3 mm; stimulator dimensions are 25 mm × 30 mm with a 6-mm thickness. (Courtesy C Varaart, Université Catholique de Louvain, Belgium.)

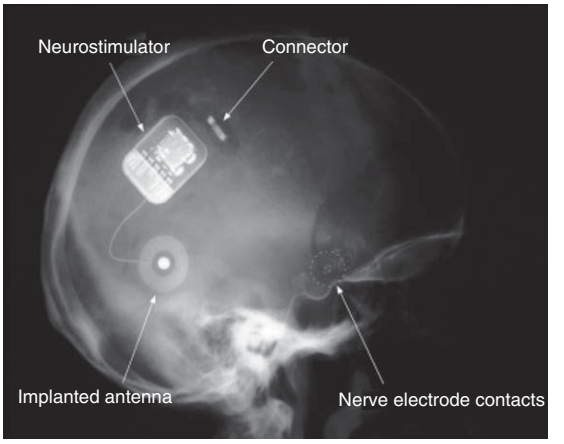


Figure 10 X-ray photograph of the implanted system. (Source: Reprinted from Veraart 2003.)

3.06.3.3 Stimulation of the Visual Cortex

Some patients have lost their vision because the fibers of the optic nerve lost their function. In these cases only a cortical implant can help. Like in epiretinal implants, the hardware components of such a system are a camera that converts light into electrical signals and an implantable interface to the nervous system. In addition, software that converts the image to an appropriate stimulation pattern, thus transforming an image to a retinotropic map, is needed. This information is transmitted to the cortical implant that has stimulation electrodes implanted in the visual cortex (**Figure 11**). By proper stimulation a phosphene pattern is produced. But there is a difference to retinal implants. A square electrode array will elicit a square grid of phosphenes (Dagnellie 2006). Stimulation of the visual cortex will lead to a

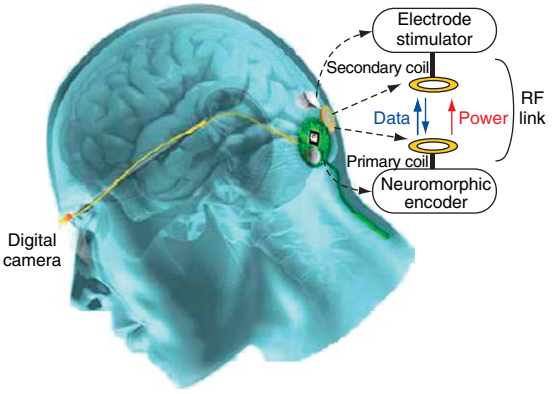


Figure 11 Schema of a cortical neuroprosthesis system. (Source: Reprinted from Piedade *et al.* 2005. © 2005 IEEE.)

phosphene pattern that is likely to be quite irregular. It is necessary to establish crude cortical maps by proper techniques before prosthesis implantation in order to know where to place the electrodes (Dagnellie 2006).

Dobelle *et al.* (1976) performed the implantation of a 64-channel platinum disk electrode array on the surface of the occipital cortex of blind patients. These electrodes were interfaced with a camera consisting of a 100×100 charge-coupled phototransistor array. Random letters were recorded by the camera. According to this information 6 out of the 64 electrodes were stimulated. The patients reported that they were able to recognize characters. However, stimulation on the brain's surface has some drawbacks: to stimulate the target cell in the brain a relatively high voltage is needed and it is not possible to stimulate small groups of cells.

These disadvantages of planar surface electrodes can be overcome by using penetrating electrodes. They can be in direct contact with the target cell. Thus, the voltage needed is drastically reduced (Maynard 2001, Troyk *et al.* 2003). **Figure 12(a)** and **12(b)** shows the configuration Troyk *et al.* (2003) used in their experiments. **Figure 13(a)** and **13(b)** shows two examples of penetrating electrodes developed at the University of Utah (Maynard 2001).

The needle electrodes are about 80 µm in diameter at the base and have a length of 1.5 mm (typically 100 in a 10×10 square grid). They are made from doped silicon. Besides the sharpened tip, the needles are electrically isolated. Each electrode is connected via a bonded, isolated 25-µm wire to a percutaneous connector. Electrode impedances of 100–500 kΩ were measured with a current of

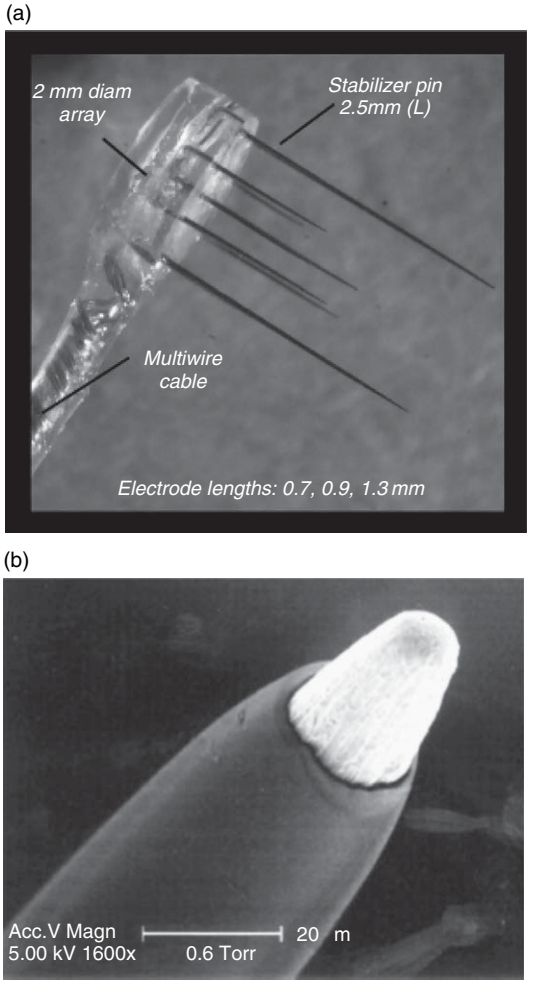


Figure 12 (a) Configuration of HMRI (Hunter Medical Research Institute) arrays used for implantation. The long stabilizer pins help to maintain the position of the array in the cortex (Troyk *et al.* 2003). (b) Scanning electron micrograph of a typical microelectrode tip showing the Parylene insulation and the exposed iridium tip (Troyk *et al.* 2003). (Courtesy P R Troyk, Illinois Institute of Technology, USA, with permission of Blackwell Publishing.)

100 nA at 1 kHz. These electrodes were implanted successfully in trained animals for 100 days.

While the cortical implant has the potential to help a greater number of people than does the retinal implant, it may require a longer investigation period because it involves implanting a device into the brain and because it has not been studied as much as the retinal implant.

3.06.3.4 Biohybrid Retinal Implants

A biohybrid retinal implant combines regenerative medicine with retinal implants. Microelectromechanical

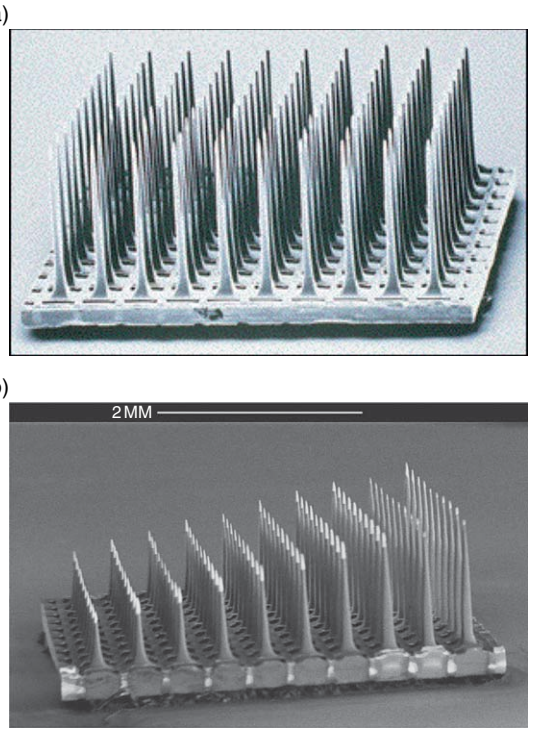


Figure 13 A typical Utah Electrode Array. (b) A modification of the Utah Electrode Array in which the length of the electrodes is uniformly graded. (Source: Maynard E 2001 visual prostheses. *Annu. Rev. Eng.* **3**, 145–68.)

systems (MEMS) is combined with nerve cells that are grown on top of stimulation electrodes. In an epiretinal approach Yagi from Nagoya University cultured nerve cells on top of electrodes (Yagi *et al.* 1999). If nerve cells and Schwann cells are brought together, the lengthening of nerve cells is promoted by factors that are produced in the Schwann cells. An artificial optic nerve can be prepared from Schwann cells and the extracellular matrix. The axons of the nerve cells are guided to the higher visual cortex. They are used as a cable to connect the MEMS electrodes with the visual cortex (**Figure 14**). It is assumed that after a connection of the nerve fibers with nerve cells of the visual cortex is formed, the nerve fibers transmit signals to the visual cortex in response to electrical stimulation of the fibers by a stimulation electrode array (Yagi *et al.* 1999). The hybrid retinal implant requires neither the retinal ganglion cells nor the optic nerve.

For successful electrical stimulation it is necessary that the stimulation electrodes be in close proximity to the nerve cell. This is being studied at Stanford University by using a subretinal approach. The nerve cells are brought in the proximity of the stimulation sites by so-called retinal migration. Retinal cells migrate rapidly (within 48–72 h) into a perforated subretinal implant while preserving their axonal connections to the retina above the implant (Palanker *et al.* 2004a, b, 2005). Within the pores in the implant, stimulating electrodes can be positioned and in this way an intimate proximity between electrodes and target cells is automatically achieved along the whole surface of the implant (**Figure 15**) (Palanker *et al.* 2005).

3.06.4 Technical Challenges

3.06.4.1 Fabrication of Flexible Retinal Implants

During surgery the implant is pushed into the eye through a tiny cut in the cornea. Flexibility and foldability of the implant allow for a smaller cut and

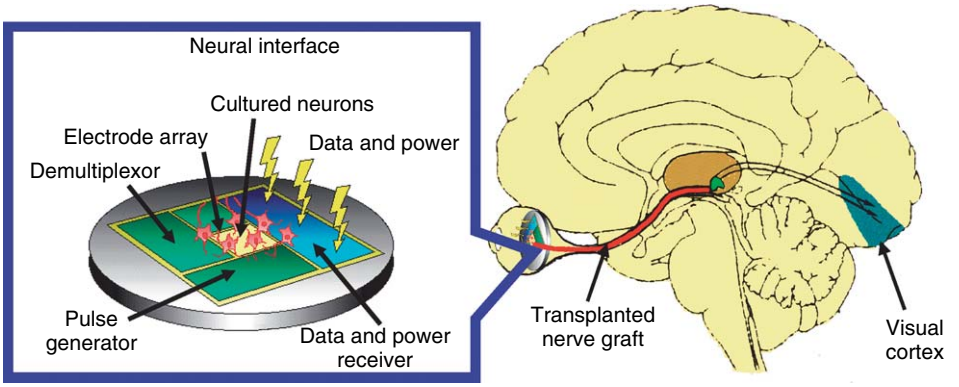


Figure 14 Schematic view of the system proposed by Yagi (Yagi *et al.* 1999). The neural interface consists of a silicon chip with circuitry that communicates with an external system. Neurons are cultured on an electrode array and will grow through a transplanted nerve graft toward the visual cortex. (Source: Yagi T *et al.* 1999 Hybrid retinal implant: Fusion of engineering and neuroscience. *Proc. IEEE Int. Conf. Syst. Man Cybern.* **4**, 382–5.)

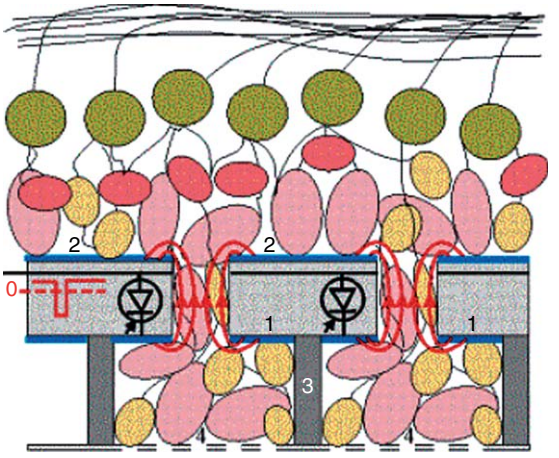


Figure 15 Concept of protruding electrodes on the subretinal array penetrating deep into the retina after migration of the retinal cells into the empty spaces between the pillars. Penetration depth is set according to the length of the pillars, which are insulated at the sides and exposed at the top. (Source: Palanker D *et al*. Design of a high resolution optoelectronic retinal hypothesis. *J. Neural Eng.* **2**, S105–20; reproduced with permission from IOP Publishing Ltd.)

hence less invasive surgery. Because of these requirements, the production of the implant by using a thin polyimide foil and a flexible electroplated coil seems to be the most promising choice (Mokwa 2004).

A typical fabrication process is described in Hungar *et al.* (2005). The polyimide foils are produced using standard and nonstandard wafer-level processes. Silicon wafers (4″) coated with a sacrificial layer are used as substrates for the production process. Polyimide is spin-coated onto the sacrificial layer and structured along the outlines of the implants and around the electrode array in the form of cloverleaves. This structuring is necessary not only to define the separation lines between implants but also to speed up the etching of the sacrificial layer at the end of the production process. In an electroplating step, gold wiring is grown before a second polyimide film is spin-coated and patterned. At this point, the trenches forming the outlines of the implant and the cloverleaves of the electrode array are structured and the openings of the pads and electrodes are etched. A second gold electroplating step is performed to add 26 µm to the height of the electrodes. At this point, the implants could in principle be assembled to fully functional telemetric eye prostheses. However, the electrode material gold has some disadvantages with regard to electrical stimulation of nerve cells. Therefore, a suitable coating material like iridium oxide is added to the electrodes. The coating is done by reactive DC magnetron sputtering from an iridium target in an argon/oxygen plasma (Slavcheva *et al.* 2004). Parylene C is then added from a gas phase to form a biocompatible protective coating. Parylene C has a long history of use in permanent medical implants, and it is an FDA-approved USP class VI polymer and is hence suitable for chronic implantation (Schmidt *et al.* 1988, Yuen *et al.* 1987). On the electrodes, holes are etched into the Parylene C film so that the electrodes can electrically contact the ganglion cells of the retina. A schematic view of the cross section of an electrode and an implant is shown in **Figure 16**.

Finally, the implants are removed from the wafer by etching away the sacrificial layer (Hungar *et al.* 2005). Then the implant is coated with silicone as an additional biocompatible protective layer. As the electrodes have to remain free from silicone, a special mould must be used for this process step. The complete encapsulated implant is shown in **Figure 6**. Biocompatibility and long-term stability of this implant were studied in a simple model (Stieglitz *et al.* 2003). No degradation was found within a period of 1 year.

Figure 17 shows a photograph of a second-generation prototype of the implant shown in **Figure 5** before encapsulation with silicone, with the main components mounted onto a flexible polyimide tape (from left to right): receiver coil, capacitor, diode,

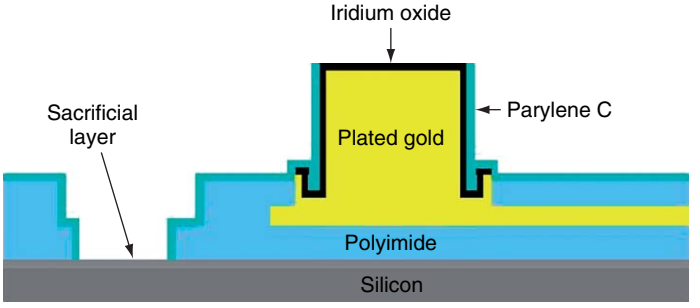


Figure 16 Schematic view of wafer-level cross section of an electrode and an implant.

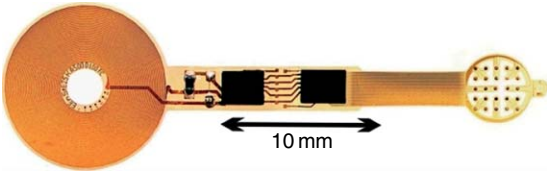


Figure 17 Second-generation prototype of a telemetric eye prosthesis. (Source: Mokwa 2004 MEMS technologies for epiretinal stimulation of the retina. *J. Micromech. Microeng.* **14**, 12–16.)

receiver chip, stimulation chip, microcable, and a 5×5 array of 3D iridium oxide-covered gold electrodes (Mokwa 2004, Wessling *et al.* 2005).

Around the electrodes the polyimide tape is patterned in the shape of cloverleaves, with each carrying four electrodes. This gives the implant additional flexibility, allowing it to be better adapted to the round shape of the retina and hence improving electrical contact between the electrodes and the retina.

3.06.4.2 Stimulation Electrodes

3.06.4.2.1 Requirements for stimulation electrodes

In view of their specific application the implantable electrodes must fulfill several very stringent requirements such as a miniaturized geometry (to match the limited space in the biological units and to ensure a good stimulation selectivity), softness and flexibility (to prevent damage to the surrounding tissue due to mechanical stress), and long-term stability and corrosion resistance (to withstand the corrosive attack of the physiological environment). Biocompatibility is another very important feature of the stimulating electrodes because stimulation must be accomplished without leakage of toxic products that cause damage to the neural cells. However, the most essential precondition for the success of functional electrostimulation (FES) is the availability of highly electroactive electrode materials capable of delivering enough charges across the phase boundary stimulating electrode/stimulated tissue as well as the low impedance of this boundary, especially in the frequency range of 10 Hz to 1 kHz, which is relevant for neural stimulation (Blau *et al.* 1997, Fröhling *et al.* 1996). It is difficult to find a material conforming to all these requirements. Thus, the practical decisions taken are often a compromise.

3.06.4.2.2 Research on stimulation electrodes

In the search for the ideal candidate, microelectrodes of different shapes and sizes have been fabricated, including planar disk electrodes (Agnew *et al.* 1983, Hesse *et al.* 2000, McCreery *et al.* 1990), tethered microwires and intrafascicular electrodes (Googall *et al.* 1993, Naughton and Horch 1996, Silva *et al.* 1998, Yoshida and Horch 1993a, b), cuff electrodes (Duysens and Stein 1978, Meier *et al.* 1986, Schuettler *et al.* 2002, Veraart *et al.* 1998), and various microfabricated 2D and 3D microelectrode arrays (Anderson *et al.* 1989, Rutten 2002, Tanghe *et al.* 1990, Weiland and Anderson 2000). The electroactive metallic materials that have been tested as stimulating electrodes are numerous, the ones most frequently used being Ti, TiN, and different Ti alloys (Blau *et al.* 1997, Haemmerle *et al.* 2002, Walter and Heimann 2000), Au (Haemmerle *et al.* 2002, Hung *et al.* 2002), Pt and some of its alloys such as Pt/W and Pt/Ir (Hesse *et al.* 2000, Lee *et al.* 2002, Yoshida and Horch 1993b), and most recently Ir and IrO_x (Anderson *et al.* 1989, Blau *et al.* 1997, Bolz *et al.* 1995, Fröhling *et al.* 1996, Fröhling *et al.* 1998, Lee *et al.* 2002, Walter and Heimann 2000, Weiland and Anderson 2000).

The performance of a metallic electrode in a physiological fluid environment is mainly determined by the properties of the phase boundary electrode/tissue. This boundary is a complex system exhibiting nonohmic behavior. From a physical point of view it acts as an electrochemical transducer exchanging electrons with ions, the current carriers in the electrode and the electrolyte (body fluid), respectively. As the stimulation of the neuronal action potential depends on the amount of charge transferred across the phase boundary, it is important to select an electrode material that is capable of delivering high charge densities reversibly, i.e., a material that possesses a high charge delivery capacity, Q_{CDC}– a parameter that describes the highest possible charge accumulation at the electrode in a defined environment. The charge delivery must be accomplished safely, which means without occurrence of irreversible processes such as gas evolution due to water electrolysis, major pH changes, metal dissolution, or oxidation of organics that cause damage to the neural cells. A reversible charge injection is possible by two basic mechanisms: capacitive charging and discharging of the electrical double layer formed on the phase boundary and Faradaic redox reactions involving species that remain bound

to the electrode surface. It is ideal from a safety point of view to stimulate only capacitively, and the initial focus has been on the search for capacitive electrodes. An example of such an electrode material is TiN. It has been extensively tested and has shown typical capacitive electrochemical behavior and a very good corrosion resistance in different electrolytes (Haemmerle *et al.* 2002, Janders *et al.* 1997, Walter and Heimann 2000). The capacitive stimulation, although safe, has one essential disadvantage. It seriously limits the transferable charge: generally only 20–30 $\mu C\,cm^{-2}$ are passed during the charging of the electrical double layer of a smooth electrode material (Cambell and Jones 1992). This layer behaves as a parallel plate capacitor and its capacitance is proportional to the electrochemically active surface.

3.06.4.2.3 Enhancement of charge delivery capacity

The miniaturization of the stimulating electrodes enhances the spatial resolution and the selectivity of the stimulation. On the other hand, the smaller the electrode the lower the charge transfer and the higher the impedance of the phase boundary. In order to overcome this contradiction an active area that is as large as possible must be integrated on an electrode of limited size. This can be achieved to some extent by preparing macroporous structures through sintering, etching, electrochemical plating, and plasma-enhanced chemical or physical vapor deposition (Bolz *et al.* 1995, Fröhling *et al.* 1996, Hung *et al.* 2002, Janders *et al.* 1997, Naughton and Horch 1996).

The Faradaic mechanism is an alternative stimulation mechanism that requires the usage of electroactive electrode materials capable of transferring high charges through reversible redox reactions in the safe water window potential range, i.e., between the potentials of H_2 and O_2 evolution. In this case the current flow across the phase boundary is due to changes in the oxidation state of the participating materials. The ability of various noble metals to act through this stimulation mechanism has been studied extensively. For instance, the silver–silver chloride electrode (Ag/AgCl) has been widely used for measuring the action potential. However, it has been found that when in contact with biological tissue, the silver chloride on the surface dissolves and causes inflammations due to its toxicity. That limits its application for stimulation purposes (Moussy and Harrison 1994, Yuen *et al.* 1987). Application of gold, another noble metal known for its biocompatibility, is also restricted (Haemmerle *et al.* 2002, Janders *et al.* 1997) due to the low current densities of its redox reactions in physiological solutions and the narrow potential range in which it shows reversible electrochemical behavior (Slavcheva *et al.* 2002). The elements of the platinum group have received special attention because they have at least two different oxidation numbers and hence higher electrochemical activity. Some of these metals like osmium are toxic, while the charge-delivering capacity of others, for example, palladium, is not high enough (Fröhling 1996). One of the metals that has been used often in FES is platinum (Anderson *et al.* 1989, Hesse *et al.* 2000, Hung *et al.* 1999, Humayun *et al.* 1996, Lee *et al.* 2002, Naughton and Horch 1996, Walter and Heimann 2000, Yoshida and Horch 1993b). It has a stable reversible electrochemical behavior in a comparatively broad potential range (from –0.6 to 1.0 V). Its charge delivery capacity is between 100 and 400 $\mu C\,cm^{-2}$ (Brummer and Turner 1977a, b, Brummer *et al.* 1983, Cambell and Jones 1992) and can be increased by electroplating platinum black on the top of a platinum electrode (Brummer *et al.* 1983, Cambell and Jones 1992, Naughton and Horch 1996). However, these electroplated platinum black films have a flimsy dendritic structure and are mechanically not stable. Hence some amount of charge is always lost due to metal dissolution, especially when used in the long term (Brummer and Turner 1977a, b, Fröhling *et al.* 1996, Hung *et al.* 2002, Janders *et al.* 1997). Some platinum alloys (platinum/tungsten, platinum/iridium) have also been tested as stimulating electrodes. For instance, 70% platinum/30% iridium has been used because of its greater mechanical strength when compared to pure platinum (Robblee *et al.* 1983). Another alloy, 90% platinum/10% iridium, has recently been used in heart pacemakers. A drawback of platinum and its alloys, according to some authors (Gielen and Bergveld 1982), is their unstable impedance in the low-frequency range. Other metals from the platinum group, namely rhodium, ruthenium, and iridium, can take up to four different oxidation states and thus they possess an increased electroactivity. However, it has been demonstrated that in the case of rhodium and ruthenium, the low initial impedance of the phase boundary electrode/electrolyte increases essentially during electrochemical loading and that their long-term stability is not adequate for FES (Fröhling *et al.* 1996). Thus, among all the tested materials, iridium seems to be the most promising

candidate for implantable stimulating electrodes. It is an attractive alternative to platinum mainly because it has four oxidation states (platinum has two) and has shown a superior corrosion resistance. Another characteristic feature of iridium that is beneficial for FES is its ability to form a contact with an aqueous environment, hydrated oxide surface layers, resulting in an increased electrochemically active surface and a higher charge delivery capacity (Robblee *et al.* 1983, Weiland and Anderson 2000). The formation of these oxide layers requires continuous potential cycling between defined lower and upper potential limits specific for the electrolyte medium. The films obtained are known as anodic/activated iridium oxide films (AIROFs). It has been unambiguously proved that AIROFs have a 100-fold higher charge delivery capacity compared to nonactivated pure iridium electrode and a very broad safe potential range (free of gas generation). The problem with these films from the FES point of view arises in connection with their long-term stability (Blau *et al.* 1997).

Another method for producing iridium oxide is by reactive sputtering of iridium in an oxygen atmosphere (Bestaoui *et al.* 1993, Wessling *et al.* 2005). The obtained films, known as sputtered iridium oxide films (SIROFs), have a highly developed fractal surface with a microporous structure and a density of about $10\,g\,cm^{-3}$ (**Figure 18**). SIROFs have unusual electrocatalytic properties and are highly efficient anodes for water electrolysis. They possess a very low overvoltage for oxygen evolution and are stable at high potentials where other catalysts corrode at a significant rate. They also exhibit an outstanding charge delivery capacity (**Table 1**).

Table 1 Charge delivery capacity of different metals

Material	Q_{CDC} ($\mu C\ cm^{-2}$)
Au	490
TiN	687
Pt	4134
Ir	17 078
IrO_x	28 450
IrO_x (after activation)	95 100

3.06.4.3 Appearance of Phosphenes

Phosphenes caused by electrical stimulation may be fuzzy, irregular in shape, varying in size and color, and limited in brightness range and gray-scale resolution; their positions may be scattered relative to an idealized grid (Dagnellie 2006). In retinal prostheses the positional irregularities can be expected to be small, but they can be widely scattered in cortical implants. Therefore, it is very important to determine these properties for each phosphene and to perform inverse image transformation to achieve the closest approximation between the original and the perceived pictures (Dagnellie 2006). For an epiretinal system, Eckmiller and others (Eckmiller 1995, 1997, Eckmiller *et al.* 2005) proposed an adaptive learning retina encoder to perform image preprocessing to optimize stimulus perception by the implant wearer (**Figure 19**). This has to be an interactive and iterative process.

The retina encoder was designed as a neural net with flexible antagonistic receptive field properties (receptive field–adaptive biology–inspired pulse

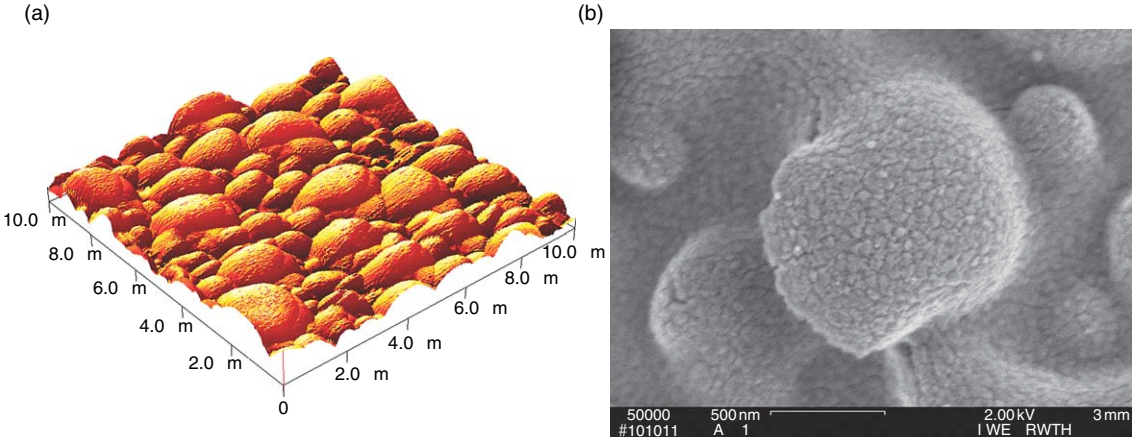


Figure 18 (a) Sputtered iridium oxide film (SIROF) surface after activation (atomic force microscopy). (b) SIROF surface after activation (scanning electron microscopy).

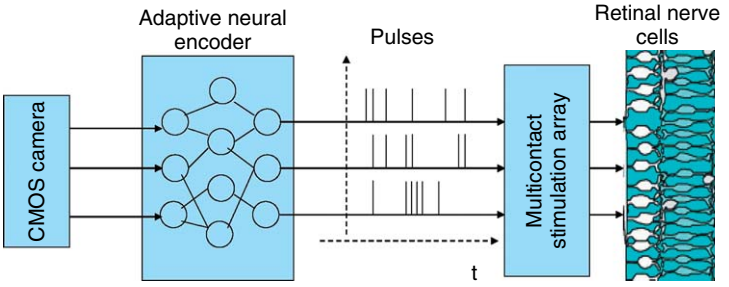


Figure 19 Working scheme of an adaptive encoder.

processing neural network, RF-BPN). The RF-BPN module allows for modification of spatial and/or temporal parameters in order to adjust the RF-BPN to the desired receptive field properties of a given ganglion cell. Specifically, in a learning process each RF-BPN module can be tuned to the receptive field properties of different cell types in the retina with regard to parameters such as RF center and time constants (Eckmiller 1995).

3.06.5 Clinical Studies

3.06.5.1 Verification of Visual Cortex Activation by Epiretinal Stimulation of Animals

In different experimental approaches it could be proven that electrical stimulation of the retina of animals leads to relevant activation of the corresponding areas of the visual cortex. Walter and Heimann (2000) proved the activation by noninvasive recording of evoked potentials after epiretinal electrical stimulation. By recording neuron activity with intracortical electrodes, it could be shown by Schanze, Eckhorn, and coworkers that the stimulation of a distinct area of the retina leads to cortical activation of the corresponding area. The cortex activations after electrical stimulation take place at the same area where the retinal ganglion cells are represented in natural sight (Eger *et al.* 2005, Schanze *et al.* 2002, 2003).

Complete epiretinal systems (compare **Figure 6**) were implanted into mini pigs and cats. Full telemetric functioning could be shown with several systems. **Figure 20** shows a cortical recording from a mini pig after telemetric stimulation of retinal ganglion cells. As expected, the response can be seen in about 25 ms after stimulation (Laube *et al.* 2004a, b).

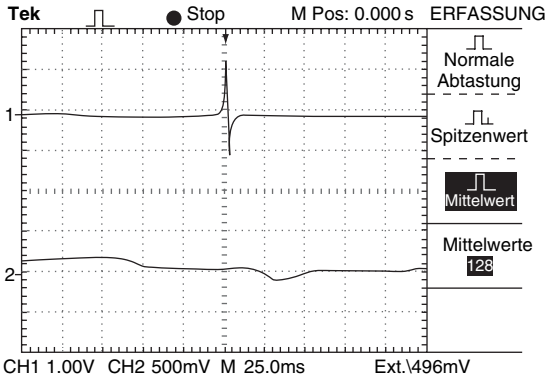


Figure 20 Cortical recording after electrical stimulation of the retinal ganglion cells of a mini pig by a retinal implant (system shown in **Figure 6**); the upper curve shows a stimulation artifact; the lower curve shows the cortical response after electrical stimulation. (Courtesy Laube, Department of Ophthalmology University Essen, Germany.)

In **Figure 21** the area of cortical activity of a cat during electrical stimulation is given. This picture was obtained by the method of optical imaging. The method distinguishes between oxygenized and deoxygenized hemoglobin. Thus, the areas of active and less active metabolism can be distinguished. The areas of increased activation in **Figure 20** correspond to the areas on the retina that had been electrically stimulated (Walter *et al.* 2005).

3.06.5.2 Experiences with Epiretinal Stimulation in Humans

Currently a lot of research activities are going on for developing implantable retina stimulators. The development of epiretinal implants seems to be relatively close to a product. Several working groups have been conducting acute trials on the question whether blind RP patients recognize visual sensations after electrical stimulation. In these trials, electrodes were

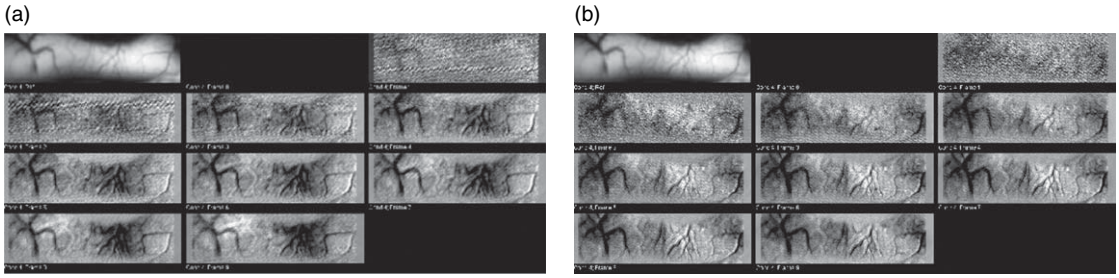


Figure 21 (a) Optical imaging from the visual cortex of a cat during electrical stimulation of the corresponding regions on the retina. The black regions in the right part of the pictures show increased oxygen consumption in areas of the visual cortex that correspond to the stimulated areas of the retina (Walter *et al.* 2005). (b) Control experiment without stimulation pulses (Walter *et al.* 2005). (Courtesy P Walter, Department of Ophthalmology, University Hospital, RWTH Aachen University, Aachen, Germany.)

placed on the retina under the conditions of vitrectomy and the patients were asked whether they had visual sensations. By this the current threshold for stimulation was obtained. The results differ very much partly because of the different experimental conditions for the electrode configuration. Values between several microamperes and milliamperes were reported (Humayun *et al.* 1996, 2003, Rizzo *et al.* 2003a, b).

The US company Second Sight has already implanted a system with 16 stimulation electrodes for chronic application in six patients. The German company IIP Technologies used their system for semichronic studies in two patients (Wickelgren 2006).

3.06.6 Conclusions

Various successful approaches for developing an artificial retina have been presented in this chapter. It is estimated that in a couple of years complete systems will be available for chronic implantation in humans. With respect to the total retina implant system there are still many unanswered and maybe unknown questions with regard to the high complexity of the system. Questions concerning a long-term stable fixation, a long-term stable stimulation, and the minimization of power consumption of the implant have to be answered before chronic implantation in humans.

These examples have shown that MEMS technology is a powerful tool for realizing small and complex electronic implants. This technology will help to increase the functionality of existing systems in a significant way. New applications that nobody thought that they were possible some years ago will be possible. These intelligent implants and prostheses will support the field of home care to a great extent. They will also have a strong impact on clinical research. Implants of the future will have an increased biological part like the biohybrids mentioned earlier.

References

Agnew W F, Yuen T G H, McCreery D B 1983 Morphologic changes after prolonged electrical stimulation of the cat's cortex at defined charge densities. *Exp. Neurol.* **79**, 397–441

Anderson D, Najafi K, Tanghe S J, Evans D A, Levy K L, Hetke J F, Xue X, Zappia J J, Wise K D 1989 Batch-fabricated thin-film electrodes for stimulation of the central auditory system. *IEEE Trans. Biomed. Eng.* **36**, 639–703

Benjamin A, Humayun M S, Hickingbotham S, de Juan E, van den Honert C 1994 Characterization of retinal responses to electrical stimulation of retinal surface of *Rana catesbeiana*. *Invest. Ophthalmol. Vis. Sci.* **35(Suppl.)**, 1832

Bestaoui N, Prouzet E, Deniard P, Brec R 1993 *Thin Solid Films* **235**, 35–42

Blau A, Ziegler C H, Heyer M, Enderst F, Schwitzgebelt G, Matties T, Stieglitz T, Meyer J U, Goepel W 1997 Characterization and optimization of microelectrode arrays for *in vivo* nerve signal recording and stimulation. *Biosens. Bioelectron.* **12**, 883–92

Bolz A, Froehling R, Schmidt K, Schaldach M 1995 Effect of smooth, porous and fractal surface structure on the properties of an interface *J. Mater. Sci. Mater. Med.* **6**, 844–8

Brindley G S, Levin W S 1968 The sensations produced by electrical stimulation of the visual cortex. *J. Physiol. (Lond.)* **196**, 479–93

Brummer S B, Turner M T 1977a Electrochemical considerations for safe electrical stimulation of the nervous system with platinum electrodes. *IEEE Trans. Biomed. Eng.* **BME-24**, 59–63

Brummer S B, Turner M T 1977b Electrical stimulation with Pt electrodes. Estimation of maximum surface redox (theoretical non-gassing) limits. *IEEE Trans. Biomed. Eng.* **BME-24**, 440–3

Brummer S B, Robblee L S, Hambrecher F T 1983 Criteria for selecting electrodes for electrical stimulation: Theoretical and practical considerations. *Ann. N Y Acad. Sci.* **405**, 159–71

Cambell P K, Jones K E 1992 In: Cahn R W, Haasen P, and Kramer E J (eds.) *Materials Science and Technology.* VCH Publishers, New York, Vol. 14, p. 346

Chow A Y 1993 Electrical stimulation of the rabbit retina with subretinal electrodes and high density microphotodiode array implants. *Investig. Ophthalmol. Vis. Sci.* **34**, 835

Dagnellie 2006 Visual prosthetics 2006: Assessment and expectations. *Expert Rev. Med. Dev.* **3(3)**, 315–25

Delbecke J, Oozeer M, Veraart C 2003 Position, size and luminosity of phosphenes generated by direct optic nerve stimulation. *Vis. Res.* **43**, 1091–102

Dobelle W H, Mladejovsky M G, Evans J R, Roberts T S, Girvin J P 1976 'Braille' reading by a blind volunteer by visual cortex stimulation. *Nature* **259**, 111–2

Duysens J, Stein R B 1978 Reflexes induced by nerve stimulation in walking cats with implanted cuff electrodes. *Exp. Brain Res.* **32**, 213–24

Eckmiller R 1995 Towards retina implants for improvement of vision in human with RP – Challenges and first results. *Proc. World Congr. Neural Networks*. INNS Press Hillsdale, NJ, Vol. 1, pp. 228–33

Eckmiller R 1996 Retina implants with adaptive retina encoders. *Proc. RESNA Res. Symp.* **21**, 21–4

Eckmiller R 1997 Learning retina implants with epiretinal contacts. *Ophthal. Res.* **29(5)**, 281–9

Eckmiller R, Neumann D, Baruth O 2005 Tunalble retina encoders for retina implants: Why and how. *J. Neural Eng.* **2**, 91–104

Eger M, Wilms M, Eckhorn R, Schanze T, Hesse L 2005 Retino-cortical information transmission achievable with a retina implant. *Biosystems* **79(1–3)**, 133–42

Fröhling G, Rzany A, Riedmueller J, Bolz A, Schaldach M 1996 Electroactive coating of stimulating electrodes. *J. Mater. Sci. Mater. Med.* **7**, 393–7

Fröhling G, Bolz A, Stroebel J, Rutz M, Lawall P, Schwerdoe H, Schaldach M, Schieffer H 1998 *PACE* **21**, 1239–46

Gielen F L H, Bergveld P 1982 Comparison of electrode impedances of Pt, PtIr (10% Ir) and Ir-AIROF electrodes used in electrophysiological experiments. *Med. Biol. Eng. Comput.* **20**, 77–83

Googall E V, Horch K W, McNaughton T G, Lybbert C M 1993 Analysis of single-unit firing patterns in multi-unit intrafascicular recordings. *Med. Biol. Eng. Comput.* **31**, 257–67

Haemmerle H, Kobuch K, Kohler K, Nisch W, Sachs H, Stelzle M 2002 Biostability of micro-photodiode arrays for subretinal implantation. *Biomaterials* **23**, 797–804

Hesse L, Schanze T, Wilms M, Eger M 2000 Implantation of retina stimulation electrodes and recording of electrical stimulation responses in the visual cortex of the cat. *Graefes Arch. Clin. Exp. Ophthalmol.* **238**, 840–5

Humayun M S, de Juan E, Dagnelie G, Greenberg R, Propst R, Phillips H 1996 Visual perception elicited by electrical stimulation of the retina in blind humans. *Arch. Ophthalmol.* **114**, 40–6

Humayun M S, Prince M, de Juan E Jr., Barron Y, Moskowitz M, Klock I B, Milam A H 1999 Morphometric analysis of the extramacular retina from postmortem eyes with retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **40(1)**, 143–8

Humayun M, Weiland J D, Fujii G Y, Greenberg R, Williamson R, Little J, Mech B, Cimmarusti V, Van Boemel G, Dagnelie G, de Juan E 2003 Visual perceptions in a blind subject with a chronic microelectronic retinal prosthesis. *Vis. Res.* **43**, 2573–81

Hung A, Zhou D, Greenberg R, Judy J W 2002 Micromachined electrodes for high density neural stimulation systems. *Proc. 2nd Annu. Int. IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine and Biology*. Madison, WI, USA, May 2–4, 2002

Hungar K, Slavcheva E, Spanier G, Mokwa W, Görtz M 2005 Production processes for a flexible retina implant. *Sens. Actuators* **123–4C**, 172–8

Janders M, Egert U, Steltze M, Nish W 1997 *Proc. 18th Annu. Int. Conf. Bridging Disciplines for Biomedicine, IEEE-EMBS*, Amsterdam, The Netherlands, pp. 245–7

Kaufman P L, Alm A 2003 *Adler's Physiology of the Eye*, 10th edn. Mosby, St. Louis, MO

Kuritz T, Lee I, Owens E T, Humayun M, Greenbaum E 2005 Molecular photovoltaics and photoactivation of mammalian cells. *IEEE Trans. Nanobiosci.* **4**, 196–200

Laube T, Brockmann C, Buß R, Lau C, Höck K, Stawski N, Stieglitz T, Richter H A, Schilling H 2004a Optical energy transfer for intraocular microsystems studied in rabbits. *Graefes Arch. Clin. Exp. Ophthalmol.* **242(8)**, 661–7

Laube T, Akguel H, Schanze T, Görtz M, Bolle I, Brockmann C, Bornfeld N 2004b First time successful epiretinal stimulation with active wireless retinal implants in Göttinger minipigs. *Arvo poster no. 4188/B649, Ford Lauderdale*

Lee I S, Wang Ch N, Choi K, Choo M S, Lee J H 2002 Characterization of iridium film as a stimulating neural electrode. *Biomaterials* **23**, 2375–80

Maynard E 2001 Visual prostheses. *Annu. Rev. Biomed. Eng.* **3**, 145–68

McCreery D B, Agnew W F, Yuen T G H, Bullara L 1990 Charge density and charge per phase as cofactors in neural injury induced by electrical stimulation. *IEEE Trans. Biomed. Eng.* **37**, 996–1000

Mokwa W 2004 MEMS technologies for epiretinal stimulation of the retina. *J. Micromech. Microeng.* **14**, 12–16

Moussy F, Harrison J 1994 Prevention of the rapid degradation of subcutaneously implanted Ag/AgCl reference electrodes using polymer coatings. *Anal. Chem.* **66**, 674–9

Naughton T G, Horch K W 1996 Metallized polymer fibers as leadwires and intrafascicular microelectrodes. *J. Neurosci. Meth.* **70**, 103–10

Palanker D, Huie P, Vankov A, Aramant R, Seiler M, Fishman H, Marmor M, Blumenkranz M S 2004a Migration of retinal cells through a perforated membrane: Implications for a high-resolution prosthesis. *Invest. Ophthalmol. Vis. Sci.* **45(9)**, 3266–70

Palanker D, Huie P, Vankov A, Freyvert Y, Fishman H, Marmor M F, Blumenkranz M S 2004b Attracting retinal cells to electrodes for high-resolution stimulation. *Opthal. Technol., SPIE* **5314**, 306–12

Palanker D, Vankov H, Huie P, Baccus S 2005 Design of a high resolution optoelectronic retinal prosthesis. *J. Neural Eng.* **2**, S105–20

Peachey N S, Chow A Y 1999 Subretinal implantation of semiconductor-based photodiodes: Progress and challenges. *J. Rehabil. Res. Dev.* **36**, 371–6

Piedade M, Gerald R, Sousa L A, Tavares G, Tomás P 2005 Visual neuroprosthesis: A non invasive system for stimulating the cortex. *IEEE Trans. Circuits Syst. I Regular Pap.* **52(12)**, 2648–62

Rizzo J F III, Wyatt J, Loewenstein J, Kelly S, Shire D 2003a Methods and perceptual thresholds for short term electrical stimulation of human retina with microelectrode arrays. *Invest. Ophthalmol. Vis. Sci.* **44(12)**, 5355–61

Rizzo J F 3rd, Wyatt J, Loewenstein J, Kelly S, Shire D Perceptual efficacy of electrical stimulation of human retina with a microelectrode array during short term surgical trials. *Invest. Ophthalmol. Vis. Sci.* **44(12)**, 5362–9

Robblee L S, Lefko J L, Brummer S B 1983 An electrode suitable for reversible charge injection in saline solution. *J. Electrochem. Soc.* **130**, 771–3

Rutten W L C 2002 Selective electrical interfaces with the nerve system. *Annu. Rev. Biomed. Eng.* **4**, 407–52

Sachs H G, Gabel V P 2004 Retinal replacement the development of microelectronic retinal prostheses – Experience with subretinal implants and new aspects. *Graefes Arch. Clin. Exp. Ophthalmol.* **242(8)**, 717–23

Santos A, Humayun M S, de Juan E, Greenburg R J, Marsh M J, Klock I B, Milam A H 1997 Preservation of the inner retina in retinitis pigmentosa. A morphometric analysis. *Arch. Ophthalmol.* **115(4)**, 511–5

Schanze T, Wilms M, Eger M, Hesse L, Eckhorn R 2002 Activation zones in cat visual cortex evoked by electrical retina stimulation. *Graefes Arch. Clin. Exp. Ophthalmol.* **240(11)**, 947–54

Schanze T, Greve N, Hesse L 2003 Towards the cortical representation of form and motion stimuli generated by a retina implant. *Graefes Arch. Clin. Exp. Ophthalmol.* **241(8)**, 685–93

Schmidt E M, McIntosh J S, Bak M J 1988 Long-term implants of Parylene C coated microelectrodes. *Med. Biol. Eng. Comput.* **26**, 96–101

Schuettler M, Riso R R, Dalmose A, Stefania D, Stieglitz T 2002 Initial chronic results of flexible sieve electrodes as interface to nerve stumps. *Biomed. Technik* **47**, 696–9

Schwahn H N, Gekeler F, Kohler K, Kobuch K, Sachs H G, Schulmeyer F, Jakob W, Gabel V P, Zrenner E 2001 Studies on the feasibility of a subretinal visual prosthesis: Data from Yucatan micropig and rabbit. *Graefes Arch. Clin. Exp. Ophthalmol.* **239(12)**, 961–7

Schwarz M, Ewe L, Hausschild R, Hosticka B J, Huppertz J, Kolnsberg S, Mokwa W, Trieu H K 2000 Micro implantable visual prostheses. In: *Proc. 1st Annu. Int. IEEE-EMBS Special Top. Conf. Microtechnologies in Medicine and Biology*, Piscataway, NJ, USA, pp. 461–5

Silva T M, Rito J E, Simoes A M P, Ferreira M G S, da Cunha Belo M, Watkins K G 1998 Electrochemical characterization of oxide films formed on Ti–6A1–4V alloy implanted with Ir for bioengineering applications. *Electrochem. Acta* **43**, 203–11

Slavcheva E, Ewe L, Schnakenberg U, Mokwa W 2002 Electrochemical characterization of different biocompatible metallic materials as planar and 3D-electrodes in neural stimulation microarrays. *Proc. 2nd Eur. Conf. Medical and Biological Engineering*, Vienna, Austria, Vol. 3(1), p. 785

Slavcheva E, Ewe L, Schnakenberg U, Mokwa W 2004 Sputtered iridium oxide films as charge injection material for functional electrostimulation. *J. Electrochem. Soc.* **151(7)**, E226–37

Stieglitz T, Haberer W, Kammer S, Schanze T 2003 A simple model to investigate the stability of flexible micromachined retina stimulators. In: *Proc 8th Annu. Int. Conf. International Functional Electrical Stimulation Society*, Maroochydore, Australia, pp. 69–73

Stone J L, Barlow W E, Humayun M S, de Juan E, Jr., Milam A H 1992 Morphometric analysis of macular photoreceptors and ganglion cells in retinas with retinitis pigmentosa. *Arch. Ophthalmol.* **110**, 1634–9

Tanghe S J, Najafi K, Wise K D 1990 A planar IrO multichannel stimulating electrode for use in neural prostheses. *Sens. Actuators* **B1**, 464–7

Troyk P, Bak M, Berg J, Bradley D, Cogan S, Erickson R, Kufta C, McCreery D, Schmidt E, Towle V 2003 A model for intracortical visual prosthesis research. *Artif. Organs* **27(11)**, 1005–15

Veraart C, Raftopoulos C, Mortimer J T, Delbeke J, Pins D, Michaux G, Vanlierde A, Parrini S, Wanet-Defalque M C 1998 Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode. *Brain Res.* **813(1)**, 181–6

Veraart C, Wanet-Defalque M C, Gerard V, Vanlierde A, Delbeke J 2003 Pattern recognition with the optic nerve visual prosthesis. *Artif. Organs* **27(11)**, 996–1004

Walter P, Heimann K 2000 Evoked cortical potentials after electrical stimulation of the inner retina in rabbits. *Graefes Arch. Clin. Exp. Ophthalmol.* **238(4)**, 315–8

Walter P, Kisvarday Z F, Roessler G F, Alteheld N M, Goertz M, Stieglitz T, Eysel U T 2005 Optical activation via an implanted wireless retinal prosthesis. *Invest. Ophthalmol. Vis. Sci.* **46(5)**, 1780–5

Weiland J D, Anderson D J 2000 Chronic neural stimulation with thin-film, iridium oxide electrodes. *IEEE Trans. Biomed. Eng.* **47**, 911–8

Wessling B, Hungar K, Slavcheva E, Mokwa W, Schnakenberg U 2005 Sputtered iridium oxide as a top layer of stimulating 3D microelectrodes in retina implants. *Biomed. Eng.* **50, Suppl, Vol. 1, Part 1**, 824–3

Wickelgren I 2006 A vision for the blind. *Science* **312**, 1124–6

Wyatt J L, Rizzo J L, Grumet A, Edell D, Jensen R J 1994 Development of a silicon retinal implant: Epiretinal stimulation of retinal ganglion cells in the rabbit. *Invest. Ophthalmol. Vis. Sci.* **35(Suppl.)**, 1380

Yagi T, Ito Y, Kanda H, Tanaka S, Watanabe M, Uchikawa Y 1999 Hybrid retinal implant: Fusion of engineering and neuroscience. *Proc. IEEE Int. Conf. Syst. Man Cybern.* **4**, 382–5

Yoshida K, Horch K 1993a Reduced fatigue in electrically stimulated muscle using dual channel intrafascicular electrodes with interleaved stimulation. *Ann. Biomed. Eng.* **21**, 709–14

Yoshida K, Horch K 1993b Selective stimulation of peripheral nerve fibers using dual intrafascicular electrodes. *IEEE Trans. Biomed Eng.* **40**, 492–4

Yuen G T, Agnew W A, Bullara L A 1987 Tissue response to potential neuroprosthetic materials implanted subdurally. *Biomaterials* **8**, 138–41

Zrenner E 2002 Will retinal implants restore vision? *Science* **295**, 1022–5

Zrenner E, Stett A, Weiss S, Aramant R B, Guenther E, Kohler K, Miliczek K D, Seiler M J, Haemmerle H 1999 Can subretinal microphotodiodes successfully replace degenerated photoreceptors? *Vis. Res.* **39(15)**, 2555

Biography

Wilfried Mokwa received the Dipl.-Phys. degree in Physics from the RWTH Aachen University, Germany, in 1977 and the Dr. rer. nat. degree in 1981 from the same university. From 1981 on he worked at the Department of Experimental Condensed Matter Physics at the RWTH Aachen University on studies of catalytic reactions on gas sensor surfaces. Then he joined the Fraunhofer Institute of Microelectronic Circuits and Systems (IMS) in Duisburg, Germany, in 1985. At the IMS he managed a group working on the monolithic integration of silicon sensors. In 1996 he became a full professor in the Faculty of Electrical Engineering at the RWTH Aachen University where he is director of chair 1 of the Institute of Materials in Electrical Engineering (IWE-1) with special interests in the field of MEMS technologies for medical applications.

3.07 Optical Communications

Dan M. Marom, The Hebrew University of Jerusalem, Jerusalem, Israel

Glossary

ATM Asynchronous transfer mode
BLSR Bidirectional Line Switched Ring
DBR Distributed Bragg Reflector
DFB Distributed Feedback
DMEMS Diffractive MEMS
DRIE Deep Reactive Ion Etch
DUV Deep Ultraviolet
DWDM Dense Wavelength Division Multiplexing
FSR Free-Spectral Range
GMPLS Generalized Multi-Protocol Label Switching
GR Generic Requirement
ILA In-Line Amplifier
LC Liquid-crystal
MARS Mechanical Anti-Reflection Switch
MEMS Micro-Electro-Mechanical System
MSPP Multi-Service Provisioning Platform
OADM Optical Add-Drop Multiplexer
OEO Optical Electrical Optical
OEOEO Optical Electrical Optical Electrical Optical
OOO Optical Optical Optical
OSNR Optical Signal-to-Noise Ratio

OTU Optical Translator Unit
OXC Optical Cross-Connect
PLC Planar Lightwave Circuit
RMS Root Mean Square
ROADM Reconfigurable Optical Add-Drop Multiplexing
SAN Storage Area Network
SOI Silicon on Insulator
SONET Synchronous Optical NETworking
TEC Thermo Electric Cooling
VCSEL Vertical Cavity Surface Emitting Laser
VOA Variable Optical Attenuator
WSS Wavelength-Selective Switch
WSXC Wavelength-Selective Crossconnect

3.07.1 Introduction

Ubiquitous broadband communication has revolutionized all aspects of our lives in the early 21st century, affecting business and personal commerce, workplace mobility and globalization, military operations and security awareness, information distribution, learning, shopping, and entertainment, leading to unheralded empowerment of the individual. The infrastructure supporting and reliably delivering all this information within and between cities, countries, and continents is based on fiber-optic transport. The significant advantage of fiber-optic transport over alternative methods is its unique ability to carry high bit-rate information streams for substantial distances. This ability is the culmination of decades of research in the field of photonics, providing us with the elements of today's optical communication systems: diverse laser sources, high-speed optoelectronic devices for modulation and detection, low-loss optical fibers, optical amplifiers, dispersion management solutions, and means for multiplexing and demultiplexing many wavelength channels. But as optical communication systems flourished, and branched out further, and with improvements in transport distances and data rates, it became necessary and more economically viable to operate the transport infrastructure as optical networks, as opposed to static optical communication links. Optical networks are remotely managed, able to respond to changes in traffic demand and system conditions, and combine electronic and optical switches at the network nodes. The migration toward optical communication networks requires new components and subsystem elements that provide the functionality to enable the operation of these networks. These elements need to provide the adaptability, reconfigurability, and control to the network at relatively slow rates ranging from approximately a millisecond scale when service protection from failure is concerned to seconds and minutes for monitoring purposes. These requirements fit well with the capabilities of optical microsystems that rely on mechanical motion of micron-scale structures.

Optical microsystems and microelectromechanical system (MEMS) actuation are highly compatible, as the MEMS feature sizes are comparable in size to the optical wavelength. This compatibility enables interference-based devices to be constructed, and whose response will change significantly for wavelength-scale motion. MEMS brings many of the advantages of macroscopic elements, such as mirrors, shutters, translation stages, and lenses, but on a submillimeter scale. They maintain the performance of their macroscopic counterparts on account of the well-established microfabrication processes that have been developed by the electronics industry. MEMS is a batch fabrication method, driving down cost in mass production. MEMS devices can be highly complex, implementing the functional element that interacts with the optical field, its support structure, actuation mechanism, and alignment features, all with micron-scale resolution and registration. Because the integration complexity is assigned to the device fabrication process, completing the optical subsystem assembly is greatly simplified. The silicon material from which most MEMS devices are constructed can serve to guide the light, as it is transparent in the optical communication wavelength range, or reflect the light after metallization of mirror surfaces. Deployed communication equipment should be compact, lightweight, and robust, which are all attributes of well-designed optical microsystems.

This chapter provides an overview of the major applications where microsystems can impact optical communication networks. It starts with a review of optical networking that highlights the functional requirements of different network elements where microsystems may contribute. The optics of microsystems, which is based on free-space propagation of single-mode radiation, is examined afterwards.

Microsystem implementations that have been developed for five exemplary applications are subsequently reviewed. The future outlook for optical microsystems is addressed in the conclusion.

3.07.2 Emergence of Optical Networking

What is the difference between optical communications and optical networking? Optical communications has been researched and developed for many decades, and has brought the transport technology to a very mature level. Its fundamental goals are to increase the transmission capacity, the transmission range, and/or the transmission capacity-range product. This has been primarily achieved by utilizing many wavelength channels concurrently on the fiber (a technique known as DWDM or dense wavelength division multiplexing), optical amplification, dispersion compensation, and efficient modulation formats. Optical networking is what happens beyond optical communications, the effective utilization of the transport layer to deliver the connectivity, accessibility, reliability, and survivability at the network level. It is about protocols for access, routing, protection, and restoration, as well as the electronic, optical, and optoelectronic (or photonic) hardware required for implementing it. This new field even has a dedicated research journal inaugurated in 2002 (*Journal of Optical Networking*). Optical networks are customized and optimized for a specific operating environment; a continental-scale backbone network for connecting cities is very different from a neighborhood access network for individual households or from an optical data network within a supercomputing center. Backbone networks, which support the data and voice traffic infrastructure, are required to establish long-term light paths with high availability and fast restoration. Due to their geographical extent, they are the most expensive to deploy and maintain (Fishman *et al.* 2006). Access network are the most price conscientious, due to the economy of scales required for deployment. To reduce deployment and operational costs, access networks utilize passive optical networking strategies (Feldman *et al.* 1998). Computer data networks are designed for communication throughput between supercomputers as well as to mass storage, to maintain the processing cycles active. Such networks operate in accordance to changing resource allocations and demand variations, and can tolerate periodic changes to the topology (Taesombut *et al.* 2006).

Of the broad optical networking categories described above, backbone networks as well as the smaller scale metropolitan networks are most receptive to deployment of components offering adaptability and reconfigurability. There are several reasons for this acceptance. First, these networks are required to operate their communication services with defined restoration speeds stipulated under quality of service contractual obligations. An example for a stringent requirement for maintaining synchronization between two end points communication with SONET (Synchronous Optical NETworking) protocol is a 50 ms restoration of service in the case of interrupts (component failures, fiber cuts, etc.). This level of restoration requires that link and path protection schemes be employed, requiring basic switching elements with millisecond-scale response times. However, such protection schemes increase the network redundancy and cost while reducing the network capacity utilization. By using more advanced network protocols and more capable switching elements, shared path protection plans can be offered with high availability assurances, as the likelihood of multiple network interrupts occurring simultaneously is extremely low. Shared protection also increases the network's capacity utilization, achieving a better return on investment for the service provider (Ye *et al.* 2000). The second reason is related to path provisioning and reconfigurability. As bandwidth demand increases with the introduction of new electronic services, new light paths that traverse the network have to be established to support the increased load. This act of provisioning light paths may sometime include rerouting of existing light paths, to more evenly spread the load across the network links. Provisioning used to be a manual process, and required a technician's visit to every network node. The manual process was also prone to human error, and requires additional visits for verification, all leading to notoriously long provisioning times (measured in months from request to completion) and high operating expenses. By introducing advanced switching elements at the network nodes, remote provisioning from a centralized control center can be performed, reducing the provisioning time to potentially seconds. The third reason is related to the efficient utilization of the fiber transport capabilities. The advances made in transport technology have enabled deployed systems to achieve transmission distances beyond 2000 km. However, most light paths supported by the network have shorter path lengths. (Rent's rule for interconnection lengths applies; the distribution of network light path lengths is inversely proportional to the length. In other words, there

are more short light paths than there are long ones.) Hence, for efficiently using an optical network, light paths carried on specific wavelengths are required to be managed, i.e., extracted and reintroduced at different nodes of the network at the command of the centralized control center. This functionality is known as reconfigurable optical add-drop multiplexing (ROADM). More recently, the ROADM equipment has acquired the additional task of light path link switching at network nodes (Bonenfant and Jones 2003). Finally, a fourth reason for the greater acceptance of adaptable components in backbone networks is that they enable the optical network to operate more efficiently, from transport as well as maintenance points of view. The optical network topography is nonunifirm, having links of different lengths due to the locations of cities and possible mix of optical fiber types. Customized equipment can be placed on the different links to achieve a more uniform response, but inventorying many variants of a certain component is an inefficient solution. Adaptive components can be used instead of carrying many variants, with tunable dispersion compensators as an example. Stockpiling different laser transmitters for compatibility with any DWDM wavelength is another example where inventorying becomes prohibitively expensive. Tunable lasers spanning the entire communication window can be used instead. Balancing the signal levels within the network after changes such as the introduction or extraction of channels maintains the optical network performing at the correct power levels, and is important for maintaining the optical signal-to-noise ratio (OSNR). The price premium for the tunable and adaptable components over fixed components more than pays for itself in performance and operational savings.

To illustrate the form and functionality provided by optical networks, an exemplary backbone optical network as well as its interface to a metropolitan network is depicted in **Figure 1**. The backbone network is typically of national or continental scale, and serves to interconnect major sites where data needs to be introduced or extracted from the network, typically at major cities. Backbone networks are usually deployed with a mesh architecture, where each network node is attached to its nearby neighboring nodes (typically two to four links fan out per node). The links in a backbone network vary in length with the distribution of the network nodes or cities, and can vary from tens to hundreds of kilometers. In the latter case, optical amplification has to be periodically provided along the link, to maintain the signal strength. An optical in-line amplifier (ILA) is typically inserted for every 60–100 km of optical fiber; the actual distance between amplifiers is a network optimization parameter balancing performance versus cost (closely distributed amplifiers offer better optical signal-to-noise characteristic at the expense of higher amplifier count). The optical amplifier may include additional equipment to balance the optical signals as well as accompanying power taps and a monitor, for maintaining the proper signal strength and network performance. At an optical add-drop multiplexer (OADM) node, most of the DWDM channels typically continue unimpeded in the optical domain while a fraction is selected to be dropped and then added back after data are extracted and new data are introduced. The added and dropped signals are converted between the electrical and optical domains at optical translator units (OTU), which consist of optoelectronic components such as lasers, modulators, detectors, and electronic circuitry. The data in the electrical domain are handled by electronic equipment operating the protocols employed by the networks. This multiservice provisioning platform (MSPP) interoperates with the transport networks, performs grooming and tributary hand-off tasks (splitting, distributing, and combining the fractional data rate signals that make up the aggregate line signal), and enable services such as Ethernet, SAN (storage area networking), ATM (asynchronous transfer mode), GMPLS (generalized multiprotocol label switching), and private line circuits. The MSPP serves as a data gateway between different optical networks, in this case between the backbone mesh network and the metropolitan ring. The metropolitan DWDM ring distributes and collects data to and from its ring nodes, which separate the data streams to finer granularity and serve as gateways to local access networks (not shown in **Figure 1**). The basic OADM architecture is based on a power splitter, which generates two replicas of the signals, one for the drop and one for the channels that will continue through the nodes. A channel blocker prevents those channels that have been selected to be dropped from continuing to the OADM output. This architecture requires a fixed demultiplexer to separate the dropped channels, with deployed OTUs to receive the dropped channels. Such architecture is restrictive because a network reconfiguration is limited on the add and drop sides to specific channels that have OTUs on them, as there is a fixed assignment between the OTU and optical channel due to the fixed optical demultiplexer. An alternative OADM architecture is based on wavelength-selective switches (WSSs) for switching the dropped channels to individual ports.

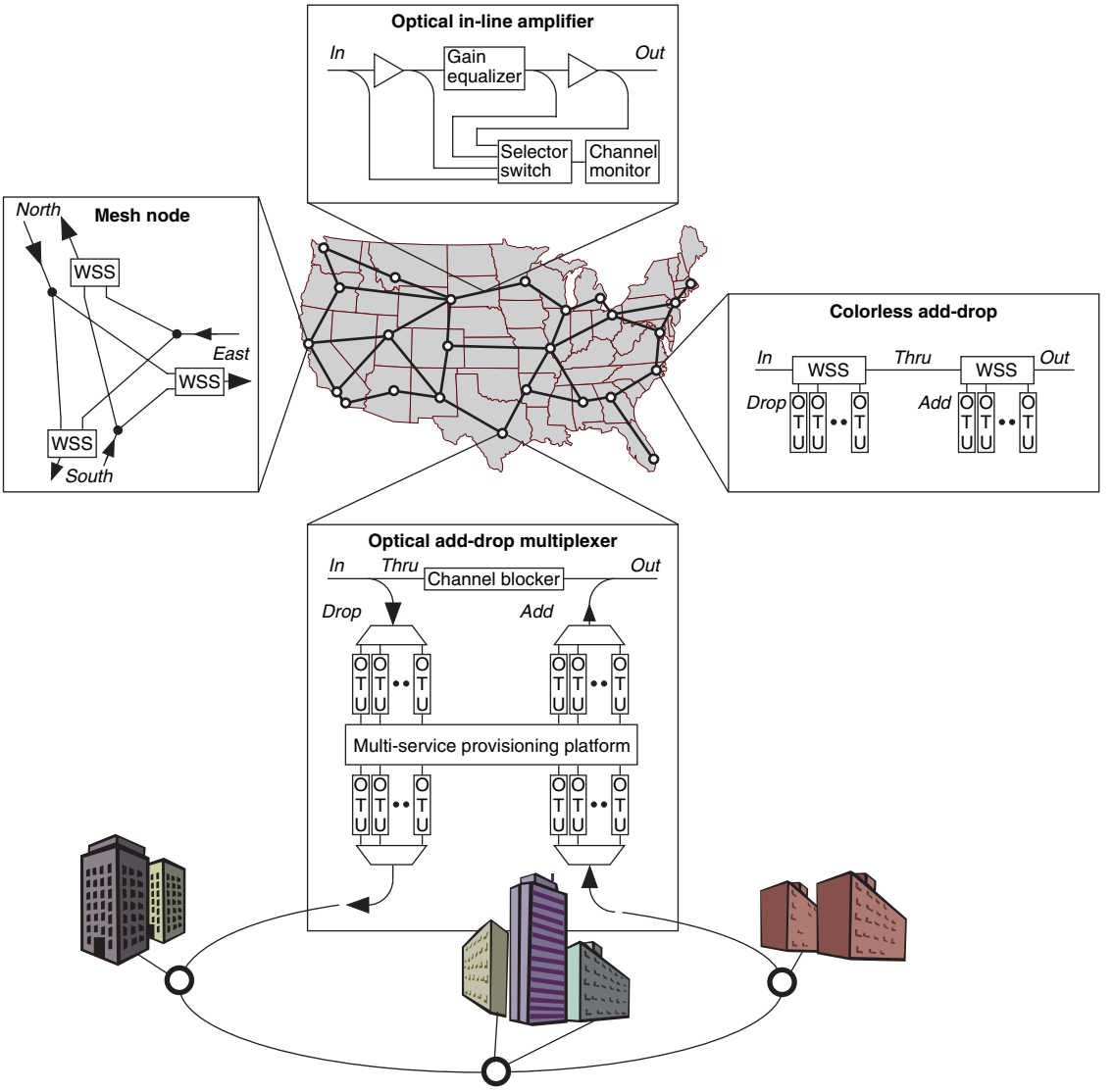


Figure 1 A typical continental-scale backbone optical network layout, demonstrating the major functions at mesh switching nodes, add-drop nodes, and its interface to a local metropolitan-scale ring.

This colorless add-drop solution is completely reconfigurable; any wavelength channel can be selected to be dropped to any fiber port. Thus, the fixed association between OTU and DWDM channel is broken, enabling greater flexibility in the network operation and necessitating that OTU support all the DWDM channel on command (the major implication being for the laser transmitter to tune to any channel; detectors have negligible wavelength dependence). The WSS port count has to match the channel drop count, with eight ports typically sufficient for most add-drop sites. The DWDM channels in the backbone mesh network traverse intermediate nodes between their ingress and egress nodes. At these intermediate nodes, the channels are path switched to the proper link that will route the channel to its intended destination. This path switching task at mesh nodes is best accomplished with WSSs, which independently route each DWDM channel to an output port that is assigned to the desired output link. Each input link at the mesh node has a WSS for splitting and combining the channels according to the routing assignment. The WSS port count has to match the node fan-out, with four ports typically sufficient for most mesh nodes.

Many telecommunication components are based on optical microsystem principles. Some of these are

adaptive, by changing certain optical parameters using MEMS or liquid-crystal (LC) devices. Others are fixed, but still benefit from microsystem construction over other implementations. Some examples of fixed optical microsystems include optical isolators (used to prevent back reflections) and circulators (for separating light propagating in the forward and backward direction), free-space optical multiplexers and demultiplexers (achieve very good thermal stability for outdoor deployment), and spectral filters using thin-film interference (for implementing fixed OADM). Examples of active optical microsystems are fiber switches (small, medium, and large varieties for implementation of various network functions, according to size), variable optical attenuators (required to maintain acceptable power levels impinging on photodetectors), tunable filters (for channel selection in broadcast networks and for optical monitoring), WSSs (for realization of ROADMs), tunable dispersion compensators, and even beam steering mirrors for free-space optical links). The active optical microsystems listed above (and some described in depth in the bulk of the chapter) all exhibit operating speeds in the millisecond timescale. Even though this characteristic speed is much slower than the data bit rate of SONET or packet rate of IP (Internet protocol), it is more than suitable for tracking network changes due to ambient conditions (variations of the network infrastructure due mostly to temperature) and for supporting light path provisioning as well as protection and restoration functions. The component marketplace is flush with competing products and technologies from many companies (~1000 companies present at the largest technical conference and trade show, *Optical Fiber Communications*), all vying for a fraction of the global component sales.

3.07.3 The Optics of Telecommunication Microsystems

The optical communications field matured considerably since the days of the first deployment in 1978, in a sequence of several generations. First-generation systems were based on 780 nm GaAs laser transmitters using multimode fibers. The systems were limited to low bit rates due to modal dispersion. Second-generation systems addressed the bit-rate limitations by using single-mode fibers and shifting the operating wavelength to 1300 nm. Modal dispersion is nonexistent when using single-mode fibers, and at 1300 nm chromatic dispersion is minimized in glass (silica) waveguide. However, lacking optical amplifier technology at the time, the communication range became signal-level limited, due to propagation losses at 1300 nm (losses attributed to Rayleigh scattering, which decrease for longer wavelengths). Third-generation systems shifted the operating wavelength to 1550 nm, where the optical fiber exhibits its minimal losses of 0.2 dB km^{-1} (attenuation increases for longer wavelengths due to silica molecular vibration absorption). Many additional advancements have been introduced to create the fourth-generation systems, some of which must be mentioned for completeness: optical amplification allowing for compensation of the fiber's attenuation, DWDM enabling multiple channels to copropagate in the same fiber, dispersion management for reducing chromatic dispersion impairments, advanced optical modulation formats for achieving higher spectral efficiencies, and nonlinear propagation. However, fiber transport largely remains around 1500-nm wavelength over single-mode fiber, which will be the focus of this section (with the exception being in short-haul links and local area networks, where range is less of an issue).

Realizing optical networks requires integrating many additional components to the transport layer. These components must exhibit low losses, as losses must be compensated for with amplification, which reduces the OSNR. Hence, the design of optical components has to be extremely well engineered to minimize these losses. For components based on interaction with unguided light, or light propagating in free space, which is the case for most optical microsystems described in this chapter, this is especially challenging as the light must radiate out of the input fiber, be manipulated by the microsystem, and then be coupled back into the output fiber. Depending on functionality, the optical microsystem enables the interaction of the MEMS element with the delivered light, in the form of refraction, reflection, diffraction, or absorption. Loss sources can creep up quickly and they are usually cumulative; hence, the microsystem's optical design has to be optimized for high coupling efficiency as well as tolerant of natural component variations, misalignments, and deployment conditions.

3.07.3.1 Gaussian Beam Optics in Free Space

The light propagating within a step-index single-mode fiber is confined by the higher core refractive index. The light has a transverse profile given by

radial Bessel functions in the core and cladding regions. However, for optical microsystem design purposes, the precise transverse profile is approximated by a Gaussian one, with an excellent fit at a beam diameter of 10.4 μm at an operating wavelength of 1.55 μm. The advantage of using the Gaussian profile approximation is that it is a solution to the paraxial Helmholtz equation, which describes light propagation in a homogeneous region (Saleh and Teich 1991). Hence, a Gaussian profile shall remain Gaussian as long as it does not encounter nonhomogeneous regions, such as an aperture.

The Gaussian beam field, $U(x, y, z)$, is described in a Cartesian coordinate system with the propagation direction along the z-axis by

$$U(x,y,z) = U_0 \frac{w_0}{w(z)} \exp\left(-\frac{x^2+y^2}{w^2(z)} - \mathrm{j}\frac{2\pi n}{\lambda} \times \left(z + \frac{x^2+y^2}{2R(z)} \right) + \mathrm{j}\phi(z) \right)$$

where $R(z)$ is the field radius of curvature varying as

$$R(z) = z\left(1 + \left(\frac{z_0}{z}\right)^2\right)$$

$w(z)$ is the beam radius defined by

$$w(z) = w_0\sqrt{1 + \left(\frac{z}{z_0}\right)^2}$$

and $\phi(z) = \tan^{-1}(z/z_0)$ is a phase delay along the propagation direction giving rise to a Guoy phase shift associated with a π phase jump when propagating past the beam waist (**Figure 2**). The beam waist is defined at $z = 0$, and is the location where the beam is at its narrowest spot, a location associated with a minimal beam radius w_0 and an infinite radius of curvature (flat phase front as would appear on a plane wave; however, the field is apodized by a Gaussian envelope). The distance z_0 is known as the Rayleigh length, defined by

$$z_0 = \pi \frac{w_0^2 n}{\lambda}$$

where n is the refractive index of the medium. The Rayleigh length measures the distance from the waist to the location where the beam grows by a factor of $\sqrt{2}$. The Gaussian beam size is symmetric about the beam waist, with the radius of curvature defining whether the beam is converging toward the waist (negative radius of curvature) or diverging from the waist (positive curvature). The fiber mode is approximated by the Gaussian beam with the waist located at the fiber tip. Hence, light radiating out of the fiber diverges, with a Rayleigh length of 55 μm (for air with $n = 1$), and the same parameters have to be satisfied for efficient coupling into a fiber.

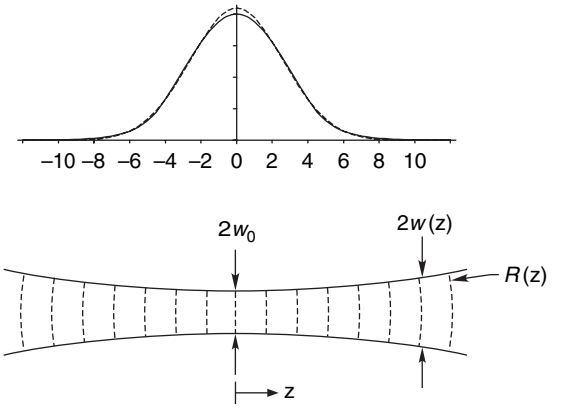


Figure 2 *Top*: The mode shape of the standard single-mode optical fiber (solid line), and its Gaussian approximation of width $2w_0 = 10.4$ μm at an operating wavelength of 1.55 μm (dashed line). *Bottom*: A Gaussian beam propagating in free-space, dashed lines show phase fronts. The beam has a minimal waist at location $z = 0$. A Gaussian beam converges toward the waist and diverges afterward. The light radiating out of a cleaved optical fiber is well approximated by a Gaussian beam.

The Gaussian beam can be manipulated in free space by mirrors, lenses, and prisms. Flat mirrors do not affect the Gaussian beam profile evolution, but do change the propagation direction. Curved mirrors have a lensing effect, with the effective focal length equal to half the mirror radius. Lenses, when treated as thin lenses, change the Gaussian beam radius of curvature by a factor of $F/(F–R)$, where F is the lens focal length and R is the Gaussian beam radius of curvature impinging on the lens. Simple beam relationships can be established with the aid of Gaussian beam propagation formulas and the lens radius transformations for cases of interest such as the maximum throw distance of a lens (i.e., for a given input Gaussian beam and lens focal length, how far and large can a Gaussian beam waist be projected from the lens). Prisms refract light in accordance to Snell's law, and can be used to expand or contract the beam in one direction, converting the circular Gaussian beam to an elliptical profile. Note that the beam propagation calculations can be in the forward or backward directions, and are valid as long as the beam does not encounter an aperture.

3.07.3.2 Misalignment Losses

A Gaussian beam can be propagated within an apertureless optical microsystem in a lossless fashion, as the

beam propagation does not take into account Fresnel reflections at dielectric interfaces (which can be minimized to ~0.1 dB per interface with antireflection coating), and finite metallic reflectivity (~0.25 dB per reflection). The forward propagation eventually leads to the output fiber end face, where the beam has to couple back into a single-mode fiber. The coupling can be evaluated with the complex ovelap integral between the incident beam wavefront $\varphi(x,y)$ and the fiber mode profile $\phi(x,y)$, which is approximated by the Gaussian distribution, and given by

$$\eta = \frac{\int_{-\infty}^{\infty} \varphi(x,y)\phi^*(x,y)\mathrm{d}x\mathrm{d}y}{\sqrt{\int_{\infty}^{-\infty} \varphi(x,y)\varphi^*(x,y)\mathrm{d}x\mathrm{d}y}\sqrt{\int_{\infty}^{-\infty} \phi(x,y)\phi^*(x,y)\mathrm{d}x\mathrm{d}y}}$$

The calculated coupling η can yield complex results, leading to spectral phase variations or chromatic distortions. The power coupling, $|\eta|^2$, is typically the quantity of interest and is often expressed in decibel scale $\left(10\cdot\log_{10}\left[|\eta|^2\right]\right)$. When using normalized field distributions, the denominator of the coupling integral is unity. An incident wavefront that exactly matches the fiber mode profile will achieve a power coupling efficiency of one. The coupling integral can be evaluated numerically for any field distribution, and analytically for the Gaussian beam case. It is instructive to inspect the effects of lateral misalignment, angular misalignment, mode mismatch, defocus, and combinations thereof (**Figure 3**). The power coupling in the presence of lateral and angular misalignments is

$$|\eta|^2 = \exp\left(-\frac{x_0^2+y_0^2}{w_0^2} - \frac{w_0^2\pi^2\left(\theta_x^2+\theta_y^2\right)}{\lambda^2}\right)$$

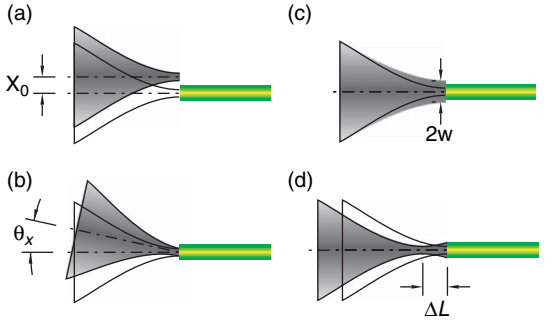


Figure 3 Mechanisms for coupling losses into single-mode fiber. (a) Lateral misalignment between incident Gaussian beam and fiber endface. (b) Angular misalignment. (c) Mode-size mismatch. (d) Waist location shift or defocus.

and has Guassian distributions in the lateral misalignment, x_0 and y_0, and angular misalignments, θ_x and θ_y. For single-mode fiber, a total lateral misalignment of 4.4 μm or a total angular misalignment of 4.5° will give rise to a 3-dB penalty (half the power is lost). Because the lateral and angular misalignment dependencies are Gaussian, or quadratic in decibel scale, a factor of $1/x$ improvement in power coupling requires a factor of $1/\sqrt{x}$ tightening in the misalignment. For example, reducing fiber misalignment losses from 3-dB to 1-dB penalty (factor 1/3 improvement) shall require a 2.5 μm lateral misalignment (factor of $1/\sqrt{3}$ tightening). Mode mismatch (different waist size incident on fiber) and defocus (longitudinal displacement of waist location) do not follow a Guassian relationship. The mode mismatch coupling loss, given by

$$|\eta|^2 = \frac{4w_0^2w_1^2}{\left(w_0^2+w_1^2\right)^2}$$

is small for waist variations as large as 20%. The defocus coupling loss is given by

$$|\eta|^2 = \frac{1}{1+\left(\frac{\Delta L}{2z_0}\right)^2}$$

where ΔL is the defocus distance and will provide 80% efficiency (or a loss of about 1 dB) at the Rayleigh distance (55 μm for single-mode fiber).

The coupling integral was evaluated above at the output fiber end face. However, as the Gaussian beam can be propagated forward as well as backward, the overlap integral can be evaluated anywhere within the optical microsystem. This propagation to a different location within the microsystem is instructive for cases involving a single aperture in the beam path.

3.07.3.3 Beam Clipping Losses

After a Gaussian beam is clipped by an aperture, the field distribution is no longer Gaussian and the beam propagation formulas provided above are no longer valid. Because all real elements in the beam path have finite extent, we shall define a criterion for a tolerable aperture size for a given Gaussian beam size later in this section. For the instructive case of a single clipping element in the beam path, we can use the forward and backward Gaussian beam propagation technique to meet at the aperture. Consider the configuration shown in **Figure 4**, where the input fiber waist is projected by a collimator lens of infinite

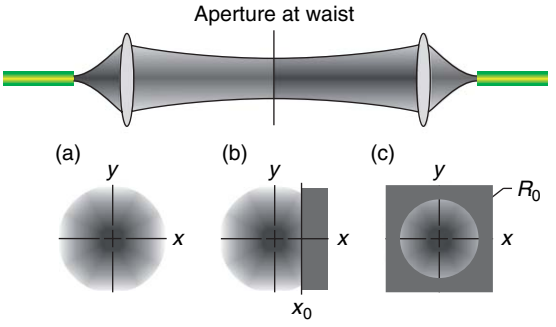


Figure 4 *Top*: Imaging arrangement between two fiber endfaces, having a wide collimated beam region in between where optical devices may be inserted. Effect of single aperture in the beam path can be analyzed: (a) circular Gaussian beam profile at aperture location; (b) shutter aperture blocking parts of beam; and (c) circular aperture.

extent to create a secondary waist of size w_1, and a similar collection lens couples the light into the output fiber. The configuration can be properly designed with ideal lenses to deliver a perfect mode to the output fiber with no coupling losses. If an edge shutter is inserted into the beam path at the Gaussian beam waist location, the beam will be clipped by the edge of shutter. The coupling integral can easily be evaluated at the location of the shutter using the ideal Gaussian waist of w_1; the impact of the shutter is introduced by adjusting the integration area. If the edge of the shutter is at position x_0, then beam regions for $x > x_0$ are clipped and the power coupling is given by

$$|\eta|^2 = \frac{1}{4}\left(1 + \mathrm{erf}\left(\frac{\sqrt{2}x_0}{w_1}\right)\right)^2$$

The power coupling approaches 1 for $x_0 \gg w_1$ (shutter far from interacting with the beam), and approaches zero for $x_0 \ll -w_1$ (shutter blocks most of the beam). It is interesting to note that when the shutter is blocking half the beam (edge is at $x_0 = 0$) the power coupling is 1/4 (or −6 dB). This somewhat surprising results occur because half the beam is clipped (1/2 power lost) and the modal overlap contributes another factor of 1/2. Clipping effects from sharp edges often occur in variable optical attenuators, blockers, and WSSs (see Sections 3.07.4.3 and 3.07.4.5).

Lenses and other optical elements are oftentimes circular, and will clip the extent of the circularly Gaussian mode. The analysis of a single beam clipped by a circular aperture can be performed in the same fashion as for the shutter edge. Adopting the configuration of **Figure 4**, a circular aperture of radius R_0 is placed at the location of the Gaussian waist of radius w_1. The aperture and beam centers are aligned, and beam regions for $r > R_0$ are clipped, defining a power coupling efficiency of

$$|\eta|^2 = \left(1 - \exp\left(\frac{2R_0^2}{w_1^2}\right)\right)^2$$

The power coupling approaches 1 for $R_0 \gg w_1$ (aperture far from interacting with the beam), and approaches zero $R_0 \to 0$ (aperture blocks most of the beam). When $R_0 = 1.5w_0$, the fiber coupling loss is nearly lossless at −0.1 dB.

When multiple apertures are present in the beam path, the analysis becomes much more complex. The Gaussian beam from the input fiber can be propagated until the first aperture. Likewise, the Gaussian mode from the output fiber can be back propagated to the last apertures. Gaussian beam propagation cannot be used for the region between the first and large apertures, and Fresnel propagation must be used instead. Because Fresnel propagation in the presence of apertures is difficult to compute, a design rule has been established stating that if the aperture is 50% larger than the Gausian beam, then clipping effects can be ignored and Gaussian beam propagation can be still be used. When the condition is satisfied $(R_0 > 1.5w)$ the light throughput of the aperture is 99%. Note that the stated condition does not consider any beam to aperture misalignment tolerance, so a slightly more stringent condition might be adopted in practice to avoid clipping effects.

The characterization of optical components intended for integration into optical communication systems is not complete without procedural tests to ascertain the compliance to generic requirements (GR) put forth by Telcordia, the advising body to the regional service providers in the United States (www.telcordia.com). The two most often cited documents are GR-1209, GR for passive optical elements, which specifies the operating conditions that components must achieve, and GR-1221 that specifies the reliability testing procedures. The listed tests include temperature cycling and temperature and humidity cycling, high-temperature storage, high-temperature and humidity storage, cold temperature storage, vibrational tests, and shock tests. Some of the tests are performed while the device is in operation, others mimic extreme conditions the component may experience in storage conditions. The components have to pass the tests to gain Telcordia compliance certification.

3.07.4 Exemplary Optical Networking Microsystems

Of the many possible networking applications for which optical microsystems have been demonstrated, at both the research or development levels, five concerted efforts are reviewed here. For each highlighted application, several microsystem implementations are reviewed, demonstrating the diversity of creativity in the face of a challenging task.

3.07.4.1 MEMS Tunable Lasers

Optical communications have been utilizing semiconductor lasers almost exclusively, from early systems lasing at 850 nm and based on GaAs, to modern systems based on InGaAsP and operating at either the 1300- or 1550-nm communication bands. The main advantages of semiconductors lasers for communications are their compactness, high power conversion efficiency, wafer-scale processing, and the capability of direct modulation of the electrical pump current, even though the latter has been abandoned as the communication bit rate has reached $10\,\mathrm{Gb\,s^{-1}}$ due to chirp effects (leading to bandwidth broadening). The basic structure of a laser consists of an optical cavity with a gain section introduced into it. The semiconductor section provides a relatively broad and continuous gain spectrum, whereas the cavity provides finite, equally spaced modes defined by the phase continuity of the cavity round trip travel for specific wavelengths. The basic Fabry–Perot cavity laser, formed from a cleaved section of a semiconductor chip with a waveguiding structure defining the transverse confinement, lases at the multiple longitudinal cavity modes under the gain spectrum. However, lasers for high bit-rate optical communications are required to lase on a single cavity mode only, to minimize chromatic dispersion-induced distortions upon propagation. To limit the laser cavity modes, a filtering component is introduced into the cavity that selects a particular cavity mode while attenuating all others. The cavity filter can be incorporated into the fabrication of the semiconductor chip, by slightly etching or longitudinally modulating the waveguide structure at the cavity ends. This periodic modulation forms a reflective filter for a wavelength component satisfying the Bragg criterion, limiting the lasing to a single frequency in each distributed feedback (DFB) laser.

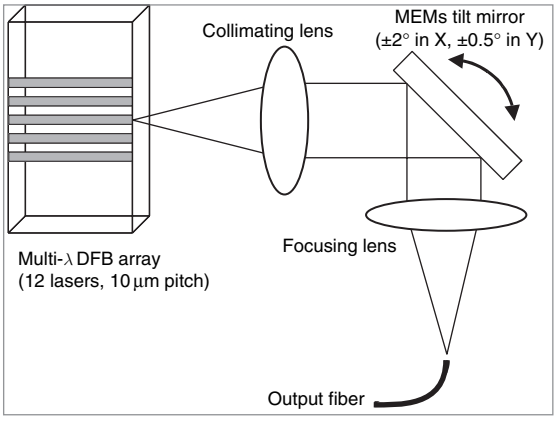
DWDM transmission necessitates discrete individual lasers to match the channel numbers in operation by the optical communication system. With DWDM channel counts of few tens to one hundred or more, having unique single-frequency lasers poses a manufacturing, inventory, and management cost to the component manufacturers, system vendors, and network operators. Tunable semiconductor lasers can eliminate the inventory cost and manufacturing variations by having a single laser serve any wavelength used in the communications band on command. The overall cost savings that can be afforded by using tunable lasers should more than offset the price premium for such lasers over their single-frequency counterparts. Optical networking introduces additional functional motivations for deploying full band tunable lasers. New light paths in the network can be established by provisioning an available wavelength channel through intermediary network nodes. Networks can respond to changes in demand patterns due to daily, weekly, or seasonal variations, releasing resources (wavelengths) from one end and reassigning them to others. These scenarios require that the laser transmitter be capable of tuning to any prescribed DWDM channel (Coldren *et al.* 2004).

The earliest tunable lasers utilized thermal tuning of the DFB laser. The DFB laser already incorporates a thermal feedback loop for stabilizing its center wavelength, so providing tunablility by tuning the temperature setting came practically free. However, the possible thermal tuning range of DFB lasers is a few nanometers only, far short of the full communication bandwidth. Full band tunable lasers require greater complexity, and the research community has demonstrated several innovative solutions. The principle behind all the tuning techniques is based on an intracavity filter that can effectively span the entire communication band and select a single cavity mode for lasing. A differentiator among the tuning technique is the tuning speed; physical property variations such as thermal or mechanical requires a few milliseconds or longer, whereas electronic techniques can switch faster than microsecond speed. For the foreseeable future, the relatively slow tuning speed should suffice for applications in optical networking. The faster switching speeds are reserved for optical packet switching applications, which are currently under the exploratory research realm. Slow tunable lasers pose fertile ground for MEMS ingenuity, as exemplified in the following.

3.07.4.1.1 DFB array selector

One of the main advantages of single-frequency DFB lasers is that the entire laser is completely integrated on a semiconductor chip, guaranteeing that the laser

cavity is very stable and robust. The laser package is completed after fiber attachment, typically via collimation optics that converts the elliptical laser output mode to the circular fiber mode for efficient coupling and also incorporates an optical isolator to prevent back reflections into the laser. One possible tuning approach is to capitalize on this mature laser technology, and to provide wide-band tuning by fabricating an array of DFB lasers in close proximity, each with an offset center wavelength that can be thermally tuned over a few nanometers. A MEMS switching mirror is placed within the free-space collimation optics and serves to select which waveguide of the waveguide array is coupled to the output fiber (**Figure 5**). Any wavelength can be set by electrically activating the single laser waveguide that can tune to the desired waveguide and thermally tuning the entire chip to the predetermined temperature (Pezeshki *et al.* 2002, Ton *et al.* 2004, Zou *et al.* 2004). The lasing characteristics are similar to those of conventional DFB lasers.



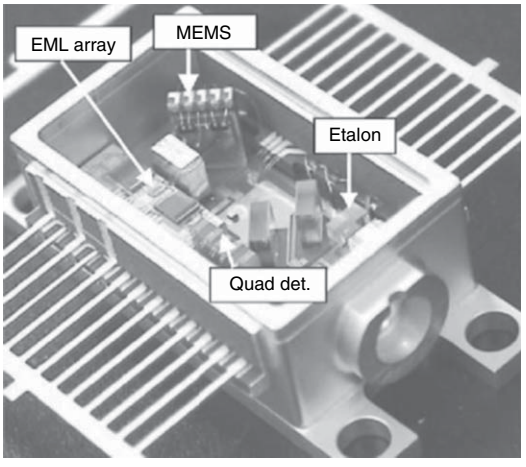


Figure 5 Tunable laser using an array of thermally tunable distributed feedback (DFB) lasers and a microelectromechanical systems (MEMS) selector switch. *Top*: Schematic of laser structure. *Bottom*: Packaged laser in standard butterfly package. (Source: Pezeshki B *et al.* 2002 20-mW Widely tunable laser module using DFB array and MEMS selection. *IEEE Photon. Technol. Lett.* **14**, 1457–9, © [2002] IEEE.)

The MEMS micromirror angle is tilted to the angle that couples the selected output laser from the DFB array to the fiber output and associated optics and isolator. The laser array consists of 12 lasers on a 10-μm pitch, each with a center wavelength shift of 2.8 nm by varying the longitudinal grating pitch. The chip size is similar to that of a fixed wavelength DFB laser, and contains no additional processing steps. To prevent erroneous wavelengths from being emitted from the laser to the optical communication system during the tuning process and thermal settling time, the MEMS mirror is intentionally misaligned from the output fiber during the tuning operation. When tuning is complete, the mirror is adjusted to optimally couple the laser output to the fiber. A 1.5 mm diameter mirror consisting of a gimbaled structure with torsional springs, is fabricated using bulk silicon micromachining, and reaches its maximum deflection angle at 120 V. The MEMS steering mirror can tilt ±2° in the waveguide selection direction, and ±5° in the orthogonal axis. The large mirror tilt range in two axes allows for loose placement of the optical components within the laser package, and using the MEMS mirror for correcting for the placement errors. A feedback loop based on 1% power dither at 50 Hz applied to the mirror maintains the optical alignment and compensates for any possible mechanical drift or creep. Shock and vibration do not influence the wavelength stability of the laser, as the cavity is self-contained within the chip, and only results in coupled output power variations.

The laser chip, MEMS mirror, and all other collimation optics and control electronics are packaged in a standard butterfly package, as well as a wavelength locker to provide a feedback signal for the chip heater. The laser module provides 20 mW fiber coupled power, tunable over the entire C-band (greater than 30 nm tuning). A typical ~1 s laser tuning speed is limited by the thermoelectric cooler (TEC) cooling capacity.

3.07.4.1.2 External cavity with grating filter

The DFB laser's small refractive index change limits their useful thermal tuning range to a few

nanometers. Different filtering techniques must be introduced into the laser cavity to extend the tuning range to cover the entire communication band. When the wide-tuning filtering mechanism cannot be implemented directly on the semiconductor gain chip due to an imcompatibility with the lithographically predefined and rigid cavity structure, external cavity laser architecture can be utilized. In an external cavity laser, the cavity consists of a semiconductor gain waveguide, typically with a cleaved facet mirror on one end and an antireflection coating on the second end, followed by a free-space propagation section with collimation optics to an end mirror defining the cavity. Within the free-space section of the external cavity laser, other tunable filtering elements can be introduced. A tunable external cavity laser can be based on a Littman–Metcalf geometry (Littman and Metcalf 1978, Liu and Littman 1981), consisting of a diffraction grating illuminated at near-grazing incidence for angularly dispersing light and a mirror on a rotation block for retroreflecting finite spectral components (**Figure 6**). Continuous tunability is achieved by rotating the mirror and selecting the wavelength that satisfies retroreflection. However, as the cavity supports equally spaced discrete modes, the cavity length typically needs to be minutely perturbed to shift the cavity mode to coincide with the desired lasing wavelength. Moreover, the filter passband has to provide sufficient loss to the adjacent cavity modes to suppress their lasing. Because the cavity mode spacing is inversely proportional to the cavity length and the external cavity

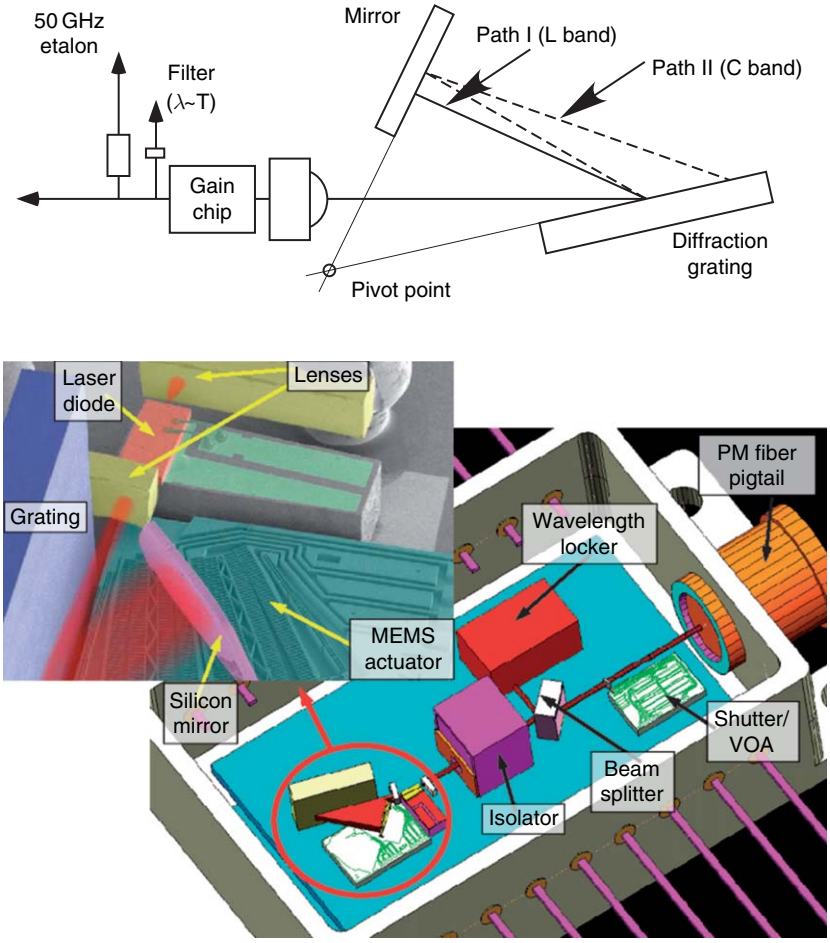


Figure 6 Tunable laser with an intracavity grating spectral filter. *Top*: Schematic of laser cavity with rotating end mirror for wavelength selection. *Bottom*: Schematic of packaged laser. *Inset*: Cavity and mirror mounted on microelectromechanical systems (MEMS) actuator. (Source: Anthon D, Berger J D Tselikov A 2004a C+L band MEMS tunable external cavity semiconductor laser. *Optical Fiber Conference (OFC)*, Los Angeles, CA, USA, paper WL2, © [2004] IEEE.)

laser has significant length due to the space necessary for collimation optics and the tunable components, the filter must have narrowband characteristics.

The retroreflection mirror can be placed on a MEMS rotation block fabricated from 85-μm-thick single-crystalline silicon using deep reactive ion etching (DRIE) (Anthon *et al.* 2004a, b, Berger *et al.* 2001). Multiple comb drives electrostatically rotate the actuator about a virtual pivot point that is at the intersection of the grating and mirror planes. The pivot point placement is critical for mode-hop free tuning per the Littman–Metcalf design, where the mirror rotation is accompanied by displacement for maintaining the overlap of the cavity mode and the filter center wavelength. In practice, and especially for full communication band tuning range, additional controls need to be added for controlling the cavity mode locations. Cavity length control is provided by a piezoelectric actuator that translates the diffraction grating. To reduce the mirror's angular rotation range requirement, which is directly related to the filter's tuning range, a double reflection–diffraction arrangement is used (the light diffracts from the grating toward the mirror, reflects back to the grating, diffracts a second time from the grating and then in incident on the mirror at retroreflection for the center wavelength). The double reflection–diffraction arrangement also increases the cavity length, which decreases the cavity mode spacing, but the double diffraction generates a narrower filter passband that more than compensates for the tighter mode spacing. The MEMS actuator rotates the mirror $\pm 2.8^\circ$ for 140 V driving voltages, providing at ± 2.5 THz laser tuning range, easily covering the communication band. Because both the cavity and the tuning element are defined by the large rotating end mirror, vibrational sensitivity of the cavity is a major concern with this design.

The laser can be tuned continuously and mode-hop free with simultaneous control of the mirror rotation angle and grating position. However, in an optical networking environment, the laser cannot emit light during tuning to prevent interference with other signals. This can be addressed by adding a MEMS optical shutter between the laser cavity and the optical fiber to blank the signal during tuning. Additionally, a wavelength locker is used to provide a feedback signal necessary for wavelength tuning the laser with the MEMS and piezoelectric actuators. Laser tuning is complete in less than 15 ms, and the laser's wavelength locker ensures stable long-term operation.

3.07.4.1.3 Vernier thermal etalons

A different approach to tuning an external cavity laser is demonstrated by utilizing dual silicon etalons within the cavity (Finot *et al.* 2004, McDonald *et al.* 2006). The standard external cavity layout is formed with a semiconductor gain chip (one facet used for the cavity end mirror and the other antireflection coated) and a second mirror in the free-space region. The cavity mode selection is accomplished by the cascaded filtering response of the two etalons that are inserted within the free-space region (**Figure 7**). Each etalon, fabricated from a polished silicon wafer of unique thickness, has a transmission function consisting of a frequency comb. The comb spacing is determined by the etalon thickness, and as the silicon etalon is much shorter than the laser cavity length, its modes are spaced much wider apart. The etalon comb mode widths and position are determined by the etalon reflectivity, or finesse (see Section 3.07.4.2 for definition), and etalon phase – the desired reflectivity achieved by proper coating of the silicon and the etalon phase controlled by thermally tuning the silicon's refractive index. Because the two etalons are slightly different in width, their cavity mode spacing is also slightly different. The etalons are engineered such that their spacing difference and finesse ensure that only a single etalon mode overlaps over the laser gain spectrum. This Vernier effect tuning determines the desired lasing wavelength, and allows the dual etalons to cover a wide bandwidth with only a small cavity mode shift. Finally, for the laser to operate at this overlap peak, it must also coincide with one of the laser cavity's modes. The laser cavity's modes are tuned by varying the cavity phase with a lithium niobate crystal.

The silicon etalons are fabricated using wafer-level MEMS processing, even though the etalons have no moving parts. A polished silicon-on-insulator (SOI) wafer is used to define the silicon disks that form the etalons. The silicon handle on the backside is removed and the disk is suspended with a SiN membrane in tension. The etalon is thus thermally isolated from the handle, reducing the heating load required to stabilize its temperature. The fundamental resonance frequency of the silicon and membrane is above 10 kHz. The etalons have integrated heaters and temperature sensors formed by two Ti/Pt conductors tracing the etalon's optical aperture circumference. Heating is achieved by the power dissipation from the heating conductor, and the temperature is sensed by circuitry that registers the resistance of the second conductor, which depends on the etalon temperature. A 30°C temperature rise is achieved with 15 mW power

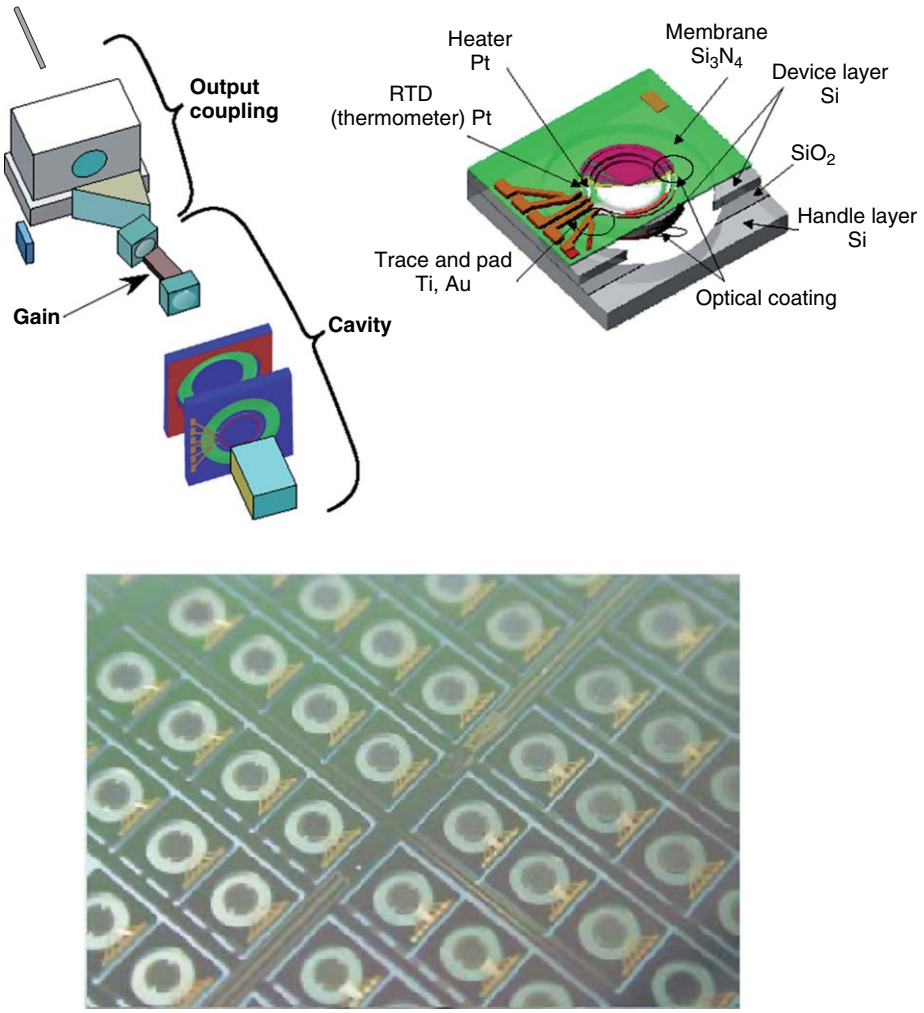


Figure 7 Tunable laser with intracavity thermally tuned etalon filters. *Top-left*: Schematic of laser cavity with two filters. *Top-right*: Details of thermally tuned silicon microfabricated etalon filter. *Bottom*: Wafer-level batch fabrication of filters. (Source: Finot M *et al.* 2004 Automated optical packaging technology for 10 Gb/s transceivers and its application to a low-cost full C-band tunable transmitter. *Intel Technol. J.* **8**; McDonald M *et al.* 2006 Wavelength filter with integrated thermal control used as an intracavity DWDM laser tuning element. *IEEE Optical MEMS Conference*, Piscataway, NJ, USA, © [2006] IEEE.)

dissipation. The thermal time constant, determined by the resistance and thermal mass of the etalon, is under 1 s. The etalon's finesse is achieved with a four-layer reflection coating.

The stability and repeatability of the etalon's response allows the etalon to serve as the wavelength reference for the laser. Laser tuning is performed by heating the two etalons to the prestored values and then adjusting the cavity phase to maximize the lasing power. The cavity phase is determined by a peak-locating algorithm with electro-optical dithering of the lithium niobate modulator. Because the inter cavity etalons serve as the wavelength reference, no wavelength locker is necessary after the laser. Moreover, the etalon wavelength tuning can be performed without lasing, so the laser cavity can be tuned in the dark state and turned on with the pump current when the temperature has settled, eliminating the need for an additional shutter.

3.07.4.1.4 Tunable cavity VCSEL

If placing a filter in an external cavity is challenging, why not do away with it entirely? VCSEL (vertical cavity surface emitting laser), in contrast to the aforementioned edge emitting lasers, have a vertical optical cavity that enables different laser designs,

fabrication sequences, and beam characteristics. VCSELs are fabricated in a single epitaxial growth sequence, starting from a bottom distributed Bragg reflector (DBR), a semiconductor gain section, followed by a top DBR mirror, resulting in a short pillar laser geometry. The DBR consists of many alternating layers of lattice-matched materials; the large number of layers is required to achieve high reflectivity from materials with similar refractive indices. The high reflectivity is crucial for reducing the cavity losses, as the small gain available from the short gain section must be larger than the losses to support lasing. The short cavity length implies that the cavity modes are spaced widely apart, and typically only a single cavity mode exists within the gain bandwidth. Hence, no mode-selection filter is required in a VCSEL. Other advantages VCSEL offer are wafer-level inspection, as the laser emits out of plane, circular beam modes compatible with optical fiber mode, and 2D array geometries important for some applications. Conversely, their output powers are much lower than those of the DFB laser, and heat removal is challenging due to their pillar structure leading to high thermal resistance.

Tunable VCSEL can be realized by thermally changing the refractive index or by changing the cavity length. Thermal tuning is compounded by the laser's inherent thermal resistance, and cavity length changes are incompatible with the monolithic pillar structure. By introducing an air gap into the cavity and integrating a MEMS actuator with the top mirror, the cavity length and lasing wavelength can be directly tuned (**Figure 8**). Many variations on this theme have been reported in the scientific literature: using a flat mirror

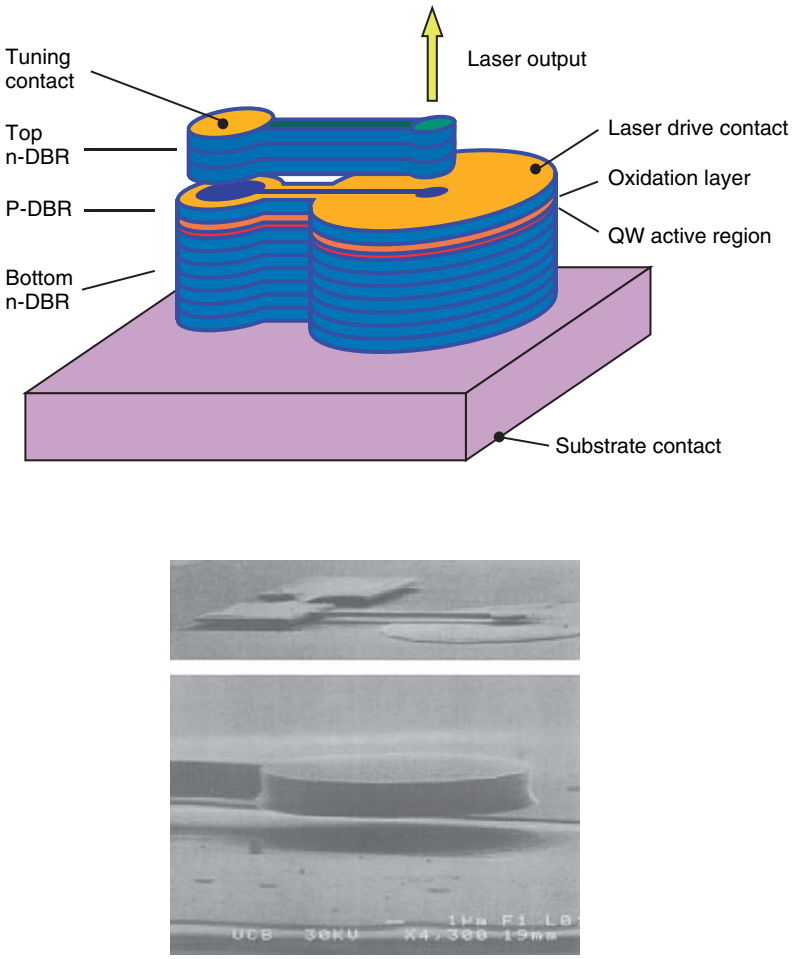


Figure 8 Tunable vertical cavity surface-emitting laser (VCSEL) with cavity length modulation. *Top*: Schematic of piston-motion VCSEL. *Bottom*: Top mirror mounted on piston motion actuator. (Source: Chang-Hasnain C J 2000 Tunable VCSEL. *IEEE J. Select. Top. Quant. Electron.* **6**, 978–87, © [2000] IEEE.)

versus a curved mirror, suspending the movable mirror with a continuous membrane versus a symmetric beam support structure or a cantilevered beam, electrostatic versus thermal actuation, one- and two-chip laser assemblies, and electrical or optical pumping of the laser. Each variation has its unique traits, but all share the principle that the cavity tunes by adjusting the air gap regions (Chang-Hasnain 2000, Kner *et al.* 2003, Li *et al.* 1998, Sugihwo *et al.* 1998, Tarraf *et al.* 2004, Tayebati *et al.* 1998).

As an exemplary VCSEL design, an electrically pumped, single-chip laser depicted in **Figure 8** is described. The entire laser is grown in one epitaxy run. The DBR mirrors are fabricated from a GaAs–AlGaAs heterostructure, and an InGaAs active region. An oxidation layer serves as a current guide, for effective electrical pumping of the gain region. The top mirror is suspended with a cantilever beam from a large, parallel plate electrostatic actuator. Because the actuator is physically decoupled from the laser section, its design can be optimized for operation at low voltages (<30 V), with a reported switching time of 1–10 ms. The mirror can displace up to a 1/3 of the initial gap size, with the lasing wavelength shifted toward shorter wavelengths as the gap size decreases upon actuation. The laser utilizes a wavelength locker for stabilizing the mirror position and emission onto the DWDM grid. The VCSEL can be directly modulated via the injection current at 2.5 Gb s^{-1} and the signal carried with a small impairment for 100 km on single-mode fiber, making it suitable for metropolitan-scale networks. However, the laser had a reported limited tuning range of ~10 nm in the 1.55 μm communication window and DC power levels approaching 1 mW. It should be noted that optically pumped VCSEL can exhibit a tuning range spanning the entire communication window and higher output powers, as less heat is generated at the active region (heat removal associated with electrical pumping remains an issue with VCSELs).

3.07.4.2 MEMS Tunable Filters

Fourth-generation DWDM optical communication systems achieve increased capacity through efficient utilization of the optical fiber bandwidth by multiplexing many wavelength channels. Optical networking uses the foundation of DWDM optical transport to interconnect geographically dispersed end users. The role of the optical network is to provide communication conduits at high or guaranteed availability between any two end points wanting to exchange information. An early optical network prototype based on DWDM used a broadcast-and-select topology (Dono *et al.* 1990). Each network user is distinguished by a unique laser wavelength, and all users transmit their signals simultaneously into a common fiber distribution network. An end user can select a communication partner by tuning his receiver to the wavelength corresponding to the desired transmitter. The receiver tuning is accomplished by placing a tunable filter in front of the optical detector. The filter permits the desired wavelength channel to go through, while blocking all the other DWDM channels. This early broadcast-and-select network topology could not support many users or reuse wavelength, and did not offer any security measures. More advanced network prototypes ensued (Kaminow *et al.* 1996), providing routing and wavelength management features present in modern networks, while continuing to advocate for tunable receivers. In today's deployed optical networks, both at the core and metro levels, tunable channel selection filters are not used for establishing connections (that task is performed by reconfigurable add-drop filters with WSSs). However, as fiber network venture deeper into access network and eventually to individual homes, tunable filters may once again be considered as they are the least expensive wavelength-tunable element. Other uses for tunable filters in today's networks are primarily for system monitoring. It is imperative to maintain the operational parameters of optical networks, and track any deviations from them. This is typically accomplished by tapping a fraction of the optical signal, and scanning the optical bandwidth with the filter to monitor the presence and power level of all channels in the fiber.

MEMS technology is very attractive for implementing tunable filters, as most tuning mechanisms are based on motion. Other competitive technologies do exist; however, the wide tuning requirement for spanning the entire communication window is typically more challenging for these alternative technologies. The tunable filter structures most commonly implemented using MEMS can be categorized by function: tunable resonant devices, deformable gratings, and spatial filters of spectrally dispersed light.

3.07.4.2.1 Tunable resonant devices

One of the simplest and well-known spectral filters is the Fabry–Perot interferometer or etalon, comprised of two partially reflective and mutually parallel surfaces separated by a finite gap. The transfer function of

the Fabry–Perot etalon consists of narrow spectral peaks with unity transmission; the complementary reflectance consists of the entire spectrum except for narrow spectral notches. The transmission/reflection is periodic in frequency, with the frequency period denoted by the free spectral range, $\mathrm{FSR} = c/(2d)$, where d is the etalon gap and c is the speed of light in vacuum, under the assumption that the gap region refractive index is unity (air) and under normal incidence. A transmission peak occurs whenever the round trip path length is equal to an integer multiple of an optical wavelength; hence, it is said to be resonant when the condition is satisfied. The half-width, Δf, of the Fabry–Perot etalon spectral peak is determined by the mirror reflectivity, R, and is defined by

$$\Delta f = \frac{c(1-R)}{2\pi d\sqrt{R}}$$

Narrow peak widths, which are necessary for channels isolation in DWDM, are achieved when the reflectivity approaches unity. The dimensionless ratio of the FSR to the half-width Δf is denoted by the finesse, $\mathcal{F}$, of the etalon and is given by

$$\mathcal{F} \equiv \frac{\mathrm{FSR}}{\Delta f} = \frac{\pi\sqrt{R}}{1-R}$$

The finesse measures the resolving power of the etalon, and is strictly determined by the mirror reflectivity. Locations of the transfer peaks are shifted by adjusting the gap separation. The FSR periodicity is inherent to resonant structures and is an undesirable feature when one wants to select a signal channel from the communication bandwidth. To overcome this effect, a telecom Fabry–Perot filter is constructed with a very narrow gap, which has a correspondingly large FSR. The requirements that need to be satisfied are that the FSR be larger than the communication bandwidth – to eliminate the periodicity within the bandwidth – and that the resolving power is sufficient to perform channel selection, implying that the gap region be on the order of a few wavelengths and that the mirrors have high reflectivity. The gap size as well as the mirror displacement requirement suggests that MEMS devices are well suited for construction of the etalons. MEMS Fabry–Perot etalons have been demonstrated in both in-plane and surface-normal arrangements (**Figure 9**). The latter is more common, as it is easier to fabricate using surface micromachining processes, and high mirror reflectivity is achieved by metallization and/or multilayer coatings. The actuation of one of the etalon's end mirrors is typically achieved with electrostatic forces or thermal bimorph displacement. The high finesse values that can be achieved with high-quality mirrors in the surface micromachining processes results in narrow transmission peaks. The MEMS surface micromachining diversity of material systems is demonstrated in the reported Fabry–Perot implementations in silicon (Niemi *et al.* 2001), GaAs (Amano *et al.* 2000, Kim *et al.* 2003a), InP (Daleiden *et al.* 2002), and elastomers (Blake 2003). In-plane Fabry–Perot implementations are based on DRIE in a thick SOI wafer (Yun *et al.* 2003). The two etalon mirrors are formed by etching vertical sidewalls in the silicon, one of which electrostatically actuates toward the other for tuning, and achieving relatively low finesse due to the low Fresnel reflectivity between silicon and air. Because it is difficult to metalize the vertical mirrors, the reflectivity can be enhanced by etching multiple vertical mirrors satisfying the Bragg criterion (Lipson and Yeatman 2005, Saadany *et al.* 2006). To achieve a significant reflection enhancement from the multiple reflectors, the surface quality of the vertical mirrors has to be improved after the deep etch process. The tuning speeds for electrostatic mirror actuation are typically submilliseconds, with a few volts required for submicron motion necessary for tuning across the communication band. Thermal bimorph actuation requires longer settling times, and consumes power in maintaining a filtering wavelength. Narrowband thermal tuning can also be used for modulating the refractive index of the gap region of a solid Fabry–Perot filter, as demonstrated with a silicon ridge waveguide with etched Bragg end mirrors (Angulo Barrios *et al.* 2004, Yun 2003). However, with extreme temperature tuning (up to 450°C), 41.7 nm was achieved for a silicon Fabry–Perot membrane with dielectric Bragg mirrors (Hohfeld and Zappe 2005).

A second category of resonant filters is implemented with microrings or microdisks disposed between two optical waveguides. Both the microdisks and microrings support propagation modes; in the microring case these modes are those of the curved waveguide structure, whereas in the microdisk case the propagation modes are azimuthal and are commonly termed whispering gallery modes (**Figure 10**). Each of the waveguides can couple energy into the microdisk/ring evanescently. The microdisk/ring form a recirculating cavity for the propagating light, and a resonance condition occurs when propagation around the disk/ring is an integer multiple for a certain wavelength of light. If the light coupled into the microdisk/ring from the input waveguide does not

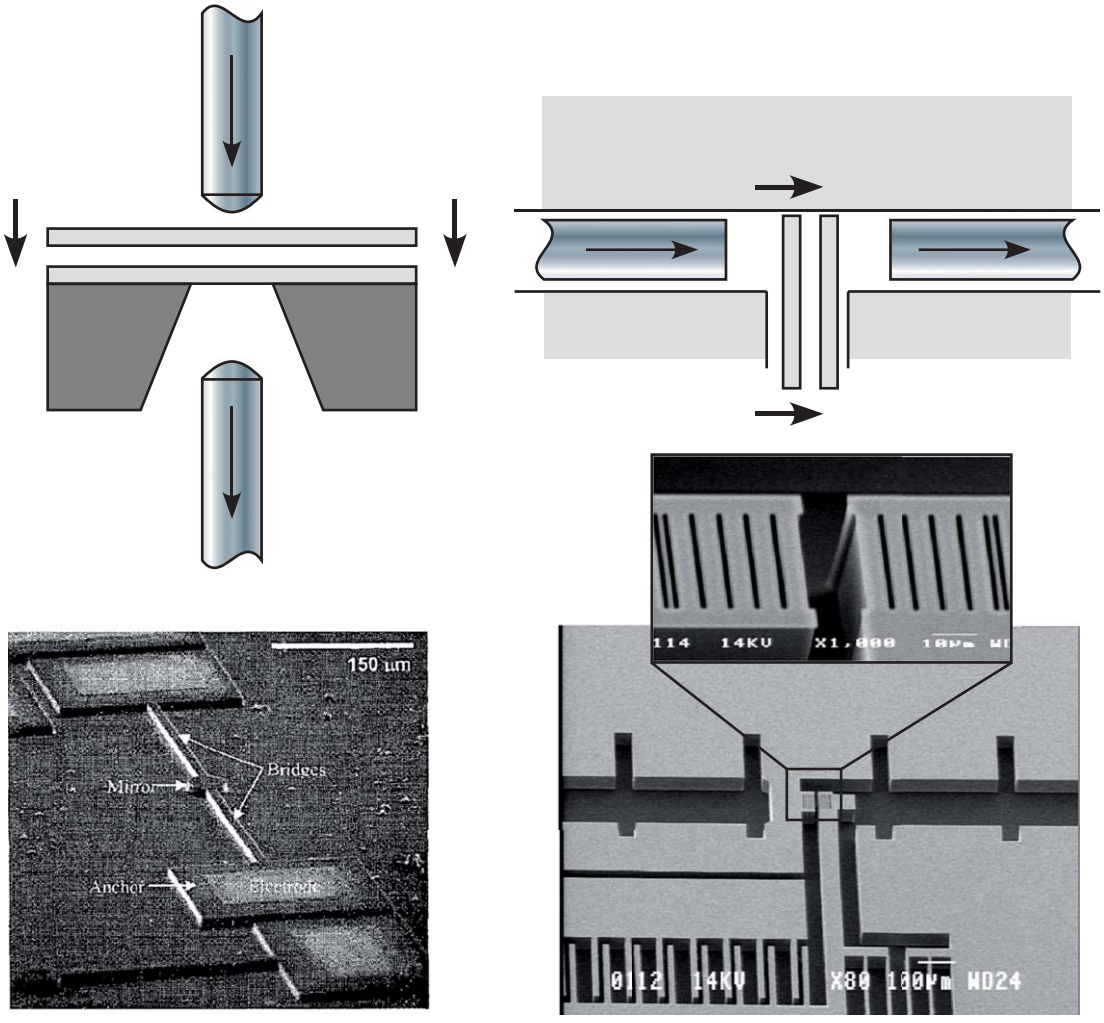


Figure 9 Fabry–Perot tunable filters. *Left*: Surface-normal arrangement and implementation. (Source: Kim C-K *et al.* 2003a 47 nm Tuning of thermally actuated Fabry–Perot tunable filter with very low power consumption. *IEEE Optical MEMS Conference*, Hawaii, HI, USA, © [2003] IEEE.) *Right*: In-plane arrangement and implementation. (Source: Saadany B *et al.* 2006 Electrostatically-tuned optical filter based on silicon Bragg reflectors. *IEEE Optical MEMS Conference*, Piscataway, NJ, USA, © [2006] IEEE.)

satisfy the resonance condition, the light will also couple out to the same waveguide and continue to propagate on the same input waveguide unimpeded. However, if the light does satisfy the resonance condition, the light will couple into the microdisk/ring from the input waveguide, and likewise will couple out of the disk/ring onto the second, drop waveguide. As in the Fabry–Perot case, the microdisk/ring bandwidth will be determined by the coupling coefficient into the cavity (similar to the partial reflectivity of the etalon), and the path length around the disk/ring. Moreover, the resonance condition is satisfied periodically in frequency, giving rise once again to a free spectral range. Their ratio is the characteristic quality factor, or Q, of the resonator. Because it is desirable to have only one resonance peak within the communication bandwidth, the microdisk/ring diameter is reduced to the order of a few microns. The described microdisk/ring resonators are fixed structures, and have been extensively explored for passive filtering applications. The operation of the microdisk/ring can be modulated by controlling the coupling into the resonator, as has been demonstrated by moving freely suspended waveguides closer to the microdisk (Lee *et al.* 2005), or by reducing the resonator's quality factor, as demonstrated by a metallic plate approaching the cavity and increasing the cavity's propagation losses (Nielson *et al.* 2005). A thermally wavelength-tunable and MEMS bandwidth-tunable microdisk cavity has been demonstrated on a silicon substrate.

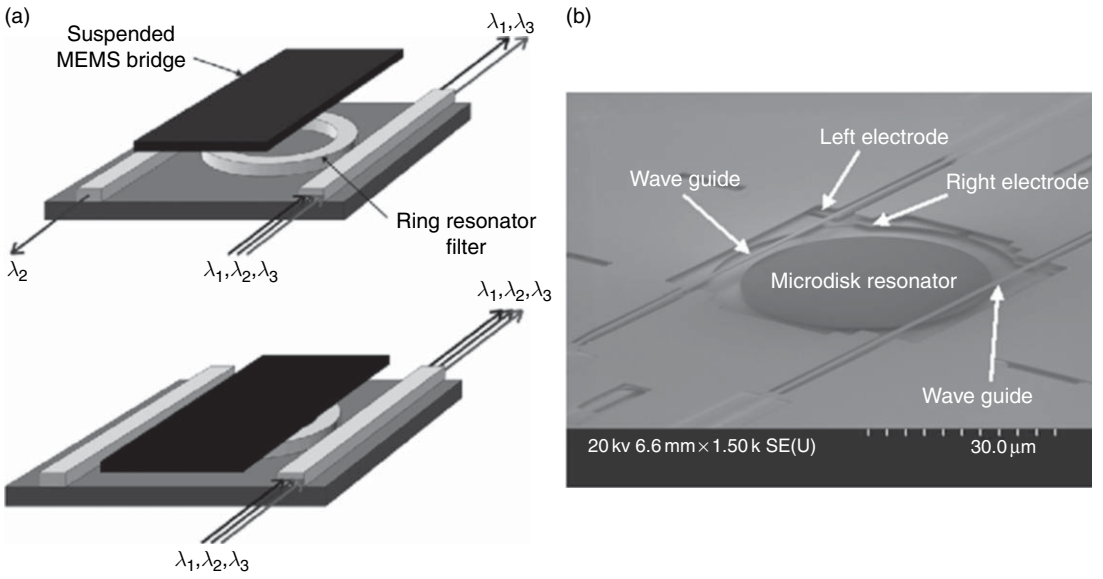


Figure 10 Microring and microdisk resonant filters. (a) Fixed-wavelength microring with modulation provided by inducing resonator losses by a metallic plate interacting with the light upon proximity. (Source: Nielson G N *et al.* 2005 Integrated wavelength-selective optical MEMS switching using ring resonator filters. *IEEE Photon. Technol. Lett.* **17**, 1190–2, © [2005] IEEE.) (b) Micrograph of microdisk with moving waveguides to control coupling coefficient. (Source: Yao J *et al.* 2005 Wavelength- and bandwidth-tunable filters based on MEMS-actuated microdisk resonators. *Optical Fiber Conference (OFC)*, Anaheim, CA, USA, paper OTuM1, © [2005] IEEE.)

Heaters placed near the microdisk increase its refractive index and coupling to the resonator with electrostatic attraction provides a tunable passband between 12 and 41 GHz (Yao *et al.* 2005).

A final class of a tunable resonant devices is based on a thin waveguiding slab with a grating pattern formed on top (Rosenblatt *et al.* 1997). When such a structure is illuminated with an incident light beam, part of the beam is directly transmitted through the structure and part is diffracted by the grating and is trapped in the waveguide layer. Some of the trapped light in the waveguide layer is then diffracted out so that it interfaces destructively with the transmitted part of the light beam. At a specific wavelength and angular orientation of the incident beam, the structure resonates; namely, complete interference occurs and no light is transmitted. For wavelength tuning purpose, the resonant filter is attached to a MEMS tiltable platform (Niederer *et al.* 2004). The reflected wavelength is tuned by changing the angle of incidence of the resonant grating filter. The grating is patterned by holographic recording and dry etching. The demonstrated filter worked over a wavelength range of 1520–1580 nm with a 0.5 nm half-width for a single polarization. The MEMS platform is DRIE processed, and electrostatically actuated over $\pm 4^\circ$ with a driving voltage of about 60 V. Alternatively, a modulated resonant grating structure has been demonstrated, in which a thin suspended silicon layer is etched with the grating structure (Kanomori *et al.* 2006). The gap separation between the patterned silicon layer and the underlying silicon can be controlled with bimorph actuators, resulting in a disruption of the resonant guiding. The silicon layer was patterned in two dimensions, resulting in polarization-independent operation.

3.07.4.2.2 Deformable gratings

A grating-based spectral filtering arrangement utilizes the dispersive property of diffraction to collect a finite bandwidth of light by an aperture, such as an optical fiber. The bandwidth captured by the fiber is a design parameter, influenced by the grating frequency and spatial dispersion. One method for wavelength tuning the filter is to change the grating frequency. The change in grating frequency changes the spatial dispersion, resulting in a different center wavelength impinging on the collecting fiber. Changing the grating frequency is typically impractical, as bulk gratings have fixed ruling. A tunable grating frequency can be achieved by implementing the diffractive element as a movable sequence of grating lines with a constant periodicity that can be continuously changed. The constant periodicity property is important for a

uniform diffraction response resulting in high resolving power; without this uniformity, the filtering response would broaden. Therefore, MEMS tunable grating frequency designs implement identical support structures for the grating elements, such that the displacement is uniformly distributed (**Figure 11**). The displacement can be achieved by thermal expansion of the inter elements' support structure (Zhang and Liu 2000), strain distribution by spring structures upon electrostatic actuation applied to the end point (Torman *et al.* 2006), and distributed expansion by a piezoelectric effect of the support structure (Wong *et al.* 2004). Both the electrostatic and piezoelectric actuation mechanisms achieve submillisecond actuation speeds. An alternative tunable grating frequency implementation uses independent analog control of the position of each of the grating elements (Li *et al.* 2006). The moving elements are blazed at an angle to achieve high diffraction efficiencies, using a process that combines KOH etching and DRIE to realize mirror surfaces with a root mean square (RMS) roughness of 26 nm and stiff structures that resist out-of-plane bending. An optical experiment using the tunable blazed grating with a small number of elements demonstrated their basic filtering functionality. The challenge in using this approach is associated with the scaling to a large number of grating elements, required for high-resolution filtering.

A different technique for controlling the diffraction response of a grating utilizes a 1D micromirror array, where each micromirror can piston out of plane for phase encoding. The independently controllable phase modulators can be set to coherently direct light of a certain frequency to a diffraction direction (Belikov and Solgaard 2003, SasBerg *et al.* 2003). Various synthesis techniques can be employed for designing different filter response functions as a rectangular frequency response. Such filters use very large mirror arrays with simple piston-motion functionality for achieving high-resolution filtering. The 1D phase modulator can also be placed within a cavity to form a segmented Gires–Tournois interferometer (Yu and Solgard 2004). The back reflection plane of the traditional interferometer is replaced with the 1D micromirror array for phase modulation, resulting in both amplitude and phase control for

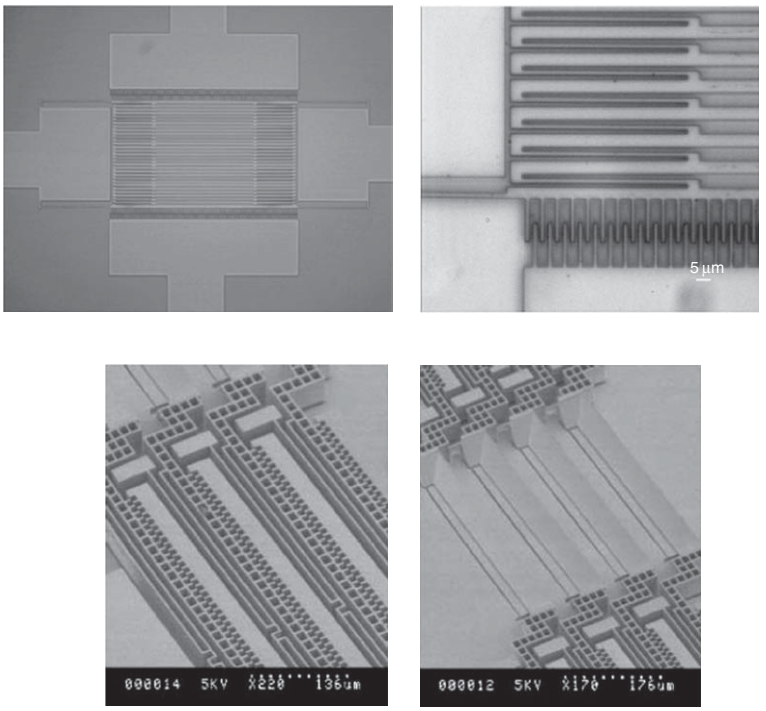


Figure 11 Deformable gratings. *Top*: Micrographs of electrostatic tunable grating frequencies with spring details on right. (Source: Torman M *et al.* 2006 Deformable MEMS grating for wide tenability and high operating speed. *J. Opt. A Pure Appl. Opt.* **8**, S337–40, © IOP.) *Bottom*: Tunable grating with independent motion of each grating element. (Source: Li *et al.* 2006 Tunable blazed gratings. *IEEE J. Microelectromech. Sys.* **15**, 597–604, © [2006] IEEE.)

each micromirror reflection. Collecting all the mirror reflections and performing an optical spatial Fourier transform results in a transfer function for the device that has the same form as a transversal filter in digital signal processing. Thus, conventional filter synthesis techniques can be used for tailoring the filter's bandpass as well as chromatic dispersion.

3.07.4.2.3 Spatial filter of spectrally dispersed light

The most common technique for wavelength tuning a grating and lens-based filter is by changing the tilt angle of the grating, resulting in a shift of the spatial dispersion and a selection of a different center wavelength impinging on the collecting fiber. However, as high-quality gratings utilize thick and thereby heavy substrates for stability, they are not compatible with MEMS actuation. However, a shift of the spatial dispersion can be similarly accomplished by a tilting micromirror inserted in the beam path (Berger *et al.* 2003). The resultant passband shape when coupling spatially dispersed light into a single-mode fiber is Gaussian, and the reported implementation achieved a 32 GHz–3 dB bandwidth, tunable over 40 nm. Similar implementations have achieved wider tuning with correspondingly wider passbands as well, demonstrating the interplay between resolving power and tuning range (Kimura *et al.* 2004). The collecting fiber can also be replaced with an apertured photodetector, for implementing a complete tunable receiver (Berger *et al.* 2004). While angle tuning of a bulk grating is difficult to implement in MEMS technology, thin-film filters are much more compatible. A thin-film filter with a 200 GHz bandpass response was mounted on a thermally actuated MEMS rotation stage (Unamuno and Uttamchandani 2005). The measured mass of the filter was 6 g, and the MEMS rotation stage achieved 1.5° rotation for a 700 mW power draw. However, thin-film filters have weak angular tuning range, in the reported case of 100 GHz only.

The Gaussian passband shape associated with the direct coupling of dispersed light to a single-mode fiber is often undesirable. An alternative approach often utilized by WSSs and blockers (see Sections 3.07.4.5 and 3.07.4.5.1) is that of dispersing the light with a grating and filtering by a finite mirror, followed by the reverse process of combining the spectral components with the same grating and coupling to the fiber. The double passing of the grating (first to disperse and second to recombine) removes any frequency dependence for spectral component beams striking the mirror far away from the mirror edges. A filter with tunable passband width and center frequency has been demonstrated using this approach together with a MEMS tilt mirror (Wilson 2006). The MEMS mirror shifts the spatially dispersed light that is incident on the spectral mirror (**Figure 12**). The reflective spectral mirror in this case is triangular, chosen to select the passband width. The center frequency of the tunable filter is selected by shifting the dispersed light along the axis of the dispersion. The passband width of the tunable filter is chosen by translating the dispersed light in the orthogonal direction, such that the dispersed light encounters a broader or narrower segment of the triangular mirror. This filter can also operate in hitless mode, which is a unique distinguishing feature compared to all the above filters. Hitless operation is important for channel add-drop, where the act of channel tuning cannot cause a disruption to the intermediate channels. The MEMS micromirror has to tilt a broad beam in the filter implementation due to resolution requirements, and the experiment used a magnetically actuated mirror with an aperture of 4.5 mm × 6 mm, with an angular range of ±1.5° for a ±2.3 V drive. The filter demonstrated wavelength tuning over 36 nm, and passband width settings from 28 to 210 GHz, with a 4.6 dB insertion loss. Transmission experiments with 10 Gb s^{-1} data rate signal showed negligible power penalty due to the grating double passing arrangement.

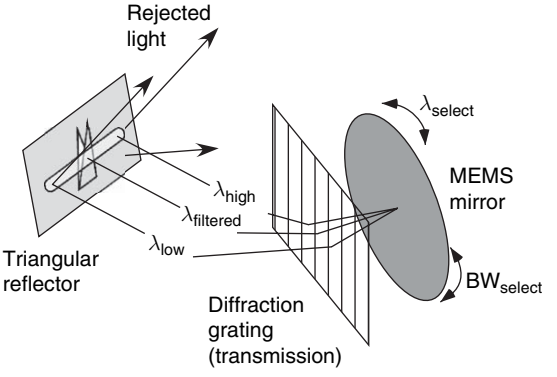


Figure 12 Wavelength and bandwidth tunable filter. Spatially dispersed signals can be scanned with microelectromechanical systems (MEMS) mirror in left–right direction to select wavelength and up–down to prescribe bandwidth. (Source: Wilson G 2006 Spectral passband filter with independently variable center wavelength, bandwidth. *IEEE Photon. Technol. Lett.* **18**, 1660–2, © [2006] IEEE.)

3.07.4.3 MEMS Variable Optical Attenuators

Terrestrial optical communications systems are required to operate over large dynamic ranges, as topography dictates irregular distribution of nodes (or cities) and amplifying stations. The path length diversity of optically amplified systems implies that the optical powers can vary by up to 20 dB, but as these are known system attributes, they can be addressed by placing fixed attenuation elements in the system. Other factors lead to changing conditions, such as the number of wavelength division multiplexing (WDM) channels, which typically increases as the network is required to handle more users and capacity, and less gain per channel is available. Furthermore, the system is designed with extra margin to allow for component aging, resulting in lower transmitter powers and amplifier gain over the lifetime of the system. This leads to a large variation in the incident power that the optical receiver might experience. On the other hand, the optical receiver, comprised of an optoelectronic detector and electrical amplifiers, typically operates over a limited dynamic range. To limit the receiver's impinging optical power to the acceptable dynamic range, a variable optical attenuator (VOA) is normally introduced as part of the receiver. The VOA is a simple device with a single input fiber and a single output fiber, where the experienced optical loss between the input and output fibers can be continuously set. VOA can also be used as replacement for simple 1×2 switches, in an arrangement consisting of a power splitter and a VOA placed on each branch, enabling broadcast and select topology at the cost of higher losses. Other uses, requiring faster VOA in optical networks, are for transient suppression of amplifier dynamics that can wreak havoc on the stability of the network, and for possible low-frequency modulation of the transmission bit stream to encode system signaling information that can be received by a low-frequency response photodetector (a technique known as pilot tones) (Darcie *et al.* 2005).

Because the VOA is integrated into each channel receiver and has additional network uses, it is a high-volume optical component, implying it must be inexpensive, compact, and preferably consume very little power. These attributes make VOA functionality well suited for a MEMS implementation.

3.07.4.3.1 Shutter VOA

Optical loss can be introduced by blocking the light radiating out from the input fiber en route to the output fiber. A movable shutter placed in the beam path can selectively block the light in proportion to the shutter's displacement. The optical arrangement can be very simple by setting a minimal gap between the input and output fibers, where the gap region is intended for the shutter placement (**Figure 13**). Such an arrangement does not require any imaging optics, greatly simplifying the packaging requirements, and has a planar form factor. The intrinsic optical loss of this arrangement is due to light diffraction between the input and output fibers, as well as a possible contribution from lateral fiber misalignment in the packaging process. To minimize the intrinsic optical losses, it is necessary to minimize the gap size; hence, the shutter width should be very small, a requirement compatible with MEMS technology. Counter to the requirement that the shutter width should be small, its lateral dimension should be comparable to the optical fiber diameter, or $\sim$125 μm. The shutter range of motion should preferably be greater than the fiber radius so that the unactuated shutter should be further from the optical fibers during the assembly process. The shutter normal should also be tilted with respect to the fiber axis, so that back reflections will not couple back into the input fiber.

The large lateral dimension and travel range requirements for the MEMS shutter and its orientation normal to the fiber plane impose design limitations. A surface micromachined optical shutter VOA has been demonstrated with the shutter element rotated 90°, normal to the substrate (Barber *et al.* 1998, Giles *et al.* 1999). The shutter element was attached via a lever arm to a parallel plate capacitor actuator. Upon actuation, the shutter elevated further from the substrate via a pivot and hinge structure. A voltage range of 0–28 V applied to the capacitor plate resulted in full attenuation, with a switching speed faster than 100 μs. The fiber-to-fiber gap size of 20 μm yielded a minimal loss of 0.8 dB. An alternative bulk micromachining process optical shutter VOA utilizes DRIE in a thick silicon layer on an insulator wafer (Marxer *et al.* 1999). The VOA uses a comb-drive actuated shutter and fiber placement grooves for simple alignment. To minimize the risk of damage during fiber placement, the large gap size of 60 μm was maintained. To reduce the effect of beam diffraction and subsequent optical losses upon coupling to the output fiber, the entire package including the gap region was filled with index matching oil. The optical losses were reduced from 2.5 to 1.5 dB, with an additional damping benefit provided by the oil, eliminating mirror ringing effects. A voltage range of 0–30 V applied to the comb-drive actuator resulted in full attenuation, with switching speed of 4.5 ms (due to the oil damping effect).

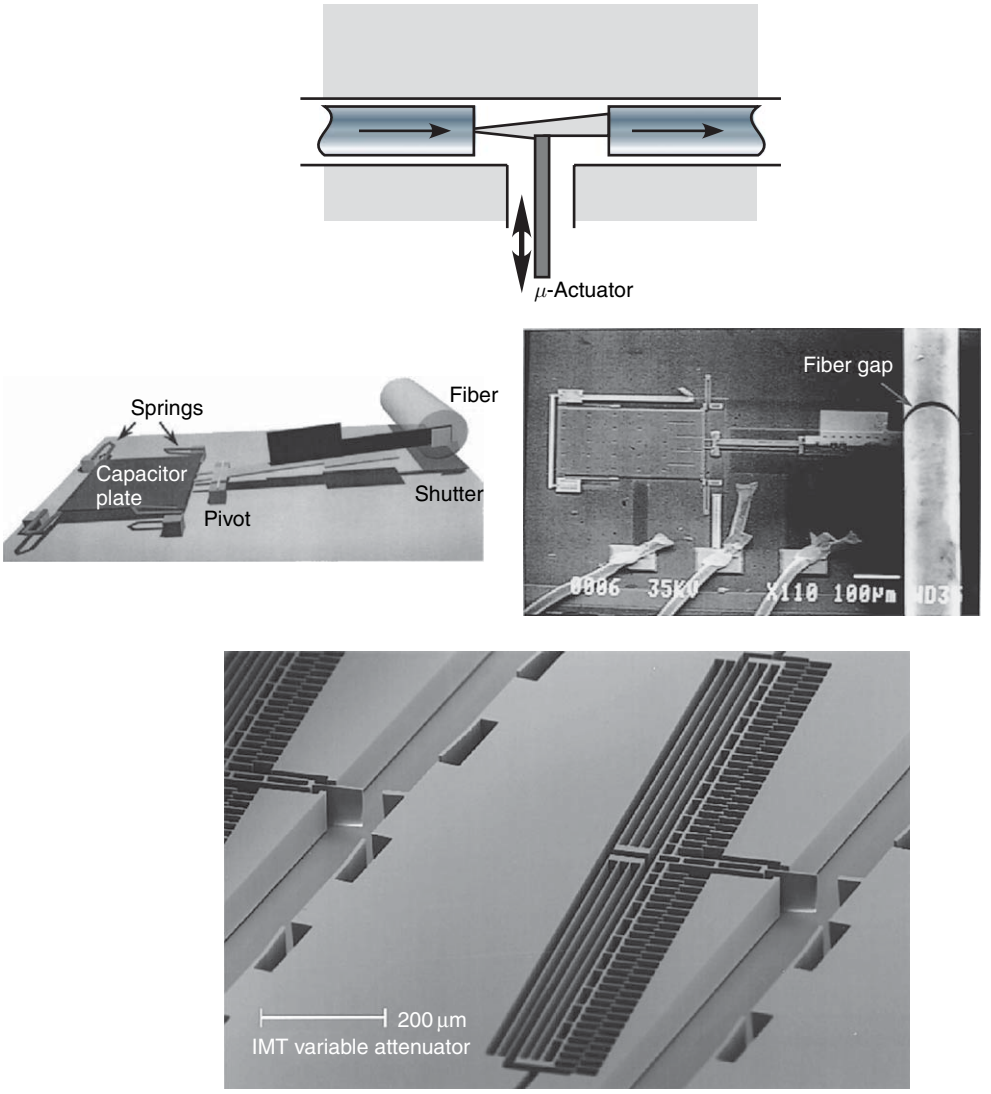


Figure 13 Shutter-type variable optical attenuator. *Top*: Schematic representation. *Middle*: Surface micromachining implementation. (Source: Barber B *et al.* 1998 A fiber connectorized MEMS variable optical attenuator. *IEEE Photon. Technol. Lett.* **10**, 1262–4, © [1998] IEEE.) *Bottom*: Bulk micromachining shutter (Source: Marxer C, Griss P, de Rooij N F 1999 A variable optical attenuator based on silicon micromechanics. *IEEE Photon. Technol. Lett.* **11**, 233–5, © [1999] IEEE.)

The VOA implementations described above require a steady voltage to maintain the attenuation settings, sometimes coupled with a feedback loop for added stability. A desired alternative is a latching switch, which requires active actuation in order to reach an attenuation state, but can then hold the state indefinitely with no power draw. A Vernier latching technique can be used to hold the shutter in the desirable attenuation state (Syms *et al.* 2004b, Unamuno and Uttamchandani 2006). The Vernier latch attenuation level progresses in discrete steps in one direction, and has a release actuator to return the VOA to the unactuated state. This operation mode is not amenable to active power level control feedback, often necessary for stable operation of optical networks. Moreover, the latch release operation and re-engagement can cause deleterious transient effects in networks.

Some concerns with polarization dependence exist when using the shutter-type actuator, due to vectorial diffraction effect from the shutter's edge. A MEMS iris alternative to the shutter has been demonstrated, with has greater symmetry about the optical fiber mode. The iris consists of four simultaneously moving blades, continuously adjusting the size of a central opening in the beam path (Syms *et al.* 2004a).

A self-powered VOA offers an interesting feature for surge suppression in optical networks (Aksyuk *et al.* 1999, Hirat *et al.*, 2005). The VOA includes a photogenerator, which provides sufficient voltage to drive the VOA in feedback mode. A power tap after the shutter drives the photogenerator in a self-governing mode; as the incident power on the detector increases, so does the photogenerator voltage, engaging the VOA shutter and regulating the output through increased attenuation.

3.07.4.3.2 Beam displacement VOA

An alternative loss-generating scheme to the optical shutter consists of an optical arrangement that displaces the beam emerging from the input fiber, thus controlling the coupling efficiency into the output fiber. The scheme suggests that a more complicated optical system is required compared to the simple fiber gap arrangement of the shutter solution. However, the optical losses associated with a well-designated optical system between the input and output optical fibers can result in lower losses compared to propagation over a finite fiber gap. MEMS micromirrors are well suited for implementing the beam displacement mechanism of the optical system.

One method for imparting beam translation utilizes a tilting micromirror in the beam path. A 3D configuration often used places the input and output fibers side by side with their polished endfaces at the front focal plane of a common lens (**Figure 14**). The MEMS tilting mirror is placed at the back focal plane of the lens. When the mirror is in the unactuated state, light from the input fiber is collimated by the lens onto the mirror, is reflected back, and subsequently focused by the common lens into the output fiber. The intrinsic coupling efficiency of this configuration is limited only by the quality of the lens (antireflection losses and aberrations) and the mirror reflectivity, and can routinely achieve better efficiency than -1 dB. Polarization effects are also minimized because the beam is almost normally incident onto the mirror. When the mirror is tilted, the focused beam is shifted from the output fiber and coupling losses increase according to fiber misalignment calculations. The relationship between the mirror tilt $\Delta\theta$ and beam shift Δx is $\Delta x = 2f\,\Delta\theta$, where f is the focal length of the lens (obtained by geometric optics using small angle approximations). Thus, larger focal lengths require smaller tilt angles for imparting the same loss values. However, the collimated optical mode size at the mirror also increases with the focal length, requiring a larger mirror with greater sensitivity to mirror curvature. The lens focal length is typically kept small (order of 1 mm), resulting in a smaller system size. The collimated beam diameter is $D = 2f\lambda/(\pi w_0)$

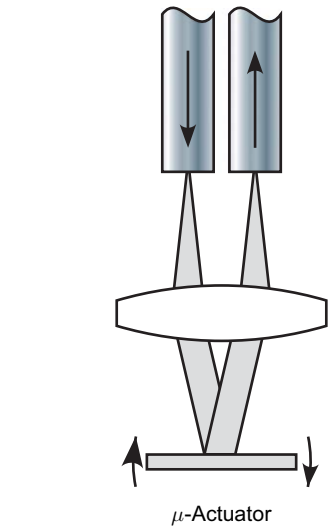


Figure 14 Beam scanning-type variable optical attenuator in a 3D layout. *Top*: Schematic representation. *Bottom*: Micromirror and support details fabricated using silicon bulk micromachining. (Source: Isamoto K *et al.* 2004 A 5-V operated MEMS variable optical attenuator by SOI bulk micromachining. *IEEE J. Selec. Top. Quant. Electron.* **10**, 570–8, © [2004] IEEE.)

($D = 200\,\mu m$ using a 1-mm lens), requiring a mirror size of $1.5D$ (or 300 μm for a 1-mm lens) for minimal insertion losses, a size that is compatible with MEMS technology. A commercial, bulk micromachining, electrostatic MEMS tilting mirror (30-μm-thick, 600 μm diameter silicon mirror) demonstrated 5 V full actuation and a 3 ms underdamped response time (Isamoto *et al.* 2004). The entire VOA package is 5.6 mm in diameter and 23 mm long. Alternative actuation techniques for the tilting micromirror can be employed (Diemeer and Dekker 2002).

A planner beam translation method places one or more micromirrors in the beam path between the input and output fibers (**Figure 15**). Both fibers have a lens attached to their end facets, creating a collimated beam with an extended propagation distance between the two lensed fibers (typically hundreds of microns). A 90° folded geometry with a single MEMS translation or rotation mirror (Bashir *et al.* 2004, Yeh *et al.* 2006), or a 180° folded goemetry with two MEMS translation mirrors have been demonstrated (Cheng *et al.* 2004). In all cases, DRIE must be employed due to the fiber diameter and the planar configuration. The different MEMS designs utilize comb-drive actuators with low driving voltage (up to 30 V) and fast response times (millisecond range). The optical arrangements have low intrinsic losses (<1 dB), but the long beam path is sensitive to fiber placement errors, sometimes requiring active alignment within the alignment grooves. Because the reflective mirror surface is the etched silicon sidewall, surface roughness contribute to the intrinsic device loss and can limit the attenuation range due to scattering.

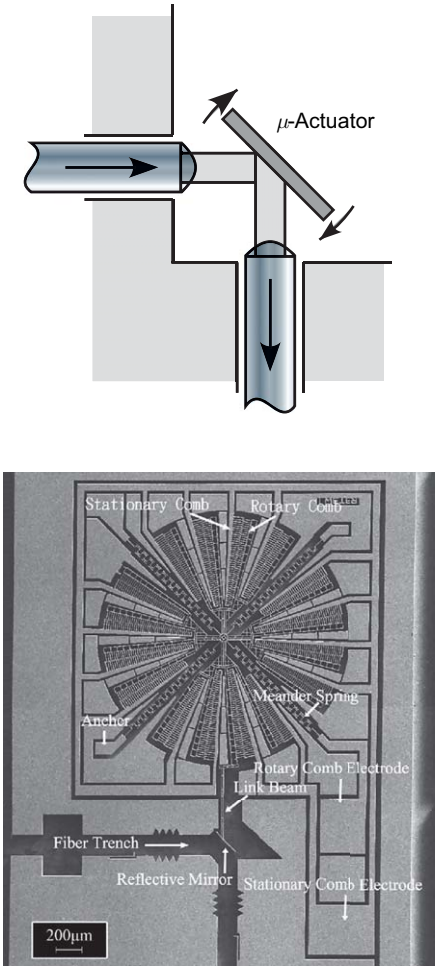


Figure 15 Planar beam scanning-type variable optical attenuator (VOA). *Top*: Schematic representation. *Bottom*: VOA with micromirror attached to a rotation stage, fabricated using bulk micromachining. (Source: Yeh J A, Jiang S-S, Lee C 2006 MOEMS variable optical attenuators using rotary comb drive actuators. *IEEE Photon. Technol. Lett.* **18**, 1170–2, © [2006] IEEE.)

An alternative beam displacement technique to a mirror in the beam path is a displacement of the input fiber with respect to the output fiber with the aid of MEMS actuators (Kanamori *et al.* 2003). One implementation utilized moving silicon waveguides suspended in air instead of optical fibers (Tilleman *et al.* 2005), with the silicon waveguides operating in the multimode regime (having a cross section of $14\,\mu m \times 14\,\mu m$). The moving waveguide is attached to a comb-drive actuator. Due to the multimode structure, the attenuation versus displacement is highly dependent on mode number and polarization. However, the fundamental mode is very well behaved. The small air gap in the waveguide introduces diffraction losses and can also introduce Fabry–Perot effects, in the form of wavelength-dependent attenuation, due to the high reflectivity in the silicon-to-air interface.

3.07.4.3.3 Variable reflectivity mirror

The previous solutions, whether based on a shutter inserted into the optical beam or beam displacement with tilting or moving mirrors, require MEMS devices with a relativity large motion range. An alternative arrangement based on mirrors whose reflectivity is adjustable can be utilized. The simple and low-loss optical arrangement with the common imaging lens is often used, with the variable reflectivity mirror replacing the tilting mirror. The variable mirrors can be based on interferometric and diffractive MEMS (DMEMS) solutions, both requiring a motion range smaller than the optical wavelength.

An interferometric mirror based on the interference between two membranes with a controllable gap size between them achieved tunable reflectivity (Ford *et al.* 1998). The mirror, termed mechanical antireflection switch (MARS), functions effectively as a dielectric coating on a silicon substrate. The active element is a silicon nitride layer one-quarter wave optical thickness, just as for a single layer dielectric antireflection coating (**Figure 16**). This silicon nitride layer is separated from the silicon substrate by $(3/4)\lambda$. At the $(3/4)\lambda$ separation, the antirefection coating becomes a single layer dielectric mirror with up to 72% reflectivity. However, the membrane is deformable. Voltage applied to electrodes on top of the membrane creates an electrostatic force, which pulls the membrane closer to the substrate. When the membrane gap is reduced to $\lambda/2$, the layer becomes an antireflection coating with close to zero reflectivity. Due to the small range of motion and the high silicon nitride film stress, the device can operate at 3 μs response times. A 28 dB attenuation dynamic range was observed for a 0–35 V operating voltage range. The device is nearly polarization independent on account of the normal incidence. However, due to the wavelength dependence in the interference effect, there is a strong wavelength-dependent attenuation. The wavelength dependence is especially pronounced about the peak attenuation value, where perfect destructive interference can only be achieved for a single wavelength.

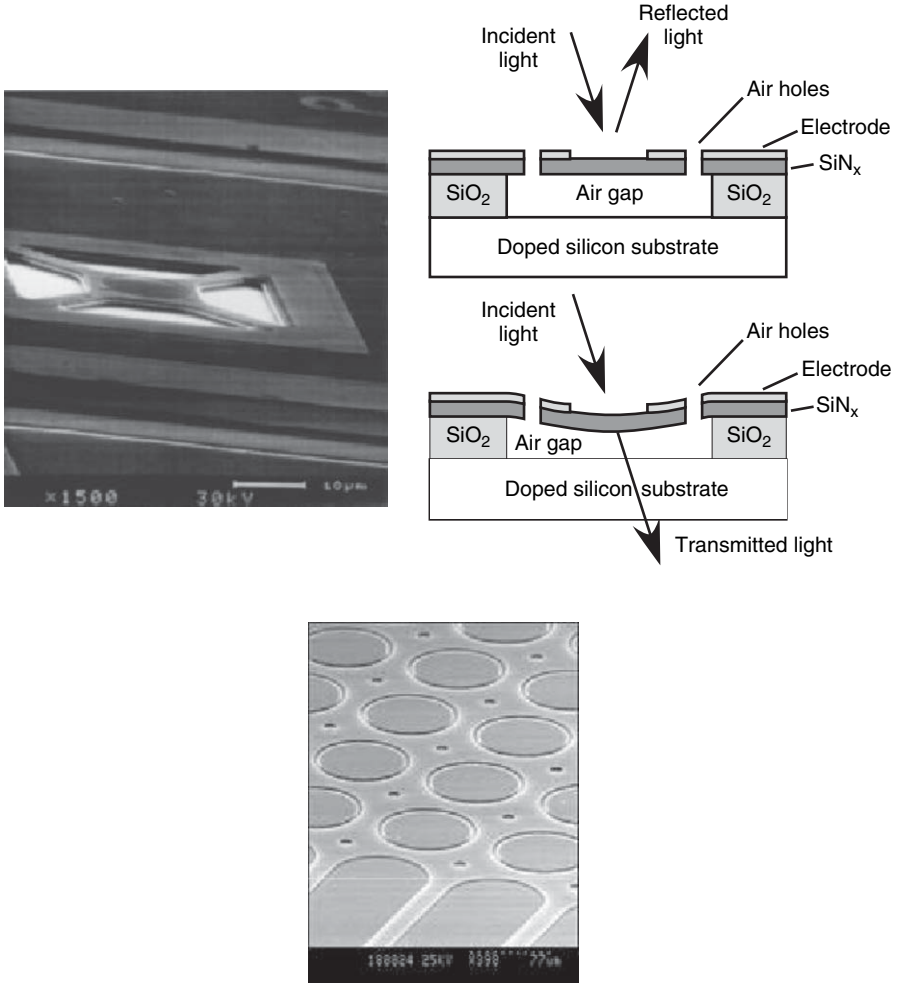


Figure 16 Variable reflectivity mirrors. *Top*: Controlled interference between moving thin membrane and substrate layer. (Source: Ford J E *et al.* 1998 Micromechanical fiber-optic attenuator with 3 μs response. *IEEE J. Lightwave Technol.* **16**, 1663–70, © [1998] IEEE.) *Bottom*: Controlled interference between moving plate and static pillars. (Source: Godil A 2002 Diffractive MEMS technology offers a new platform for optical networks. *Laser Focus World*, May.)

DMEMS mirrors can also be used in a VOA application, by utilizing the zeroth diffraction order for coupling back into the output fiber (Godil 2002). However, the traditional DMEMS, consisting of periodic ribbons where half can be actuated, does not offer the necessary performance for optical communication application. Ignoring the gap region in the DMEMS, the power in the zeroth diffraction order is proportional to $(1/2)\,(1+\cos(\Delta\phi))$, where $\Delta\phi$ is the phase difference between the reflected signal from the actuated versus unactuated ribbons. The attenuation can be set to zero when the phase difference, $\Delta\phi$, is π ($\lambda/4$ translation), but the condition is satisfied for a single wavelength only due to the wavelength dependence in the phase delay. To circumvent the wavelength dependence, the DMEMS device can be optimized taking three parameters into account; the phase and period fractions of the movable ribbons, the stationary ribbons, and the exposed substrate in the gap regions. The extra gap factor allowed for a design that is nearly achromatic in the attenuation state. The ribbon structure gives rise to polarization dependence, due to the inherent orientation asymmetry. By creating a grating structure that is perfectly symmetrical in the two directions, polarization-independent operation is achieved, and the achromatic response criterion is also included in the design. The reflective surface is comprised of fixed round disks, a movable membrane around the disks, and a gap region between the membrane and the disks. Incident light is reflected on all three surfaces, and the geometrical layout ensures low polarization and wavelength dependence. The packaged VOA is 5.4 mm in diameter, and 18 mm long, and operates at a 0–5 V voltage range. According to a commercial press release, as of March 2005, 40 000 VOAs have been shipped to customers, demonstrating the large market demand for VOA modules.

3.07.4.4 MEMS Optical Fiber Switches

Optical fiber switches route the entire guided optical signal from one fiber to another, and scale in size from fundamental 1×2 and 2×2 switches to $N \times N$ switches, where N can be as large as one thousand. When possible, it is more desirable to switch signals in the optical domain due to the cost and power savings over optoelectronic equipment. Their uses in optical communication systems and networks are as diverse as the scale range they encompass. The basic 1×2 switch is often used for protection against equipment failure. For example, metropolitan-scale fiber rings often utilize fiber redundancy using a BLSR (bidirectional line switched ring) topology. The 1×2 switch is used to switch from the main fiber to the backup fiber when a loss-of-signal occurs (e.g., fiber cut). A fundamental 2×2 switching element can be used as a stand-alone switch or within a multistage interconnection network architecture (such as a Benes network) for constructing larger switch fabrics. The 2×2 switch can be used for implementing optical add-drop architecture on a per-channel basis. A $1 \times N$ switch, which routes from one fiber to one of any N fibers, can be used for efficient equipment sharing, such as optical monitoring at an amplification site. Finally, large $N \times N$ switches, often referred to as optical cross-connects (OXCs), are used to establish a desired connectivity pattern across many fibers. An OXC is therefore sometimes referred to as an automated patch panel, whose connectivity can be changed without the need for a technician's visit to the switch site. The OXC can be used to route individual WDM channels at a network node using opaque or transparent operating modes. A transparent switch, also termed OOO, is one where an input *optical* signal is switched in the *optical* domain and exits the switch in the *optical* domain. The photons flow through the switch unimpeded. OOO terminology contrasts the transparent switch to OEO switches, where an input *optical* signal is switched in an *electronic* switch fabric and exits in the *optical* domain again. An opaque switch uses the same switching fabric as the transparent switch, but isolates the optical switch from the optical input and output signals with electronic conversion. Hence, the opaque switch with OEOEO functionality regenerates the optical signal and does not suffer from wavelength contention issues.

While the smaller 1×2 and 2×2 optical fiber switches can be constructed using many competitive technologies, the larger switches are implemented exclusively using MEMS technology. The most common switching method uses micromirrors, yet other MEMS alternatives exist as well. Companies offering MEMS optical fiber switches have received much attention, as well as investment, during the telecom boom years. Very few of these companies survived, as the demand and deployment of large OXC switches never materialized. An opaque OXC architecture, while offering complete freedom of WDM channel routing restriction due to the signal regeneration, was too expensive to deploy. Transparent OXC functionality could be more effectively implemented using WSSs.

3.07.4.4.1 Small lensless MEMS switches

Small switches require simple optical arrangements, and the simplest optical arrangements have no imaging optics between the input and output fibers. The distance between the input and output fiber end faces in such nonimaging solutions has to be very small, so that the coupling losses due to beam diffraction will be minimal. The gap region between the fibers can be minimized if no intervening elements need to be placed in it. Hence, the basis for a simple 1×2 switch is a movable input fiber, which can be laterally actuated to a position in front of either output fibers (**Figure 17**). The positional accuracy or placement in front of the fiber is very important for minimizing the coupling loss. Therefore, moving fiber switches rely on bistable designs with mechanical stops that define the parking positions (Field *et al.* 1995, Herding *et al.* 2003, Hoffmann *et al.* 1999). The optical fibers are placed in alignment grooves, with the single fiber typically placed onto a moving thermal actuator owing to the required large fiber bending forces. While the power consumption of the thermal actuator is high (~1 W), it is only required in actuation, due to the bistable design. Reported switching time is 400 ms, which is too slow for a protection application. An electrostatic switch using a metalized fiber offers an alternative actuation method. The electrostatic actuation at 75 V achieved a 10 ms switching time. One disadvantage associated with the moving fiber switch is that the three fibers are manually inserted into their alignment grooves, and offer no axial alignment mechanism. Hence, the gap region is intentionally larger than the minimal clearance (typical values of 25 μm) and the gap will be different for the two output fibers, resulting in uneven losses. The moving fiber concept can also be implemented with moving waveguides, in which case the gap region can be photolithographically defined (Bakke *et al.* 2002, Ollier 2002, Stuart *et al.* 2003). The moving waveguide approach using electrostatic actuation has been demonstrated in buried glass waveguides with metallization around the cladding region, in polymer waveguides deposited on top of a conducting SOI actuator, in GaAs waveguides, and in silicon ridge waveguides. The moving waveguide approach typically achieves submillisecond switching times and actuation voltages of a few tens of volts. Due to Fresnel reflections at the waveguide–air interface, the gap region forms a Fabry–Perot cavity. The gap region has to be precisely etched to ensure that its width will coincide with the cavity transmission peak.

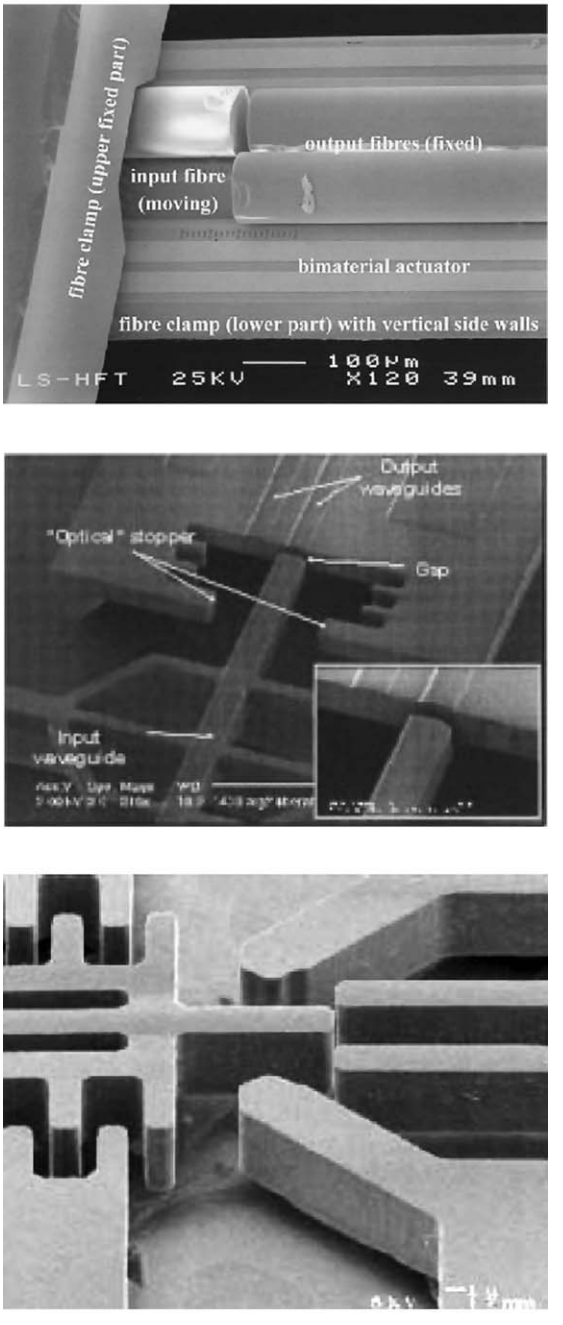


Figure 17 Lensless 1×2 switch with moving fibers/waveguides. *Top*: Fiber moving switch in a silicon actuator. (Source: Hoffmann M, Kopka P, Voges E 1999 All-silicon bistable micromachincal fiber switch based on advanced bulk micromachining. *IEEE J. Select. Top. Quant. Electron.* **5**, 46–51, © [1999] IEEE.) *Middle*: Silica-on-silicon moving waveguide. (Source: Ollier E 2002 Optical MEMS devices based on moving waveguides. *IEEE J. Select. Top. Quant. Electron.* **8**, 155–62, © [2002] IEEE.) *Bottom*: GaAs moving waveguide. (Source: Bakke T *et al.* 2002 Planar microoptomechanical waveguide switches. *IEEE J. Select. Top. Quant. Electron.* **8**, 64–72, © [2002] IEEE.)

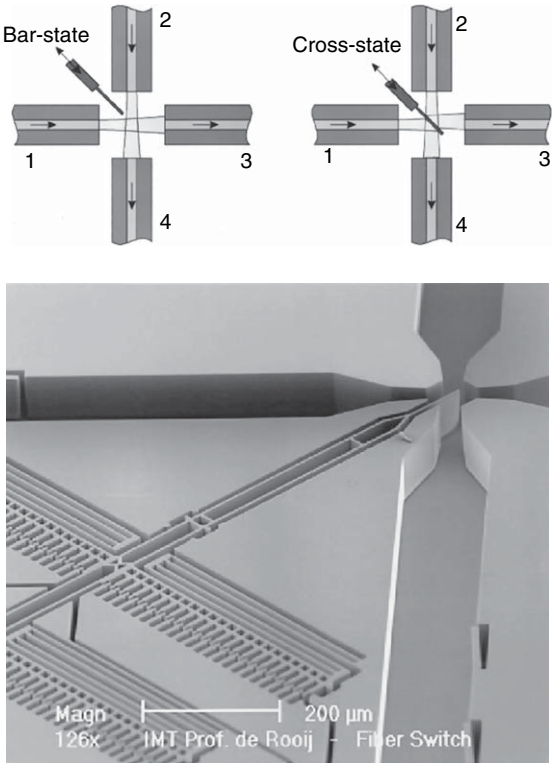


Figure 18 Lensless 2 × 2 fiber crossbar switch with inserted mirror. *Top*: Schematic representation of switch. *Bottom*: Image of microelectromechanical systems (MEMS) switch fabricated in silicon using bulk micromachining. (Source: Marxer C, de Rooij N F 1999 Micro-opto-mechanical 2 × 2 switch for single-mode fibers based on plasma-etched silicon mirror, electrostatic actuation. *IEEE J. Lightwave Technol.* **17**, 2–6, © [1999] IEEE.)

A more complicated arrangement is required for lensless 2 × 2 optical fiber switches, which is counter to the minimal gap requirement. The lensless approach brings the four fibers together in a cross configuration (**Figure 18**). Light from the two input fibers, labeled 1 and 2, can free-space propagate across the junction region and couple to two output fibers, labeled 3 and 4. When a mirror is placed across the junction region, the light is reflected from the mirror and the switch is placed in the cross state. The mirror is inserted into the beam path in a similar fashion to the optical shutter in bulk micromachined VOA. A DRIE in a thick silicon layer on insulator wafer forms the shutter, the comb-drive actuator, and fiber placement grooves (Marxer and de Rooij 1999). The etched silicon layer is 75 µm thick, sufficient to cover the entire core region when standard single-mode fiber is used. The fiber gap size is required to be at least as large as the fiber's outer diameter, due to geometric constraints associated with placing four fibers at 90°. However, the insertion losses due to diffraction after propagation over such large gaps are large. The gap size can be reduced by using fibers with reduced cladding, resulting in a smaller outer diameter. Losses can be further reduced by placing index matching oil in the gap region, reducing the diffraction angles. The index matching oil also serves to dampen the mirror response, eliminating mirror ringing. The mirror's actuation voltage was 60 V, and the response time was about 0.5 ms. Insertion losses at the bar state (no mirror inserted) was below 1.2 dB (best achieved value of 0.55 dB). In the cross state (with mirror reflection) losses increased to a typical value of 1.8 dB (best achieved value of 1.55 dB). The added loss originates from the mirror quality, as a vertically etched mirror will exhibit roughness impacting the reflectively.

3.07.4.4.2 *Sharing 1 × N* fiber switches

As the number of fiber ports in a switch increases, imaging optics have to be introduced in order to collect the light radiating out of the input fiber and efficiently couple the light to the output fiber. $1 \times N$ optical switches, which switch light from an input port to one of N output ports, have an optical system that accommodates $N+1$ optical fibers. Typical values for the number of fibers N is 4 or 8. Such switches are typically used in reverse, as $N \times 1$ switches, for sharing a common resource among the N fibers. Beam displacement with one tilting micromirror in the beam path is an effective switching technique for $1 \times N$ switches. The arrangement is similar to the dual fiber collimator employed by beam displacement VOA, with the $N+1$ fibers in the lens's front focal plane and a tilting micromirror in the back focal plane (**Figure 19**). Each of the N fibers can be addressed by a unique mirror angle. Because the beam size and the micromirror size scale with the lens' focal length, it is desirable to use a short focal length lens. However, as opposed to the two fibers required for VOA, $N+1$ fibers occupy a larger lens field of view that will result in aberrations playing a significant factor in the coupling to fibers that are farther off-axis. One way to reduce this effect is to use a longer focal length, compound lens, which is better corrected for aberrations. This approach has been used in the demonstration of a 1×160 fiber switch (Ford and DiGiovanni 1998). However, as the beam size increases with the focal length, the resulting mirror size is beyond the size of MEMS devices, and enters the realm of precision mechanical mechanisms. A

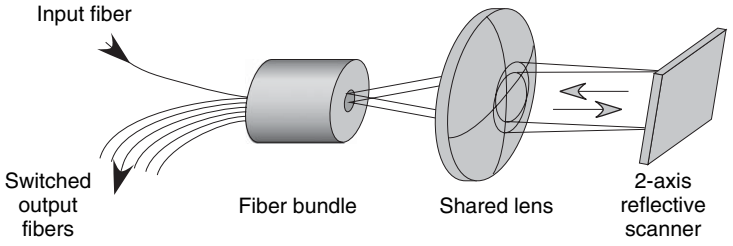


Figure 19 Sharing $1 \times N$ fiber switch. Schematic representation of beam scanning for selecting output fiber. (Source: Ford J E, DiGiovanni D J 1998 $1 \times N$ fiber bundle scanning switch. *IEEE Photon. Technol. Lett.* **10**, 967–9, © [1998] IEEE.)

large switch implementation used galvanometer rotation actuators with prisms in a reflective arrangement to accomplish the beam scanning, with a 20 ms switching time and a worst-case insertion loss of 3.5 dB. A second way to reduce the effects of the off-axis aberrations is to place a deformable mirror instead of the micromirror (Peter *et al.* 2002). The mirror shape can be optimized to reduce the coupling losses for each fiber. The reported demonstration packed 32 fibers in a linear array, placing the input fiber at the array end and on the lens axis for studying the aberration limitations. Thus, the results are equivalent to a 62-fiber switch with 1D packing, or 3019 fibers using 2D fiber packing. Resulting fiber coupling losses ranged from 3 to 6 dB. Possible alternatives to address the aberrations from large count fiber switches include the use of a reduced cladding fiber, for minimizing the fiber bundle lateral extent that is dominated by the fiber's outer diameter, or the placement of a fiber collimator for each fiber. The advantage of placing collimators is that the resulting mode size on the MEMS mirror decreases, resulting in the fiber coupling being less mirror angle dependent.

An alternative approach to the construction of a $1 \times N$ switch is to cascade several 1×2 switches in a binary tree arrangement. The cascade can be from individual 1×2 switches, whereas in the case of moving waveguides the entire switch can be monolithically integrated onto a single wafer device (Ollier 2002).

3.07.4.4.3 Optical cross-connects ($N \times N$)

Large $N \times N$ fiber switches or OXCs were believed to be the best method of managing the exponentially growing traffic in optical networks at the height of the telecom bubble. The established core network vision foresaw these large switch fabrics routing WDM channels at network switching nodes, offering new wavelength services and providing shared protection services (Al-Salameh *et al.* 2002). However, sales of these large switches did not meet the high expectations. The large switches required an upfront investment for handling future traffic demand. Two main OXC designs were pursued by many companies. The first is a beam scanning switch using analog multistate mirrors, starting from sizes of 64×64 and growing to fiber port counts of 1296×1296. The second is a crossbar switch implementation with digital bistate mirrors, most of these switches were 8×8 in size, with some as large as 32×32. The smaller switch modules could be used in a Clos architecture to construct large switch equivalents. Yet even the smaller switch versions saw their viability wither with the downturn in service providers' infrastructure investment and as interest shifted to switching WDM channels using WSSs.

Small optical fiber switches of the crossbar design are planar in structure, with the input linear fiber array arranged at one facet of the switch and the output linear fiber array arranged at a second facet of the switch, at 90° to the first facet (**Figure 20**). The fiber arrays are mated to matching collimating lens arrays, such that the beams from the input and output fibers generate a regular grid. Each beam from an input fiber intersects every beam of the output fiber. For an input/output array of N fibers, there are thus N^2 such beam intersections. Switching from any input fiber to any output fiber in the crossbar design is accomplished by placing a mirror at the respective beam intersection position. Each of these mirrors is bistable; in one state it is not in the beam path, whereas in the second state it is at 45° and completes the connection. Several factors impact the design and performance of crossbar switch. There is no way to fine-tune the mirror position to maximize the coupling efficiency. Because each mirror will have an independent misalignment from an ideal setting, these switches have considerable loss variation for each possible input/output connection. Additionally, even with an ideal mirror position, an additional source of loss variation in the switch is the optical path length

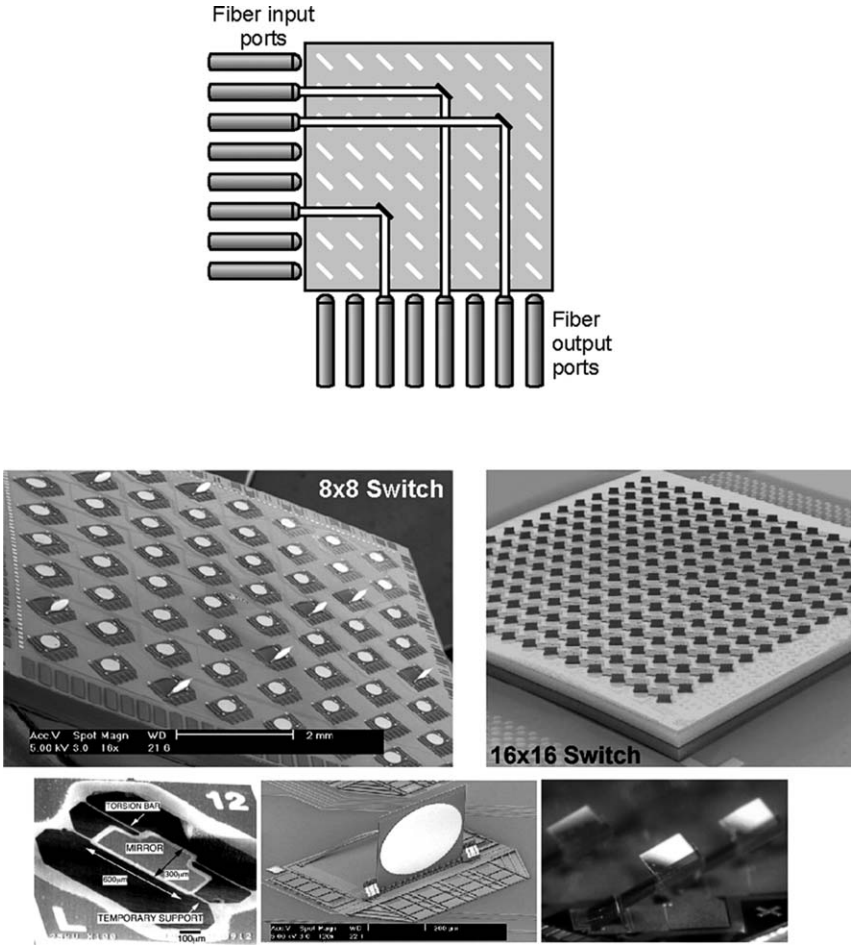


Figure 20 Planar crossbar $N \times N$ fiber switch. *Top*: Schematic representation. *Middle*: Image of 2D matrix of digital microelectromechanical systems (MEMS) mirrors. *Bottom*: Details of exemplary MEMS digital mirrors. Left mirror rotates 90° out of plane with electrostatic actuation, middle mirror rotates with scratch drive, and right mirror lifts with a seesaw actuator. (Source: Lin L Y, Goldstein E L, Tkach R W 1999 Free-space micromachined optical switches for optical networking. *IEEE J. Select. Top. Quant. Electron*. **5**, 4–9, © [1999] IEEE; Toshiyoshi H, Fujita H 1996 Electrostatic micro torsion mirrors for an optical switch matrix. *J. Microelectromech. Syst.* **5**, 231–7, © [1996] IEEE; Lin L Y, Goldstein E L, Tkach R W 1998 Free-space micromachined optical switches with submillisecond switching time for large-scale optical crossconnects. *IEEE Photon. Technol. Lett.* **10**, 525–7, © [1998] IEEE; De Dobbelaere P *et al.* 2002 Digital MEMS for optical switching. *IEEE Commun. Mag.* **40**, 88–95, © [2002] IEEE.)

differences for each possible connection, with the difference between the shortest beam path to the longest being twice the length of the fiber array. Finally, as the beam strikes the metalized surface of the mirror at 45°, the reflection adds some polarization-dependent loss (~0.4 dB). The optical path length of the crossbar switch scales with the product of the fiber count, N, and the fiber pitch, P. Placing a requirement on the beam size, w_0, to accommodate for the propagation distance, L, one finds that $L \propto \left(w_0^2/\lambda\right) \propto N \times P$. Because the fiber pitch is also linearly proportional to the beam size, we can conclude that $N \propto (w_0/\lambda)$, or that the collimated beam size also scales linearly with N, to accommodate for the longer propagation distance. The major implication of having the fiber array length as well as the fiber pitch scale linearly with N, is that the physical extent of the crossbar switch will scale with N^2, making the crossbar switch impractical for large fiber port counts due to wafer size limitations. An exemplary design of a 16×16 crossbar switch leads to a mirror diameter of 320 μm and a lateral length of 8 mm (Ryf *et al.* 2006). Doubling the fiber port count to a 32×32 switch size doubles the mirror size to 640 μm and quadruples the lateral length to 32 mm. One unique feature of the crossbar switch is the ability to utilize both sides of the micromirror for establishing a second

simultaneous connection (Lin *et al.* 1999). Such architecture has been proposed for implementing add-drop functionality with no additional hardware requirements (**Figure 21**).

The digital MEMS micromirrors of the crossbar switch are required to have large motion comparable to the beam size. Such a large motion is best achieved with long lever arms used to amplify the actuator's motion. The originally demonstrated crossbar switch used polysilicon torsion mirrors supported by long slender rods that enabled the mirrors to rotate 90° out of plane using electrostatic actuation and a mechanical stopper (Toshiyoshi and Fujita 1996). Later pop-up mirrors achieved 90° rotation from the substrate using a translation distance of 22 µm from a scratch drive actuator with submillisecond switching time (Lin *et al.* 1998). The insertion losses of these early research prototypes were 10 dB or higher. Subsequent commercialization of crossbar switches reduced the losses to 2–3 dB by using MEMS micromirrors with better pointing accuracy (Dobbelaere *et al.* 2002). The improvement is provided by decoupling the mirror angle, critical for efficient fiber coupling, from the mirror motion. This was achieved by utilizing mirror translation as opposed to rotation, using a seesaw structure.

Large optical fiber switches of the beam scanning design utilize the free-space volume of the switch, with an input 2D fiber array at one facet of the switch and an output 2D fiber array at the opposing facet (**Figure 22**). The fiber arrays are mated to matching collimating lens arrays, forming light beams from the input fibers onto the output fibers. To establish a connection from any input fiber to any output fiber, the beam from the input is directed to the desired output and then realigned to efficiently couple into the output fiber. Thus, every connection is required to have two beam scanning elements. Because there are N input and output fibers in an $N \times N$ switch, the number of actuators in the switch is $2N$, in comparison to N^2 of the crossbar switch. However, each actuator has to assume one of N possible positions for each switching state. Mirrors can be introduced to fold the propagation length between the input and output fiber arrays, resulting in a more compact switch form factor. When MEMS micromirrors perform the beam scanning, they can be part of the folding mechanism. The path length variation in the beam steering switch is relatively small, as the distance between the input and output fiber arrays, L, is much larger than the arrays' lateral extent. The fiber arrays' lateral extent scales as the product of the pitch, P, and the square root of port count, $\sqrt{N}$, due to the 2D array layout. The beam scanning range $\Delta\theta$ is proportional to the ratio of the lateral extent to the distance between the arrays, or $\Delta\theta \propto \left(P\sqrt{N}/L\right)$. Placing a requirement on the beam size, w_0, to accommodate for the propagation distance, L, one finds that $L \propto \left(w_0^2/\lambda\right)$. Because the pitch, P, is also proportional to w_0, by substitution we find the scaling of the beam scan range $\Delta\theta \propto \left(\sqrt{N}\lambda/P\right)$, of the

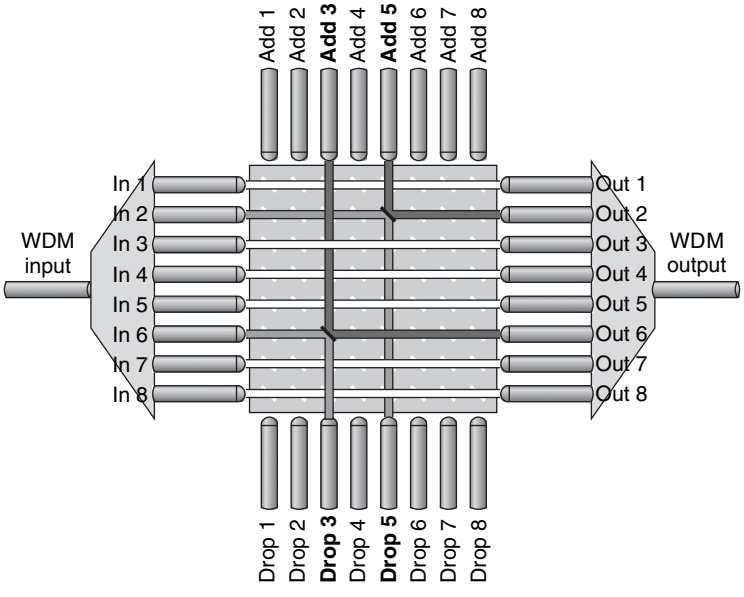


Figure 21 Use of planar crossbar switch for wavelength add-drop functionality.

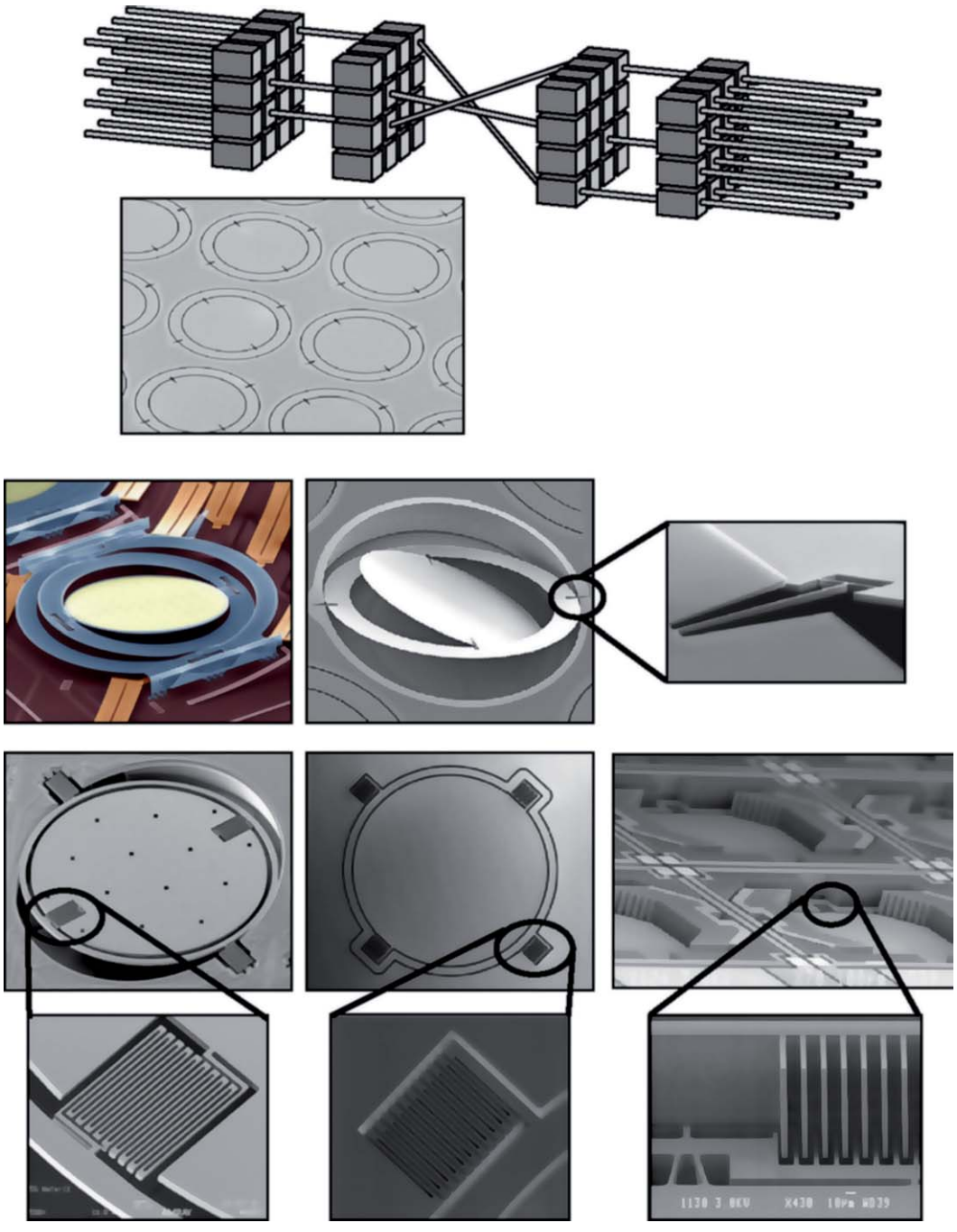


Figure 22 Three-dimensional $N \times N$ fiber switch *Top*: Schematic representation. *Middle*: Image of high fill factor 2D array of scanning microelectromechanical systems (MEMS) mirrors. *Bottom*: Details of exemplary MEMS scanning mirrors and their spring design detail. Top left mirror fabricated in surface micromachining process, other mirrors using silicon bulk processing. Spring design details show torsion rod, serpentine springs, and V-shaped mirror support, designed to provide the motion range, resonance frequencies, and in-plane stiffness. (Source: Aksyuk V A *et al.* 2003 Beam-steering micromirrors for large optical cross-connects. *IEEE J. Lightwave Technol.* **21**, 634–42, © [2003] IEEE; Chu P B *et al.* 2005 Design and nonlinear servo control of MEMS mirrors and their performance in a large port-count optical switch. *IEEE J. Microelectromech. Syst.* **14**, 261–73, © [2005] IEEE; Zheng X *et al.* 2003 Three-dimensional MEMS photonic cross-connect switch design and performance. *IEEE J. Select. Top. Quant. Electron.* **9**, 571–8, © [2003] IEEE; Yano M *et al.* 2003 Optical MEMS for photonic switching – compact and stable optical crossconnect switches for simple, fast, and flexible wavelength applications in recent photonic networks. *IEEE J. Select. Top. Quant. Electron.* **11**, 383–94, © [2003] IEEE.)

propagation distance $L \propto \left(N\lambda/(\Delta\theta)^2\right)$, and of the lateral dimension $P\sqrt{N}$. The physical extent of the switch scale more gracefully with the fiber count N; linearly in propagation distance and as the square root in the lateral dimension. An exemplary design of a 256×256 OXC with a 1 mm pitch requires a propagation distance of 86 mm, a lateral dimension of 16 mm × 16 mm, and a mirror tilt range of ~10° (Ryf *et al.* 2006). Additional optical elements, such as lenses or microprisms, can be introduced in the beam scanning switch to eliminate the tilt angle dependence on the spatial locations of the input and output fibers.

Several methods can perform the beam scanning operation. The most common is the use of two MEMS micromirrors in the beam path, in a folded Z-shaped configuration (Kim *et al.* 2003b, Neilson *et al.* 2004a, Ryf *et al.* 2006). Because the micromirrors are required to rotate about two axes, they are typically supported by a gimbal structure. The gimbal is not a working port of the mirror, hence, the mirror fill factor is reduced. The fill factor impacts the design as the fiber pitch has to increase beyond the limit imposed by the beam size. Moreover, the beam has to be well contained within the mirror's aperture to eliminate beam clipping losses. The MEMS micromirrors can be fabricated using surface micromachining processes of polysilicon or bulk processing of single-crystalline silicon. Typical mirror parameters are 500–800 µm diameter, ~100–200 V electrostatic actuation using three or four electrodes placed underneath the mirror, and 5–10 ms switching times (Aksyuk *et al.* 2003, Chu *et al.* 2005, Zheng *et al.* 2003). When properly designed, the losses of a 238×238 MEMS micromirror beam scanning OXC are in the 1–2 dB range. Loss contributors are the collimators' aberrations and antireflection coatings, and the micromirrors's reflectivity, curvature, and clipping. Other beam scanning methods are based on lateral displacement of the fiber or collimator with respect to each other (**Figure 23**). The collimator lens can be placed on an electrostatically actuated x–y stage

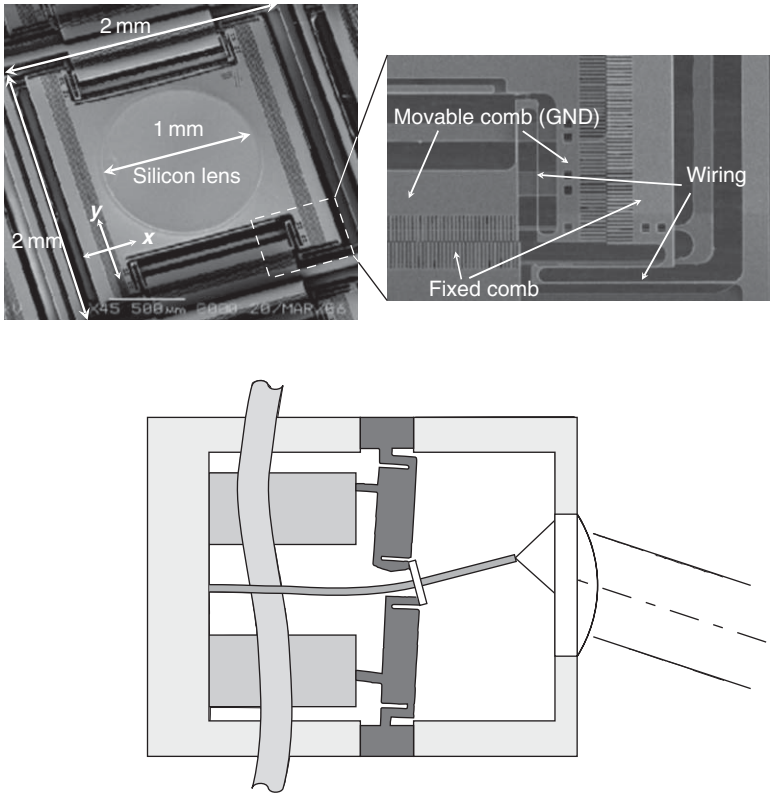


Figure 23 Alternative beam-scanning techniques for $N \times N$ fiber switch. *Top*: Lateral movement of the collimator lens. (Source: Takahashi K *et al.* 2006A 3D optical crossconnect using microlens scanner with topologically layer switching architecture. *IEEE Optical MEMS Conference*, Piscataway, NJ, USA, © [2006] IEEE.) *Bottom*: Lateral movement of the fiber tip. (Source: Hagood N W *et al.* 2004 Beam-steering optical switching apparatus. *US Pat.* 6 738 539, May 18.)

to achieve scanning (Takahashi *et al.* 2006, Toshiyoshi *et al.* 2003). However, the fill factor of the lens aperture to the fiber pitch is poor due to comb-drive actuation mechanism surrounding the lens, and the amount of lens displacement is limited. As an alternative, the fiber tip can be translated with respect to the lens (Hagood *et al.* 2004, Truex *et al.* 2003). A large displacement of the fiber tip is achieved by tilting the fiber from a pivot point farther away from the lens. Due to the relatively large forces required to bend an optical fiber, piezoelectric actuators are used. The optical path in this arrangement is extremely clear, comprised only of the two collimators. Measured optical losses of the connections average an impressive 1 dB.

3.07.4.5 MEMS Wavelength-Selective Switches

DWDM transmission has provided greatly enhanced data transport capacity on a single fiber, with today's state-of-the-art systems capable of carrying multiple terabits per second. The economic basis behind DWDM is the sharing of the fiber resources, components, and optical amplifiers across many wavelength channels. To maximize the efficiency of the transport layer, the data traffic should remain in the optical regime and undergo optical to electronic to optical (OEO) conversion only when necessary, which is either when the data in a particular channel has to be extracted for consumption or regenerated due to OSNR limitations. The system capability that is required for maximizing the efficiency is the ability to aggregate and disaggregate wavelength channel in the optical domain, allowing the remaining data paths to continue in the optical domain. This function, referred to as OADM, was initially implemented with an optical demultiplexer and multiplexer with a fiber patch panel disposed between them. The requirement to support remote provisioning and fast restoration functions in the OADM were addressed by the more flexible ROADM, which replaced the fiber patch panel with an optical fiber switch fabric. However, the architecture composed of a demultiplexer, switch, and multiplexer suffers from accumulated losses and channel passband filtering, which limit the potential reach of the transmission. A complete ROADM solution should provide wide, flat passband as well as dynamic power equalization across the channels for meeting the physical layer requirements.

WSSs are transparent optical switching subsystems, which can route and attenuate wavelength channel in meeting with ROADM requirements. A WSS contains some means of separating or demultiplexing the wavelength channels at the ingress ports, followed by an optical switch fabric and means of recombining or multiplexing back the wavelengths into one or more egress ports, following the switch function. There is no wavelength conversion or OEO inside a WSS; the switch is optically transparent for the photons carrying the data. The simplest WSS is a channel blocker, with a single input and output fiber, having the capability to power equalize or completely attenuate the WDM channels. The more capable $1 \times K$ WSS has a single input and K output fibers, adding the capability to independently route the individual WDM channels among the K fibers. Finally, the $K \times K$ wavelength-selective cross-connect (WSXC) handles K input and K output fibers with the ability to switch any wavelength from any input fiber to any output fiber, provided there is no wavelength contention. The drawback of the WSXC is its sensitivity to failures; if the switch malfunctions or needs to be replaced, then all the WDM traffic flowing on the K fibers is halted. The $1 \times K$ WSS module, which can be used to construct an equivalent $K \times K$ WSXC, is a more robust solution as the traffic is partitioned according to originating fiber port, and only a subset of the traffic will be affected by a single failure.

3.07.4.5.1 Channel blocker

The channel blocker is based on a free-space optical system that spectrally resolves the optical communication bandwidth, originally conceived for an amplifier gain flattening application (Ford and Walker 1998). However, for independent DWDM channel access, the optical system is designed to provide higher resolution (Neilson *et al.* 2004b). The light from an input/output fiber is collimated by a long focal length lens onto a diffraction grating, such that a large region of the grating is illuminated (**Figure 24**). The diffraction grating generates angular dispersion on account of the wavelength dependence of the diffraction angle. The diffracted light is refocused by the long focal length lens onto the same plane of the fiber tip, but displaced from it, and is incident onto a reflective pixelated modulator. The optical layout is identical to that of a spectrometer, which results in a spatially dispersed optical signal. The dispersive imaging arrangement demultiplexes the input DWDM channels in space, such that the information bandwidth of each wavelength channel is well contained within an active pixel

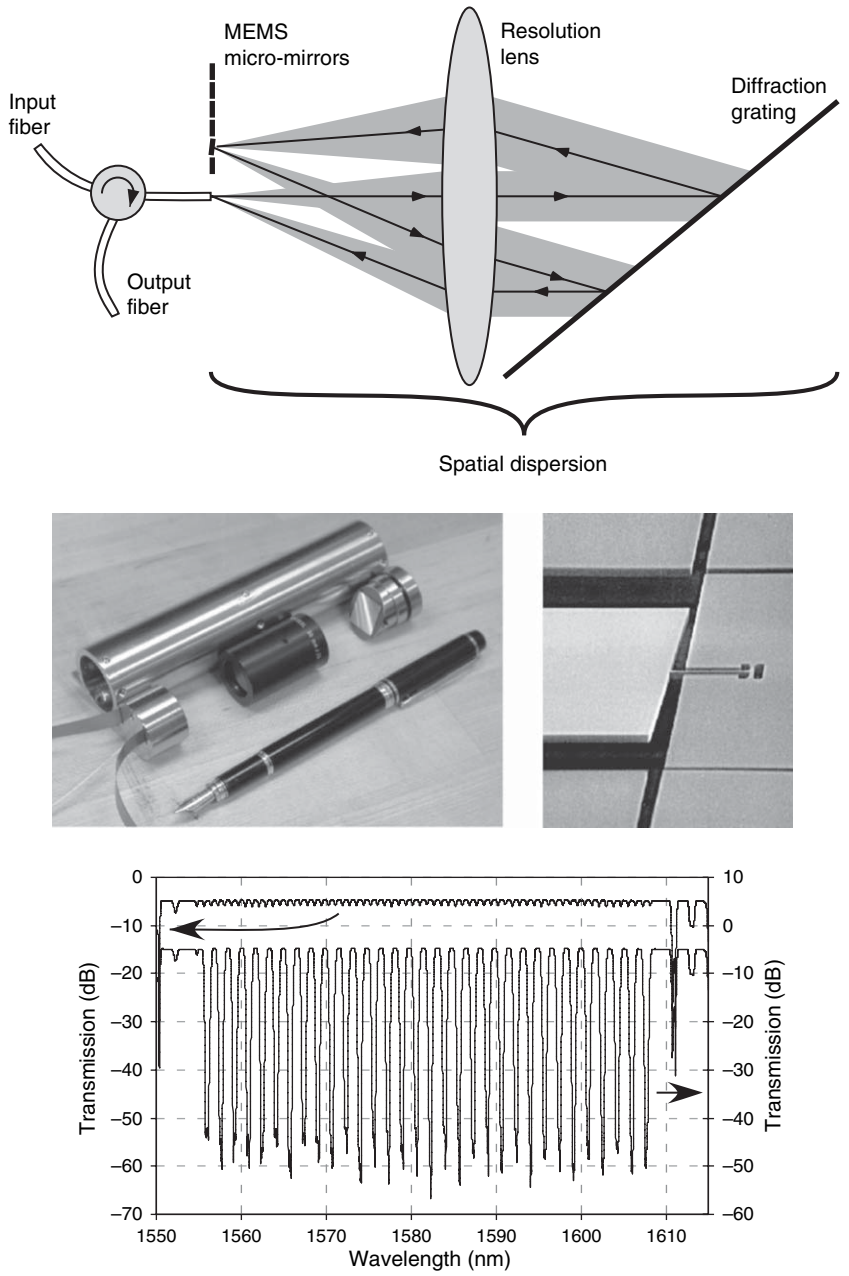


Figure 24 Wavelength blocker. *Top*: Schematic layout. *Middle*: Images of blocker package and microelectromechanical systems (MEMS) torsion mirror array. *Bottom*: Optical power spectrum of all wavelength division multiplexing (WDM) channels in through and alternating blocking patterns. (Source: Neilson D T *et al.* 2004b Channel equalization and blocking filter utilizing microelectromechanical mirrors. *IEEE Select. Top. Quant. Electron.* **10**, 563–9, © [2004] IEEE.)

element. The light is collected back from the active modulating elements and coupled into an output fiber after back propagating through the dispersive arrangement (reversing the original spatial dispersion). Ensuring that the mode size on the device array plane is less than the active pixel width yields the desired wide passband characteristics. This can be achieved by a judicious choice of the diffraction grating frequency and effective focal length of the imaging setup. Furthermore, a wide spatial margin between the incident beam and the pixel edges results in guard bands, which allow for drift in laser

source, optical system misalignment, and provides a more abrupt interchannel transition. The active pixelated device can be chosen from multiple available technologies, although LC modulators (Bouevitch *et al.* 2002, Rhee *et al.* 2001), MEMS tilting mirror arrays (Neison *et al.* 2004), and DMEMS (Ramani 2006) are often the technology of choice. The loss mechanism of each of these technologies is different; MEMS tilting mirrors change the propagation direction of the reflected beams, which introduces a loss when coupling back to the output single-mode fiber, LC attenuates the intensity of the light transmitted through the pixels by polarization rotation and extinction via a crossed polarization (reflective designs are also possible, without loss of generality), and DMEMS control the reflected intensity by adjusting the phase difference between reflective ribbons (with the remaining light diffracting outside the numerical aperture of the optical system).

A 100 GHz channel-spacing blocker based on an array of MEMS tilting micromirrors demonstrates the attractiveness of channel blockers for handling 64 DWDM channels (Neilson *et al.* 2004b). The channel passbands provided by the switch are very wide and flat, and the insertion loss of the switch is less than 5 dB. Channel blocking with greater than 40 dB attenuation was measured. More importantly, the 0.5 dB passband is 58 GHz wide, and the 3 dB bandwidth is 87 GHz. The measured chromatic dispersion within the passband was less than 2 ps nm^{-1}. Polarization-dependent losses were less than 0.35 dB. Such optical channel performance can handle signals with 10 and 40 Gb s^{-1} data rates with minimal impairment, even when the signal traverses several channel blockers in cascade. The MEMS micromirrors of the channel blocker were based on a single-crystal silicon fabrication process, patterned with a 248-nm deep ultraviolet (DUV) stepper to define submicrometer-wide springs and very small lateral gaps between the mirrors (0.4 μm). Torsion rods are used as springs and the 1-μm-thick mirrors have a tilt range of 8° at 150 V. The mirrors are actuated using electrostatic force from electrodes on an electrode chip placed below the mirrors chip. The mirror surfaces are aluminum coated to give high reflectively at 1.55 μm, and exhibit mirror resonant frequencies exceeding 20 KHz. MEMS-based channel blockers have not fared well in the marketplace, as LC-based modulators obtain similar performances with the LC modulating element and drive electronics costing a fraction of the MEMS solution.

3.07.4.5.2 Wavelength-selective $1 \times K$ switches

The $1 \times K$ WSS expands the free-space optical system of the channel blocker to allow for the introduction of $K+1$ fiber ports. The WSS design is based on two major subassemblies; the first used to spatially overlap the beams from the individual input and output fibers, to allow for switching between multiple ports, and the second to spectrally resolve the channels and introduce the wavelength selectivity (**Figure 25**). The role of the first subassembly is to image the input and K output optical fibers end faces onto an angular multiplexed common magnified spot. This subassembly converts the distinct spatial locations of the fibers to unique angular propagation directions. The second subassembly introduces the desired wavelength-selectivity property with the use of a diffraction grating. It spatially disperses the input magnified common spot, consisting of the N WDM channels, onto the MEMS micromirror array, such that each channel is imaged upon a separate mirror in the array for independent addressing. Each micromirror in the array can be tilted to a desired angle, which subsequently determines the selected output fiber array, on a WDM channel basis. The first subassembly determines the optical beam magnification ratio, the fiber array layout, and the required mirror tilts to reach each output fiber. The second subassembly determines the amount of spatial dispersion for separating the WDM channels and obtaining the necessary passband characteristic.

The optical subassembly responsible for imaging the optical fiber end faces onto a common magnified spot is comprised of a fiber array, a matching microlens array, and a condenser lens whose aperture subtends all the beam apertures from the fibers. The fiber array consists of $K+1$ fibers, where one fiber is assigned to carry the input signal and the remaining K fibers are the output fibers. The microlens array and fibers are coaxially aligned and result in collimated beams arranged in a 1D array to accommodate mirrors with a single tilt axis. Gaps are introduced between some of the lens apertures to allow for a variable attenuation feature by the beam scanning method without giving rise to crosstalk. The condenser lens focuses the $K+1$ collimated beams at its back focal plane, where the beams are superimposed and differentiated only in propagation direction. The optical arrangement of the first subassembly implements a telescopic imaging system of the fiber modes via the lenses from the microlens array and the condenser lens. The imaging operation magnifies the

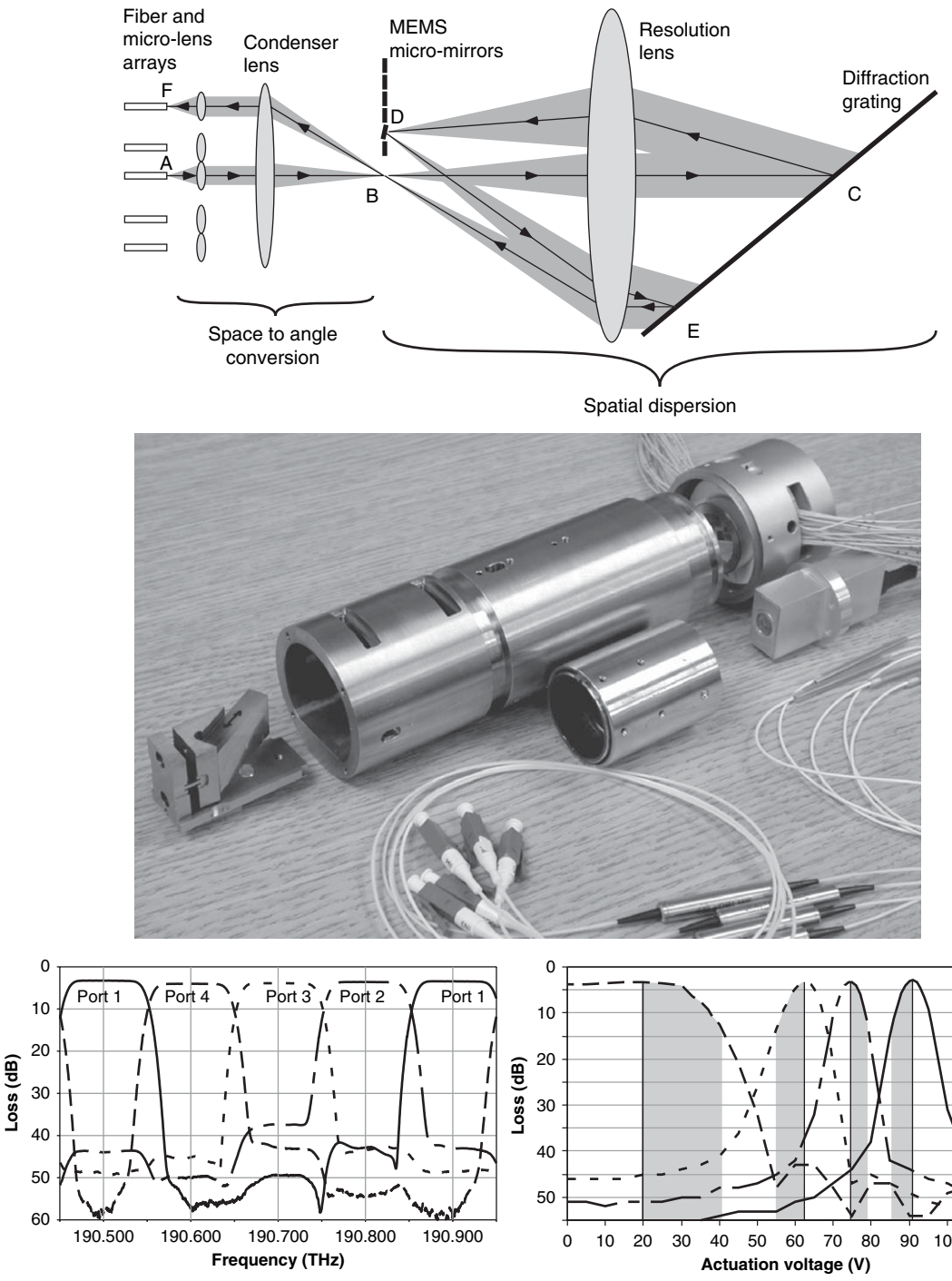


Figure 25 Wavelength-selective $1 \times K$ switch. *Top*: Schematic layout. *Middle*: Image of wavelength-selective switches (WSSs) package. *Bottom – Left*: Optical power spectrum detail of dense wavelength division multiplexing (DWDM) channels demonstrating switching to four output ports with wide passbands. *Bottom – Right*: Voltage versus coupling curve for one channel (mirror), scanning the four output ports with support for 10 dB dynamic range of attenuation (gray zones) and blocking. (Source: Marom D M *et al.* 2005a Wavelength-selective $1 \times K$ switches using free-space optics and MEMS micromirrors: Theory, design, and implementation. *IEEE J. Lightwave Technol.* **23**, 1620–30, © [2005] IEEE.)

optical beam emerging from the single-mode fiber by the ratio of the condenser lens to the microlens focal lengths. Anamorphic optics can be introduced into the optical subassembly to increase the number of fiber ports without increasing the magnified spot size in the dispersion direction, which sets the resolution requirements. Additionally, polarization diversity elements are integrated into the optics for use with high-efficiency diffraction gratings that only support one polarization.

The second optical subassembly disperses the magnified mode that was generated by the first subassembly and images it on the micromirror array. Its design is similar to a spectrometer with a Littrow mounted diffraction grating. A single lens collimates the light that is then incident on the grating. The diffracted light, which is propagating back toward the lens and is angularly dispersed, is imaged by the same lens onto the micromirror array. As is well known from classical spectrography, the spectral resolution of the instrument increases with increasing focal length and grating frequency, and with decreasing input slit size. In the WSS, the magnified mode size is equivalent to a spectrograph slit size. The spectral performance of the free-space WSS is directly related to the dimensionless ratio ξ, which is defined as the ratio of MEMS micromirror size to the magnified spot mode size (Marom *et al.* 2005). All the optical system parameters of the WSS contribute to this dimensionless ratio; the focal lengths of the lenses, the grating frequency, and the fiber mode. The ratio measures how well the Gaussian mode is confined within the micromirror. The channel passband performance of the WSS is derived by solving the frequency-dependent, power-coupling efficiency integral. Using a simplifying assumption that the MEMS array has a 100% fill factor, the frequency-dependent coupling efficiency is defined by

$$|\eta(\nu)|^2 = \frac{1}{4}\left\{\mathrm{erf}\left[\sqrt{2}\xi\left(1-\frac{2\nu}{\nu_{\mathrm{ch}}}\right)\right] + \mathrm{erf}\left[\sqrt{2}\xi\left(1+\frac{2\nu}{\nu_{\mathrm{ch}}}\right)\right]\right\}^2$$

where ν is the temporal frequency variable, ν_{ch} is the DWDM channel separation, and erf is the error function. Thus, passband requirements can be accommodated in the design of the WSS by varying the ratio of the mirror size to the magnified Gaussian mode size, which sets ξ, in fulfillment of the optical transport requirements.

The first reported WSS was configured as a 1×4 switch with support for 128 channels on a 50-GHz grid (Marom *et al.* 2002). The switch had a worst-case insertion loss of 5 dB, with 1 dB passband width of 37 GHz, and 3 dB width of 44.5 GHz. The broad flat-top filter shape enables the switch to be cascaded in transparent systems carrying $10\,\mathrm{Gb\,s^{-1}}$ data rates. Further WSS development was focused on 1×4 configuration, based on a 100-GHz grid, with support for 64 channels and $40\,\mathrm{Gb\,s^{-1}}$ data rates, along with 10 dB spectral equalization dynamic range (Marom *et al.* 2003a). These later WSSs were compact and fully packaged, had a 4 dB insertion loss, 77 GHz wide passbands at the -1 dB level, polarization-dependent losses below 0.3 dB, and chromatic dispersion within the passband below $10\,\mathrm{ps\,nm^{-1}}$. Two different MEMS micromirror arrays were designed for the switch; one based on surface micromachining of polysilicon and the other on bulk processing of a SOI wafer (Greywall *et al.* 2003, López *et al.* 2002). In the polysilicon approach, a double-hinge activation mechanism is defined. An actuation plate, anchored at one edge, is tilted to small angles via an underlying parallel plate electrode and a 4 μm gap. The mirror is attached to the free edge of the actuation plate and to the substrate with unequal arm lengths, allowing large mirror tilt angles out of plane via angle amplification. The mirror, actuator plate, and springs are etched in 1.5-μm-thick polysilicon. Spring features are typically 0.5 μm wide, and the gap between adjacent mirrors is 0.7 (~99% fill ratio). In the SOI approach, a 10-μm-thick layer of polysilicon is deposited over the patterned 1-μm-thick single-crystalline silicon and is used to define the actuator electrodes and ground shields. The electrode attracts the short actuator arm via an electrostatic fringe field, resulting in mirror rotation out of plane about the torsion springs. The monolithic structure has a further advantage that it does not exhibit rotational snap-down. Both the polysilicon and SOI micromirrors achieved 6.5° mechanical rotation at a driving voltage of 100 V, and had approximately millisecond scale switching times. Several other free-space WSS and 1D MEMS mirror array efforts have been reported, with similar functionality to that described above (Bernstein *et al.* 2004, Ducellier *et al.* 2002, Marom *et al.* 2006, Tsuboi *et al.* 2004, Wu *et al.* 2002). When using a mirror array with a single rotation axis, transient effects might occur on intermediate fiber ports as the beam is scanned along it (**Figure 26**). To achieve hitless switching, mirrors with two rotation axes have been developed, so that the beam can be first moved off the array, then to the position closest to the desired fiber, and finally moved back to couple only to the desired port. This

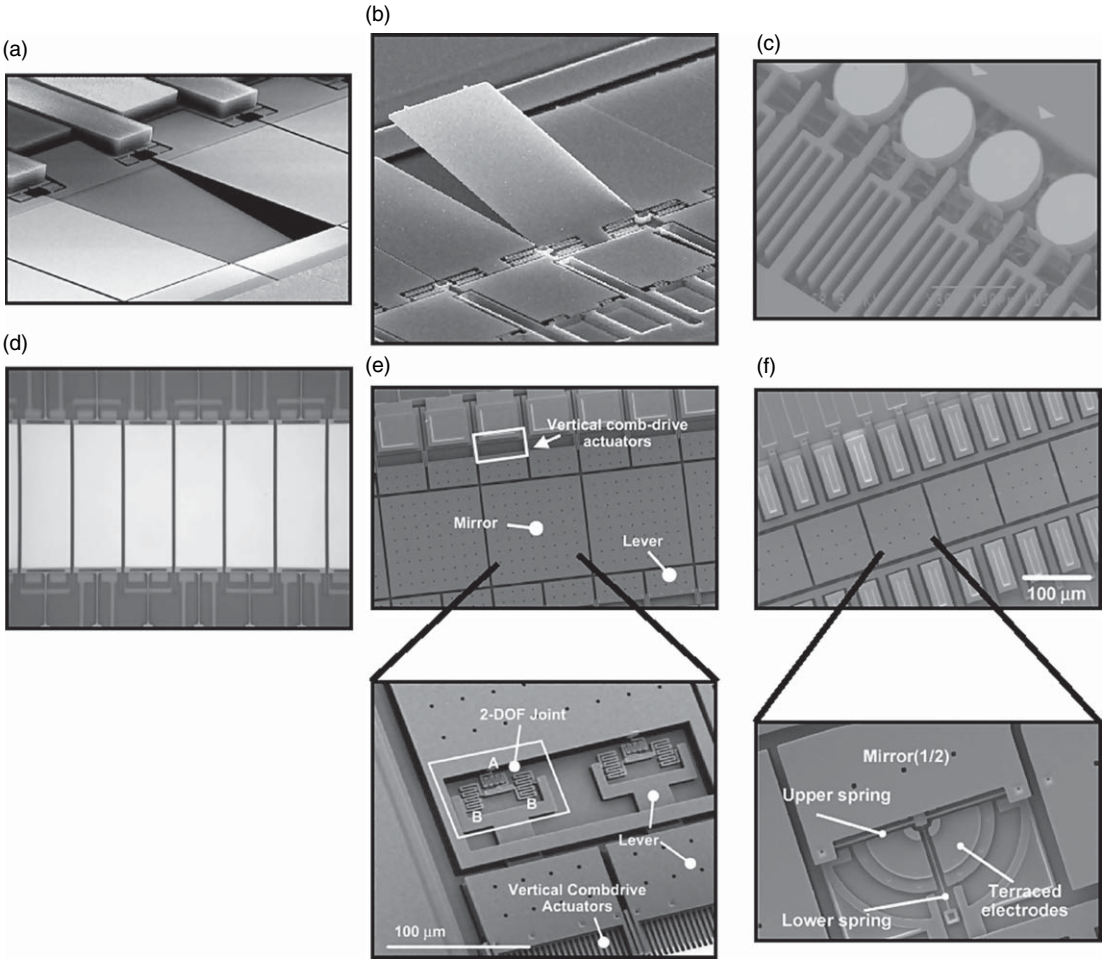


Figure 26 One-dimensional, high fill factor, microelectromechanical systems (MEMS) micromirror arrays for wavelength-selective switching applications. Mirrors can tilt preferentially with rotation axis parallel to dispersion (a–c), or normal to dispersion (d), or in both axis with underlying actuators (e, f) for hitless operation. (Source: López D *et al.* 2002 Monolithic MEMS optical switch with amplified out-of-plane angular motion. IEEE Optical MEMS conference, Lugano, Switzerland, © [2002] IEEE; Greywall D S *et al.* 2003 Monolithic, fringe-field-activated crystalline silicon tilting mirror devices. *IEEE J. Microelectromech. Syst.* **12**, 702–7, © [2003] IEEE; Tsuboi O *et al.* 2004 A high-speed comb-driven micromirror array for $1 \times N$ 80-channel wavelength selective switches. *IEEE Int. Conference on Optical MEMS*, Takamatsu, Japan, © [2004] IEEE; Bernstein J J *et al.* 2004 MEMS tilt-mirror spatial light modulator for a dynamic spectral equalizer. *IEEE J. Microelectromech. Syst.* **13**, 272–8, © [2004] IEEE; Tsai J-c, Wu M C 2005 Gimbal-less MEMS two-axis optical scanner array with high fill-factor. *IEEE J. Microelectromech. Syst.* **14**, 1323–8, © [2005] IEEE; Tsai J-c, Wu M 2006a A high post-count wavelength-selective switch using a large scan-angle, high fill-factor, two-axis MEMS scanner array. *IEEE Photon. Technol. Lett.* **18**, 1439–41, © [2006] IEEE.)

poses a design challenge as the mirror array is still required to fulfill the high fill ratio requirement. Several such two-axis micromirror implementations have been reported (Kouma *et al.* 2005, Tsai and Wu 2005, 2006a). Furthermore, with two-axis mirror technology, the number of fiber ports K served by the WSS can be increased by arranging the fibers in a 2D array (Tsai and Wu 2006b, Tsi *et al.* 2006).

While the free-space WSS arrangement performs very well, its size and packaging of discrete optical elements is challenging from an optomechanical point of view. Many of the optical elements can be eliminated by utilizing a hybrid planar lightwave circuit (PLC) and free-space optics arrangement (Marom *et al.* 2004, 2005b, Ducellier *et al.* 2004). The input and output fibers of the WSS are attached to the PLC, which distributes the light into an array of waveguides with constantly increasing path lengths. The light radiating out of the waveguide array is angularly dispersed, and with a single lens is converted to spatial

dispersion that is incident onto the MEMS micromirror array. Reflected light is incident into a desired waveguide array on the PLC for coherently recombining the light into the output fiber. Introducing the PLC platform into the WSS enables to customize each waveguide array, and in one example introduce a demultiplexer directly into the WSS. Other approaches attempt to combine the entire WSS functionality – consisting of the demultiplexing, switching with MEMS micromirrors, and remultiplexing – onto a single PLC (Fuchs 2004, Chi *et al.* 2005).

3.07.4.5.3 Wavelength-selective $K \times K$ switches

The optical switching requirement at network nodes originates from K fibers, carrying N DWDM channels. The number of individually switched channels is therefore $K \times N$, which can quickly scale to many hundreds of channels. For precisely that reason, large port count OXCs were developed. However, the switching functionality can also be provided by using $2K$ modules of $1 \times K$ WSS (or even K modules as in **Figure 27**) (Marom *et al.* 2003b). This elegant solution provides all the switching functionality carried by $K(K-1)$ interconnecting fibers between the WSS modules. If we assume that four fibers are introduced, $K=4$, and each fiber carries 80 channels, $N=80$, then all the traffic is routed on 12 fibers using four WSS modules. The OXC alternative requires $K \times N = 320$ fibers in both input and output sides, a fiber management nightmare! The entire architecture can even be implemented in planar optics with specially integrated free-space tilting micromirrors on a single silicon wafer (Chi *et al.* 2006). Another alternative is the wavelength-selective $K \times K$ switch, or WSXC. The WSXC internally handles the demultiplexing, switching, and

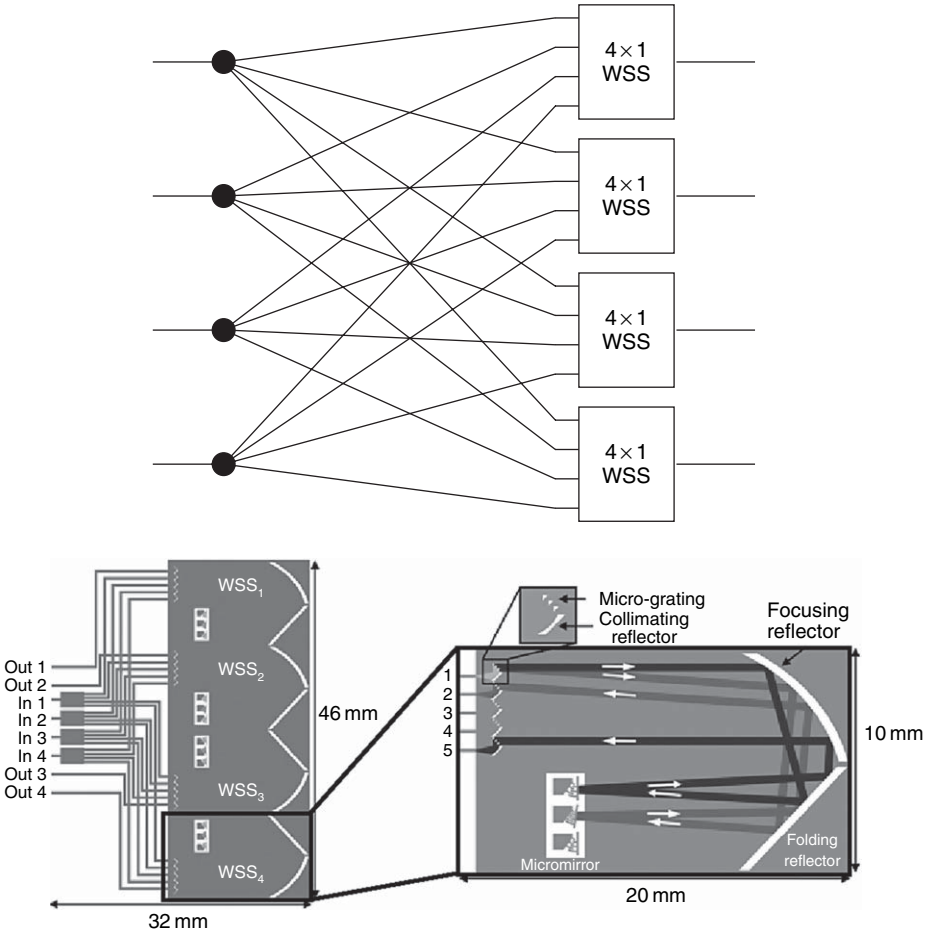


Figure 27 *Top*: Wavelength-selective $K \times K$ cross-connect architecture comprising individual $1 \times K$ wavelength-selective switches (WSSs) in a broadcast and select architecture. *Bottom*: Planar implementation of 4×4 wavelength-selective cross-connect architecture on a single silicon wafer. (Source: Chi C H *et al.* Silicon-based, monolithic, 4×4 wavelength-selective switch with on-chip micromirrors. *Optical Fiber Conference (OFC 2006)*, Anaheim, CA, March 2006, © [2006] IEEE.)

remultiplexing of all K input/output fibers, further reducing the fiber management complexity, at the expense of complex switch architecture.

A demonstrated WSXC introduced PLCs with demultiplexer and multiplexer functionality into a transparent OXC fabric (Ryf *et al.* 2001). The custom PLC placed the demultiplexed waveguides on a pitch $P = 1.25$ mm matched to a microlens array and the micromirror array. The collimated beams were incident onto the micromirrors for switching, much the same way as the collimated beams a from a fiber array are generated. The passband performance of the WSXC is dictated by the demultiplexer/multiplexer implementation, which are typically not as wide and flat as those achieved by regular WSS. Thus, cascadability using this solution is questionable. A reported alternative performs the demultiplexing with a free-space grating, and selects the bandwidth for switching by the MEMS micromirror array (Solgaard *et al.* 2002). With this solution, the bandwidth is defined by the micromirror array fill factor and shape, much the same as for the WSS. However, there is little information on any implementation attempts of this WSXC.

3.07.5 Conclusions

Optical microsystems with MEMS actuators can offer the functionality, flexibility, low power consumption, low insertion loss, and compactness required for optical communication networks. Reported optical microsystems have demonstrated most of the passive and slow active enabling elements for optical networking, such as ROADM, fiber cross-connect switching, channel selection filters for monitoring, and tunable dispersion compensating elements. The maturation of MEMS technology fortuitously coincided with, or even helped to enable, the transition to optical networking, by demonstrating the feasibility of performing the required tasks for reconfigurable networks. This has led to intensive research and development efforts in the last 10 years, by established corporate research centers, academia, and start-up companies. Unfortunately, many of these efforts have since ceased, as the forecasted market demand never materialized and consolidation and investor pull-out resulted in company shut downs. This downturn came as a predictable counter response to the hype and overselling by the investment community, leading to the infamous telecom bubble of 1999–2001.

While most of the private and institutional capital poured into telecom microsystem development may have been written off, the industry effort helped to further advance microoptics, microfabrication, packaging, and related fields. This has benefited the related microsystem scientific and industrial communities, with innovation sprouting out in sensing, bioengineering, and nanotechanology applications. Moreover, telecom industry spending is now experiencing steady growth once again on a more rational base spending level, due to a recovery of the industry (**Figure 28**), with the strongest growth in the metropolitan and fiber access spaces. Companies that have survived the downturn are growing once again, and there are signs of a resurgence of start-up activity. The renewed interest and more stable and sustainable marketplace ensure that optical communication networks will continue to serve as a driving force for innovations in optical microsystems.

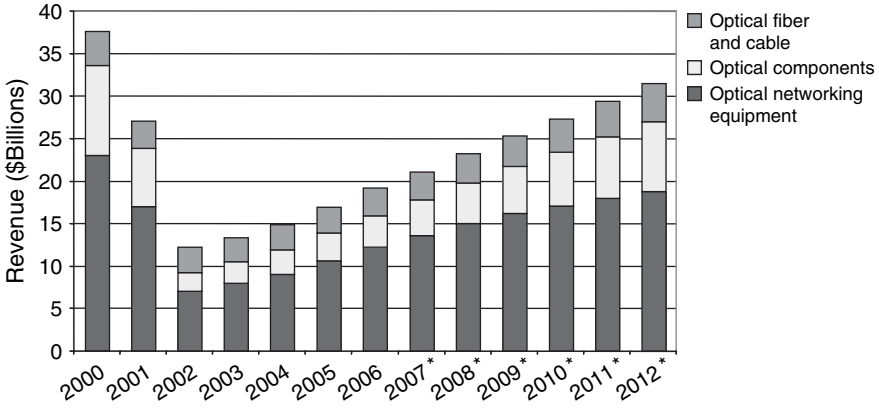


Figure 28 Global communications revenue and forecast by segment, 2000–2012. (Source: OIDA Optoelectronics Industry Development Association.)

References

Aksyuk V A, Giles C R, Dentai A, Burrus E C, Burrus C A, Stulz L, Bishop D 1999 Optically-powered optical power limiter for use in lightwave networks. *Micro Electro Mechanical Systems (MEMS)*, Orlando, FL, USA

Aksyuk V A, Pardo F, Carr D, Greywall D, Chan H B, Simon M E, Gasparyan A, Shea H, Lifton V, Bolle C, Arney S, Frahm R, Paczkowski M, Haueis M, Ryf R, Neilson D T, Kim J, Giles C R, Bishop D 2003 Beam-steering micromirrors for large optical cross-connects. *IEEE J. Lightwave Technol.* **21**, 634–642

Al-Salameh D Y, Korotky S K, Levy D S, Murphy T O, Patel S H, Richards G W, Tentarelli E S 2002 Optical switching in transport networks: Applications, requirements, architecture, technologies, and solutions. In: Kaminow I and Li T (eds.) *Optical Fiber Telecommunications IV A. Chap. 7*. Academic Press, San Diego, CA

Amano T, Koyama F, Hino T, Arai M, Mastutani A 2000 Multiple wavelength micromachined GaAlAs/GaAs vertical cavity filter array. *IEEE Optical MEMS Conference*, Kauai, HI, USA

Angulo Barrios C, Almeida V R, Panepucci R R, Schmidt B S, Lipson M 2004 Compact silicon tunable Fabry-Pérot resonator with low power consumption. *IEEE Photon. Technol. Lett.* **16**, 506–508

Anthon D, Berger J D, Tselikov A 2004a C + L band MEMS tunable external cavity semiconductor laser. *Optical Fiber Conference (OFC)*, Los Angeles, CA, USA,paper WL2

Anthon D, King D, Berger J D, Dutta S, Tselikov A 2004b Mode-hop free sweep tuning of a MEMS tuned external cavity semiconductor laser. *Conference on Lasers and Electro-Optics (CLEO)*, San Francisco, CA, USA

Bakke T, Tigges C P, Lean J J, Sullivan C T, Spahn O B 2002 Planar microoptomechanical waveguide switches. *IEEE J. Select. Top. Quant. Electron.* **8**, 64–72

Barber B, Giles C R, Askyuk V, Ruel R, Stulz L, Bishop D 1998 A fiber connectorized MEMS variable optical attenuator. *IEEE Photon. Technol. Lett.* **10**, 1262–1264

Bashir A, Katila P, Ogier N, Saadany B, Khalil D A 2004A MEMS-based VOA with very low PDL. *IEEE Photon. Technol. Lett.* **16**, 1047–1049

Belikov R, Solgaard O 2003 Optical wavelength filtering by diffraction from a surface relief. *Opt. Lett.* **28**, 447–449

Berger J D, Zhang Y, Grade J D, Lee H, Hrinya S, Jerman H 2001 Widely tunable external cavity diode laser using a MEMS electrostatic rotary actuator. *Optical Fiber Conference (OFC)*, Anaheim, CA, USA,paper TuJ2

Berger J D, Ilkov F, King D, Anthon D 2003 Widely tunable, narrow optical bandpass Gaussian filter using a silicon microactuator. *Optical Fiber Conference (OFC)*, Atlanta, GA, USA ,paper TuN2

Berger J, Anthon D, Drake J, Ilkov F, Shreeram J, Verdonck C 2004 MEMS-tunable 10 Gb/s APD receiver for broadcast and select CATV networks. *Optical Fiber Conference (OFC)*, Los Angeles, CA, USA ,paper ThQ3

Bernstein J J, Dokmeci M R, Kirkos G, Osenar A B, Peanasky J, Pareek A 2004 MEMS tilt-mirror spatial light modulator for a dynamic spectral equalizer. *IEEE J. Microelectromech. Syst.* **13**, 272–8

Blake M 2003 C-MEMS filter aims at monitoring. *EE Times*, March 24

Bonenfant P A, Jones M L 2004 OFC 2003 workshop on wavelength selective switching based optical networks. *IEEE J.* **22**, 305–9

Bouevitch O, Touahri D, Morgan J P, Panteleev S, Reimer C 2002 Channel-power equalizer and dynamic gain equalizer based on the optical bench platform. *2002 Dig. IEEE/LEOS Summer Topical Meetings*, Mont Tremblant, QC, Canada, pp. MD1-4–5

Chang-Hasnain C J 2000 Tunable VCSEL. *IEEE J. Select. Top. Quant. Electron.* **6**, 978–87

Cheng C, Lee C, Yeh J A 2004 Retro-reflection type MOEMS VOA. *IEEE Photon. Technol. Lett.* **16**, 2290–2

Chi C-H, Tsai J-C, Lee M C M, Hah D, Wu M C 2005 Integrated 1×4 wavelength-selective switch with on-chip MEMS micromirrors. *Conference Quantum Electronics and Laser Science (QELS)*, Baltimore, MD, USA

Chi C-H, Tsai J-C, Hah D, Mathai S, Lee M C M, Wu M C 2006 Silicon-based, monolithic, 4×4 wavelength-selective switch with on-chip micromirrors. *Optical Fiber Conference (OFC 2006)*, Anaheim, CA, March 2006

Chu P B, Brener I, Chuan P, Lee S-S, Dadap J I, Park S, Bergman K, Bonadeo N H, Chau T, Chou M, Doran R A, Gibson R, Harel R, Johnson J J, Lee C D, Peale D R, Tang B, Tong D T K, Tsai M J, Wu Q, Zhong W, Goldstein E L, Lin L Y, Walker J A 2005 Design and nonlinear servo control of MEMS mirrors and their performance in a large port-count optical switch. *IEEE J. Microelectromech. Syst.* **14**, 261–73

Coldren L A, Fish G A, Akulova Y, Barton J S, Johansson L, Coldren C W 2004 Tunable semiconductor lasers: A tutorial. *IEEE J. Lightwave Technol.* **22**, 193–202

Daleiden J, Irmer S, Rangelov V, Romer F, Tarraf A, Prott C, Strassner M, Hillmer H 2002 Record wavelength tuning of 127 nm for vertical cavity Fabry–Perot filter. *IEEE Optical MEMS Conference*, Lugano, Switzerland

Darcie T E, Driessen P F, Osusky M, Lin W 2005 Optical network control overlay using silicon VOA arrays. *IEEE Photon. Technol. Lett.* **17**, 513–15

De Dobbelaere P, Falta K, Gloeckner S, Patra S 2002 Digital MEMS for optical switching. *IEEE Commun. Mag.* **40**, 88–95

Diemeer M B J, Dekker R 2002 MEMS VOA with polymeric thermal microactuators. *Eur. Conf. Optical Communication (ECOC)*, Copenhagen, Denmark

Dono N R, Green P E Jr., Liu K, Ramaswami R, Tong F F-K 1990 A wavelength division multiple access network for computer communication. *IEEE J. Select. Areas Commun.* **8**, 983–94

Ducellier T, Bismuth J, Roux S F, Gillet A, Merchant C, Miller M, Mala M, Ma Y, Tay L, Sibille J, Alavanja M, Deren A, Cugalj M, Ivancevic D, Dhuler V, Hill E, Cowen A, Shen B, Wood R 2002 The MWS 1×4: A high performance wavelength switching building block. *Eur. Conf. Optical Communications (ECOC)*, Copenhagen, Denmark

Ducellier T, Hnatiw A, Mala M, Shaw S, Mank A, Touahri D, McMullin D, Zami T, Lavigne B, Peloso P, Leclerc O 2004 Novel, high-performance, hybrid waveguide-MEMS, 1×9 wavelength-selective switch in a 32 cascade loop experiment. *Eur. Conf. Optical Communication (ECOC)*, Stockholm, Sweden

Feldman R D, Harstead E E, Jiang S, Wood T H, Zirngibl M 1998 An evaluation of architectures incorporating wavelength division multiplexing for broad-band fiber access. *IEEE J. Lightwave Technol.* **16**, 1546–59

Field L A, Burriesci D L, Robrish P R, Ruby R C 1995 Micromachined 1×2 optical fiber switch. *Proc. Tech. Dig. 8th Int. Conf. Solid-State Sensors Actuators (Transducers '95).* Stockholm, Sweden, Vol. 1, pp. 344–7

Finot M, McDonald M, Daiber A, Chapman W B, Li D, Epitaux M, Zbinden E, Bennet J, Kozlovsky W J, Verdiell J-M 2004 Automated optical packaging technology for 10 Gb/s transceivers and its application to a low-cost full C-band tunable transmitter. *Intel Technol. J.* **8**, 101–14

Fishman D A, Correa D L, Goode E H, Downs T L, Ho A Y, Hale A, Hofmann P, Basch B, Gringeri S 2006 The rollout of optical networking: LambdaXtreme national network deployment. *Bell Labs Tech. J.* **11**, 55–63

Ford J E, DiGiovanni D J 1998 $1 \times N$ fiber bundle scanning switch. *IEEE Photon. Technol. Lett.* **10**, 967–9

Ford J E, Walker J A 1998 Dynamic spectral power equalization using micro-opto-mechanics. *IEEE Photon. Technol. Lett.* **10**, 1440–2

Ford J E, Walker J A, Greywall D S, Goossen K W 1998 Micromechanical fiber-optic attenuator with 3 μs response. *IEEE J. Lightwave Technol.* **16**, 1663–70

Fuchs D T, Doerr C R, Aksyuk V A, Simon M E, Stulz L W, Chandrasekhar S, Buhl L L, Cappuzzo M, Gomez L, Wong-Foy A, Laskowski E, Chen E, Pafchek R 2004 A hybrid MEMS-waveguide wavelength selective cross connect. *IEEE Photon. Technol. Lett.* **16**, 99–101

Giles C R, Aksyuk V, Barber B, Ruel R, Stulz L, Bishop D 1999 A silicon MEMS optical switch attenuator and its use in lightwave subsystems. *IEEE J. Select. Top. Quant. Electron.* **5**, 18–25

Godil A 2002 Diffractive MEMS technology offers a new platform for optical networks. *Laser Focus World* **38**, 181–5

Greywall D S, Pai C-S, Oh S-H, Chang C-P, Marom D M, Busch P A, Cirelli R A, Taylor J A, Klemens F P, Sorsch T W, Bower J E, Lai W Y-C, Soh H T 2003 Monolithic, fringe-field-activated crystalline silicon tilting mirror devices. *IEEE J. Microelectromech. Syst.* **12**, 702–707

Hagood N W, Ghandi K, Cloutier B, Borgen M 2004 Beam-steering optical switching apparatus. US Pat. 6 738 539, May 18

Herding M, Richardt F, Woias P 2003 A novel approach to low-cost optical fiber switches. *IEEE Optical MEMS Conference*, Waikoloa, HI, USA

Hirata T, Mitama I, Abe M, Makita K, Shiba K, Hane K, Sasaki M 2005 Development of MEMS-based optical surge suppressor. *Optical Fiber Communication Conference (OFC)*, Anaheim, CA, USA

Hoffmann M, Kopka P, Voges E 1999 All-silicon bistable micromachincal fiber switch based on advanced bulk micromachining. *IEEE J. Select. Top. Quant. Electron.* **5**, 46–51

Hohlfeld D, Zappe H 2005 Thermally tunable optical thin-film filters with sub-nanometer resolution, 41.7 nm continuous tuning range. *IEEE MEMS*, Miami, FL, USA

Isamoto K, Kato K, Morosawa A, Chong C, Fujita H, Toshiyoshi H 2004 A 5-V operated MEMS variable optical attenuator by SOI bulk micromachining. *IEEE J. Select. Top. Quant. Electron.* **10**, 570–8

Kaminow I P, Doerr C R, Dragone C, Koch T, Koren U, Saleh A A M, Kirby A J, Ozveren C M, Schofield B, Thomas R E, Barry R A, Castagnozzi D M, Chan V W S, Hemenway B R Jr., Marquis D, Parikh S A, Stevens M L, Swanson E A, Finn S G, Gallager R G 1996 A wideband all-optical WDM network. *IEEE J. Select. Areas Commun.* **14**, 780–799

Kanamori Y, Aoki Y, Sasaki M, Hosoya H, Wada A, Hane K 2003 Driving of optical fiber by surface-micromachined cam-micromotor for the applications to variable optical attenuator. *IEEE Optical MEMS Conference*, Waikoloa, HI, USA

Kanamori Y, Kitani T, Hane K 2006 Movable guided-mode resonant grating filters by four bimorph actuators for wavelength selective dynamic reflection control. *IEEE Optical MEMS Conference*, Piscataway, NJ, USA

Kim C-K, Lee M-L, Jun C-H, kim Y T 2003a 47 nm tuning of thermally actuated Fabry–Perot tunable filter with very low power consumption. *IEEE Optical MEMS Conference*, Waikoloa, HI, USA

Kim J, Nuzman C J, Kumar B, Lieuwen D F, Kraus J S, Weiss A, Lichtenwalner C P, Papazian A R, Frahm R E, Basavanhally N R, Ramsey D A, Aksyuk V A, Pardo F, Simon M E, Lifton V, Chan H B, Haueis M, Gasparyan A, Shea H R, Arney S, Bolle C A, Kolodner P R, Ryf R, Neilson D T, Gates J V 2003b 1100/spl times/1100 port MEMS-based optical crossconnect with 4-dB maximum loss. *IEEE Photon. Technol. Lett.* **15**, 1537–9

Kimura T, Fujimura N, Kanbara N 2004 Wide range tunable filter module based on MEMS tilt mirror. *IEEE Optical MEMS Conference*, Takamatsu, Japan

Kner P, Kageyama T, Boucart J, Stone R, Sun D, Nabiev R F, Pathak R, Yuen W 2003 A long-wavelength MEMS tunable VCSEL incorporating a tunnel junction. *IEEE Photon. Technol. Lett.* **15**, 1183–5

Kouma N, Tsuboi O, Soneda H, Ueda S, Sawaki I 2005 Fishbone-shaped vertical comb actuator for dual axis 1-D analog micromirror array. *Solid State Sensors, Actuators, and Microsystems (Transducers '05)*. Seoul, Korea

Lee M-C M, Wu M.C. 2005 MEMS-actuated microdisk resonators with variable power coupling ratios. *IEEE Photon. Technol. Lett.* **17**, 1034–6

Li M Y, Yuen W, Li G S, Chang-Hasnain C J 1998 Top-emitting micromechanical VCSEL with a 31.6-nm tuning range. *IEEE Photon. Technol. Lett.* **10**, 18–20

Li X, Antoine C, Daesung L, Jen-Shiang W, Solgaard O 2006 Tunable blazed gratings. *IEEE J. Microelectromech. Sys.* **15**, 597–604

Lin L Y, Goldstein E L, Tkach R W 1998 Free-space micromachined optical switches with submillisecond switching time for large-scale optical crossconnects. *IEEE Photon. Technol. Lett.* **10**, 525–7

Lin L Y, Goldstein E L, Tkach R W 1999 Free-space micromachined optical switches for optical networking. *IEEE J. Select. Top. Quant. Electron.* **5**, 4–9

Lipson A, Yeatman E M 2005 Free-space MEMS tunable optical filter on (110) silicon. *IEEE Optical MEMS Conference*, Oulu, Finland

Littman M G, Metcalf H J 1978 Spectrally narrow pulsed dye laser without beam expander. *Appl. Opt.* **17**, 2224–7

Liu K, Littman M G 1981 Novel geometry for single-mode scanning of tunable lasers. *Opt. Lett.* **6**, 117–18

López D, Simon M E, Pardo F, Aksyuk V, Klemens F, Cirelli R, Neilson D T, Shea H, Sorsch T, Ferry E, Nalamasu O, Gammel P L 2002 Monolithic MEMS optical switch with amplified out-of-plane angular motion. *IEEE Optical MEMS conference*, Lugano, Switzerland

Marom D M, Neilson D T, Greywall D S, Basavanhally N R, Kolodner P R, Low Y L, Pardo F, Bolle C A, Chandrasekhar S, Buhl L, Giles C R, Oh S-H, Pai C S, Werder K, Soh H T, Bogart G R, Ferry E, Klemens F P, Teffeau K, Miner J F, Rogers S, Bower J E, Keller R C, Mansfield W 2002 Wavelength selective 1×4 switch for 128 WDM channels at 50 GHz spacing. *Optical Fiber Communication Conference (OFC)*, Anaheim, CA, USA pp. 857–9

Marom D M, Neilson D T, Greywall D S, Aksyuk V A, Simon M E, Basavanhally N R, Kolodner P R, Low Y L, Pardo F, Bolle C A, Pai C S, López D, Taylor J A, Bower J E, Leuthold J, Gibbons M A, Giles C R 2003a Wavelength selective 4×1 switch with high spectral efficiency, 10 dB dynamic equalization range and internal blocking capability. ECOC 2003, Rimini, Italy, September 21–25, 2003, Paper Mo3.5.3

Marom D M, Neilson D T, Leuthold J, Gibbons M A, Giles C R 2003b 64 Channel 4×4 wavelength-selective cross-connect for 40 Gb/s channel rates with 10 Tb/s throughput capacity. *Eur. Conf. Optical Communications (ECOC)*, Rimini, Italy

Marom D M, Doerr C R, Basavanhally N R, Cappuzzo M, Gomez L, Chen E, Wong-Foy A, Laskowski E 2004 Wavelength-selective 1×2 switch utilizing a planar lightwave circuit stack and a MEMS micromirror array. *Optical MEMS 2004*, Takamatsu, Japan,August 2004

Marom D M, Neilson D T, Greywall D S, Pai C-S, Basavanhally N R, Aksyuk V A, López D O, Pardo F, Simon M E, Low Y, Kolodner P, Bolle C A 2005a Wavelength-selective $1 \times K$ switches using free-space optics and MEMS micromirrors: Theory, design, and implementation. *IEEE J. Lightwave Technol.* **23**, 1620–1630

Marom D M, Doerr C R, Cappuzzo M, Chen E, Wong-Foy A, Gomez L 2005b Hybrid free-space and planar lightwave circuit wavelength-selective 1×3 switch with integrated

drop-side demultiplexer. *Eur. Conf. Optical Communication (ECOC)* 993–4

Marom D M, Möller L, Su L, López D, Pardo F, Klemens F, Bower E, Ferry E 2006 Compatibility of flat-passband, 200GHz-wide wavelength-selective switch for 160 Gb/s transmission rates. *Optical Fiber Conference (OFC)*, Anaheim, CA, USA,paper OTuF6

Marxer C, de Rooij N F 1999 Micro-opto-mechanical 2×2 switch for single-mode fibers based on plasma-etched silicon mirror, electrostatic actuation. *IEEE J. Lightwave Technol.* **17**, 2–6

Marxer C, Griss P, de Rooij N F 1999 A variable optical attenuator based on silicon micromechanics. *IEEE Photon. Technol. Lett.* **11**, 233–235

McDonald M, Daiber A, Finot M, Su X. 2006 Wavelength filter with integrated thermal control used as an intracavity DWDM laser tuning element. *IEEE Optical MEMS Conference*, Piscataway, NJ, USA

Neilson D T, Frahm R, Kolodner P, Bolle C A, Ryf R, Kim J, Papazian A R, Nuzman C J, Gasparyan A, Basavanhally N R, Aksyuk V A, Gates J V 2004a 256×256 Port optical cross-connect subsystem. *IEEE J. Lightwave Technol.* **22**, 1499–1509

Neilson D T, Tang H, Greywall D S, Basavanhally N R, Ko L, Ramsey D A, Weld J D, Low Y L, Pardo F, López D O, Busch P, Prybyla J, Haueis M, Pai C S, Scotti R, Ryf R 2004b Channel equalization and blocking filter utilizing microelectromechanical mirrors. *IEEE Select. Top. Quant. Electron.* **10**, 563–569

Niederer G, Herzig H P, Shamir J, Thiele H, Schnieper M, Zschokke C 2004 Tunable, oblique incidence resonant grating filter for telecommunications. *Appl. Opt.* **43**, 1683–1684

Nielson G N, Seneviratne D, Lopez-Royo F, Rakich P T, Avrahami Y, Watts M R, Haus H A, Tuller H L, Barbastathis G 2005 Integrated wavelength-selective optical MEMS switching using ring resonator filters. *IEEE Photon. Technol. Lett.* **17**, 1190–1192

Niemi T, Uusimaa M, Tammela S, Heimala P, Kajava T, Kaivola M, Ludvigsen H 2001 Tunable silicon etalon for simultaneous spectral filtering and wavelength monitoring of a DWDM transmitter. *IEEE Photon. Technol. Lett.* **13**, 58–60

Ollier E 2002 Optical MEMS devices based on moving waveguides. *IEEE J. Select. Top. Quant. Electron.* **8**, 155–162

Peter Y-A, Gonte F, Herzig H P, Dandliker R 2002 Micro-optical fiber switch for a large number of interonnects using a deformable mirror. *IEEE Photon. Technol. Lett.* **14**, 301–303

Pezeshki B, Vail E, Kubicky J, Yoffe G, Zou S, Heanue J, Epp P, Rishton S, Ton D, Faraji B, Emanuel M, Hong X, Sherback M, Agrawal V, Chipman C, Razazan T 2002 20-mW widely tunable laser module using DFB array and MEMS selection. *IEEE Photon. Technol. Lett.* **14**, 1457–1459

Ramani C M 2006 Optical MEMS: Boom, bust and beyond. *National Fiber Optics Engineers Conference (NFOEC)*, Anaheim, CA, USA

Rhee J-K, Garcia F, Ellis A, Hallock B, Kennedy T, Lackey T, Lindquist R G, Kondis J P, Scott B A, Harris J M, Wolf D, Dugan M 2001 Variable passband optical add-drop multiplexer using wavelength selective switch. In: *27th Eur. Conf. Optical Communication (ECOC'01).Amsterdam, The Netherlands, Vol. 4*

Rosenblatt D, Sharon A, Friesem A A 1997 Resonant grating waveguide structures. *IEEE J. Quant. Electron.* **33**, 2038–2059

Ryf R, Bernasconi P, Kolodner P, Kim J, Hickey J P, Carr D, Pardo F, Bolle C, Frahm R, Basavanhally N, Yoh C, Ramsey D, George R, Kraus J, Lichtenwalner C, Papazian R, Gates J, Shea H R, Gasparyan A, Muratov V, Griffith J E, Prybyla J A, Goyal S, White C D, Lin M T, Ruel R, Nijander C, Amey S, Neilson D T, Bishop D J, Pau S, Nuzman C, Weis A, Kumar B, Lieuwen D, Aksyuk V, Greywall D S, Lee T C, Soh H T, Mansfield W M, Jin S, Lai W Y, Huggins H A, Barr D L, Cirelli R A, Bogart G R, Teffeau K, Vella R, Mavoori H, Ramirez A, Ciampa N A, Klemens F P, Morris M D, Boone T, Liu J Q, Rosamilia J M, Giles C R 2001 Scalable wavelength-selective crossconnect switch based on MEMS and planar waveguides. *Eur. Conf. Optical Communication (ECOC)*. Amsterdam, The Netherlands

Ryf R, Neilson D T, Aksyuk V A 2006 MEMS based optical switching. In: El-Bawab T S (ed.) *Optical Switching*. Springer, New York

Saadany B, Malak M, Marty F, Mita Y, Khalil D, Bourouina T 2006 Electrostatically-tuned optical filter based on silicon Bragg reflectors. *IEEE Optical MEMS Conference*, Piscataway, NJ, USA

Sagberg H, Lacolle M, Johansen I-R, Lovhaugen O, Solgaard O, Sudbo A S 2003 Configurable spectral filter with an array of diffraction gratings. *IEEE Optical MEMS Conference*, Waikoloa, HI, USA

Saleh B E A, Teich M C 1991 *Fundamentals of Photonics*. Wiley, New York

Solgaard O, Heritage J P, Bhattarai A R 2002 Multi-wavelength cross-connect optical switch. US Pat.6 374 008, April 16

Stuart H R, Baumann F H, Wong-Foy A 2003 Monolithic integration of optical waveguides, MEMS-based switching in silicon-on-insulator. *Proc. SPIE 5347* 137–144

Sugihwo F, Larson M C, Harris J S 1998 Micromachined widely tunable vertical cavity laser diodes. *IEEE J. Microelectromech. Syst.* **7**, 48–55

Syms R R A, Zou H, Stagg J 2004a MEMS iris mechanism. *IEEE Optical MEMS Conference*, Takamatsu, Japan

Syms R R A, Zou H, Stagg J, Moore D F 2004b Multistate latching MEMS variable optical attenuator. *IEEE Photon. Technol. Lett.* **16**, 191–193

Taesombut N, Uyeda F, Chien A A, Smarr L, DeFanti T A, Papadopoulos P, Leigh J, Ellisman M, Orcutt J 2006 The OptIPuter: High-performance, QoS-guaranteed network service for emerging E-science applications. *IEEE Commun. Mag.* **44**, 38–45

Takahashi K, Kwon H N, Mita M, Saruta K, Lee J H, Fujita H, Toshiyoshi H 2006A 3D optical crossconnect using microlens scanner with topologically layer switching architecture. *IEEE Optical MEMS Conference*, Piscataway, NJ, USA

Tarraf A, Riemenschneider F, Strassner M, Daleiden J, Irmer S, Halbritter H, Hillmer H, Meissner P 2004 Continuously tunable 1.55-mm VCSEL implemented by precisely curved dielectric top DBR involving tailored stress. *IEEE Photon. Technol. Lett.* **16**, 720–722

Tayebati P, Wang P, Vakhshoori D, Lu C-C, Azimi M, Sacks R N 1998 Half-symmetric cavity tunable microelectromechanical VCSEL with single spatial mode. *IEEE Photon. Technol. Lett.* **10**, 1679–1681

Tilleman M M, Kozhekin A, Cohen R, Haronian D, Nusinsky I 2005 Coupling losses in multimode waveguides due to misalignment. *IEE Proc. Optoelectron.* **152**, 145–150

Ton D, Yoffe G W, Heanue J F, Emanuel M A, Zou S Y, Kubicky J, Pezeshki B, Vail E C 2004 2.5-Gb/s Modulated widely tunable laser using an electroabsorption modulated DFB array and MEMS selection. *IEEE Photon. Technol. Lett.* **16**, 1573–1575

Torman M, Peter Y-A, Niederman P, Hoogerwerf A, Stanley R 2006 Deformable MEMS grating for wide tunability and high operating speed. *J. Opt. A Pure Appl. Opt.* **8**, S337–S340

Toshiyoshi H, Fujita H 1996 Electrostatic micro torsion mirrors for an optical switch matrix. *J. Microelectromech. Syst.* **5**, 231–237

Toshiyoshi H, Su G-D J, LaCosse J, Wu M C 2003 A surface micromachined optical scanner array using photoresist

lenses fabricated by a thermal reflow process. *IEEE J. Lightwave Technol.* **21**, 1700–1708

Truex T A, Bent A A, Hagood N W 2003 Beam-steering optical switch fabric utilizing piezoelectric actuation technology. *National Fiber Optics Engineers Conference (NFOEC)*, Orlando, FL, USA

Tsai J-c, Wu M C 2005 Gimbal-less MEMS two-axis optical scanner array with high fill-factor. *IEEE J. Microelectromech. Syst.* **14**, 1323–1328

Tsai J-c, Wu M C 2006a A high port-count wavelength-selective switch using a large scan-angle, high fill-factor, two-axis MEMSscanner array. *IEEE Photon. Technol. Lett.* **18**, 1439–1441

Tsai J-c, Wu M C 2006b Design, fabrication, characterization of a high fill-factor, large scan-angle, two-axis scanner array driven by a leverage mechanism. *IEEE J. Microelectromech. Syst.* **15**, 1209–1213

Tsai J-c, Huang S T-Y, Hah D, Wu M C 2006 $1 \times N^2$ wavelength-selective switch with two cross-scanning one-axis analog micromirror arrays in a 4-f optical system. *IEEE J. Lightwave Technol.* **24**, 897–903

Tsuboi O, Kouma N, Soneda H, Okuda H, Mi X, Ueda S, Sawaki I 2004 A high-speed comb-driven micromirror array for $1 \times N$ 80-channel wavelength selective switches. *IEEE Int. Conf. Optical MEMS*, Takamatsu, Japan

Unamuno A, Uttamchandani D 2005 Hybrid MOEMS tunable filter for interrogation of fiber Bragg grating sensors. *IEEE Photon. Technol. Lett.* **17**, 202–204

Unamuno A, Uttamchandani D 2006 MEMS variable optical attenuator with Vernier latching mechanism. *IEEE Photon. Technol. Lett.* **18**, 88–90

Wilson G, Chen C J, Gooding P, Ford J E 2006 Spectral passband filter with independently variable center wavelength, bandwidth. *IEEE Photon. Technol. Lett.* **18**, 1660–1662

Wong C W, Jeon Y, Barbastathis G, Kim S-G 2004 Analog piezoelectric-driven tunable gratings with nanometer resolution. *IEEE J. Microelectromech. Syst.* **13**, 998–1005

Wu M C, Tsai J-c , Huang S, Hah D 2002 MEMS WDM routers using analog micromirror arrays. *Lasers and Electro-Optics Society, 2002 (LEOS 2002)*, The 15th Annual Meeting of the IEEE, Glasgow, Scotland, November 10–14, 2002, Vol. 2, pp. 582–3, www.telcordia.com. See website for documentation availability details

Yano M, Yamagishi F, Tsuda T 2003 Optical MEMS for photonic switching – compact and stable optical crossconnect switches for simple, fast, and flexible wavelength applications in recent photonic networks. *IEEE J. Select. Top. Quant. Electron.* **11**, 383–394

Yao J, Lee M-C M, Leuenberger D, Wu M C 2005 Wavelength- and bandwidth-tunable filters based on MEMS-actuated microdisk resonators. *Optical Fiber Conference (OFC)*, Anaheim, CA, USA,paper OTuM1

Ye Y, Dixit S, Ali M 2000 On joint protection/restoration in IP-centric DWDM-based optical transport networks. *IEEE Commun. Mag.* **38**, 174–183

Yeh J A, Jiang S-S, Lee C 2006 MOEMS variable optical attenuators using rotary comb drive actuators. *IEEE Photon. Technol. Lett.* **18**, 1170–1172

Yu K, Solgaard O 2004 Tunable optical transversal filters based on a Gires–Tournois interferometer with MEMS phase shifters. *IEEE J. Select. Top. Quant. Electron.* **10**, 588–597

Yun S-S, Lee J-H 2003 A micromachined in-plane tunable optical filter using the thermo-optic effect of crystalline silicon. *J. Micromech. Microeng.* **13**, 721–725

Yun S-S, Jo K W, Lee J H 2003 Crystalline Si-based in-plane tunable Fabry-Perot Filter with wide tunable range. *IEEE Optical MEMS Conference*, Waikoloa, HI, USA

Zhang X M, Liu A Q 2000 A MEMS pitch-tunable grating add/drop multiplexers. *IEEE Optical MEMS Conference*, Kauai, HI, USA

Zheng X, Kaman V, Yuan S, Xu Y, Jerphagnon O, Keating A, Anderson R C, Poulsen H N, Liu B, Sechrist J R, Pusarla C, Helkey R, Blumenthal D J, Bowers J E 2003 Three-dimensional MEMS photonic cross-connect switch design and performance. *IEEE J. Select. Top. Quant. Electron.* **9**, 571–578

Zou S Y, Olson R W, Pezeshki B, Vail E C, Yoffe G W, Rishton S A, Emanuel M A, Sherback M A 2004 Narrowly spaced DFB array with integrated heaters for rapid tuning applications. *IEEE Photon. Technol. Lett.* **16**, 1239–1241

Biography

Dan Marom earned the B.Sc. degree in Mechanical Engineering (1989) and the M.Sc. degree in Electrical Engineering (1995), both from Tel-Aviv University. He was awarded the Ph.D. degree in Electrical Engineering by the University of California, San Diego, in 2000. In his doctoral research, under the guidance of Prof. Yeshaiahu (Shaya) Fainman, Dan explored nonlinear wave mixing for optical signal processing of ultrafast waveforms.

From 2000 until 2005, Dr. Marom was employed as a Member of the Technical Staff at Bell Laboratories, then part of Lucent Technologies, where he pioneered the work on Wavelength-Selective Switches (WSS), a critical element for the realization of reconfigurable optical networks using MEMS (Micro-Electro-Mechanical-System) technology.

Dr. Marom joined the faculty of the Applied Physics Department in the fall of 2005, where he is pursuing his research interests in creating photonic devices for switching and manipulating optical signals.

Dr. Dan Marom is active in the scientific community, serving or has served on program committees of leading conferences such as OFC, CLEO, IEEE LEOS, and IEEE Optical MEMS. Dan is the first recipient of the Peter Brojde Excellence in Engineering Award (2006), received the Golda Meir Fellowship (2006), was a Fannie and John Hertz Foundation Fellow during his doctoral studies (1996–2000), and received a Powell Foundation Fellowship (1995).

3.08 Micro-Imaging Systems

Christine Harendt and Heinz-Gerd Graf, Institut für Mikroelektronik, Stuttgart, Germany

3.08.1 Introduction

Visualization is preferred in many technical applications, because the image of a subject or scene provides essential information, which is difficult to obtain with other sensors. Nowadays, digital cameras are available in small sizes and can be integrated in instruments. The rapid miniaturization in semiconductor technology and related techniques enabled the development of miniaturized camera systems. Small-sized imagers are present in everyday life as well as in scientific applications today. In the future various new products can be expected.

This chapter describes miniaturized cameras and their integration into microsystems. First, in Section 3.08.2, general aspects of image sensors such as principles of operation, technology, and general architecture are summarized. System integration and packaging aspects are described in Section 3.08.3. The different applications of integrated microcameras are discussed in Section 3.08.4.

3.08.2 Image Sensors for Microsystems

Image sensors are devices consisting of a photosensitive solid-state sensor array and addressing and readout circuitry. Since the development of the charge-coupled device (CCD) principle at Bell Labs in 1969 (Boyle and Smith 1970) and the demonstration of its first application in Esser (1972), sensors have been developed in small volumes for scientific and space applications (Janesik 2001). In the mid-

1970s, CCDs began to appear in commercial applications such as in fax machines, and later in camcorders.

Some of the initial applications of complementary metal oxide semiconductor (CMOS) image sensors have also been shown by Weckler (1967), but at that time no commercial product was feasible due to the size of CMOS transistor structures. Since then, the advances in performance, image quality, and the price of these imagers made them the major device used in cameras, in photography, and in other imaging applications. The structure of size reduction in semiconductor technology enables the development of miniaturized image sensors and the integration into microsystems. This chapter deals exclusively with silicon photodetectors, which are the major devices used in microimagers.

3.08.2.1 Interaction of Photons with Silicon

The interaction of a photon with silicon can result in the reflection, the transmission, or the absorption of the photon. The following section describes the phenomena briefly and detailed information is given in Tredwell (1995), Pankove (1971), and Sze (1981).

3.08.2.1.1 Absorption

The absorption of a photon and the generation of an electron–hole pair is the effect used for photodetection. Light absorption in semiconductor materials normally occurs by transferring a valence band electron into the conduction band (**Figure 1**). This process depends on the density of occupied states in the valence and the empty states in the conduction band. Also, the absorption coefficient is a function of the transferred energy and the wavelength of the light. Because of the high density of states, the

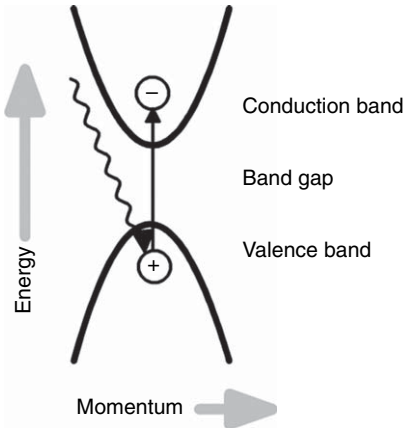


Figure 1 Semiconductor band model.

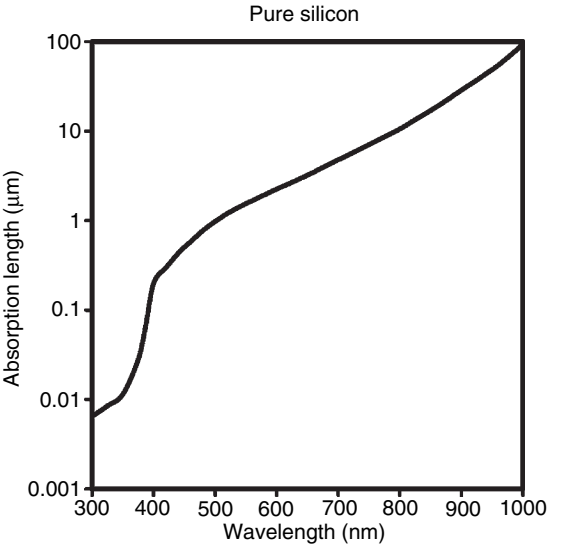


Figure 2 Absorption of light in silicon.

absorption (**Figure 2**) is high for high-energy UV light and low for low-energy near-infrared (NIR) light. For incident light with energies lower than the bandgap energy, there is no transfer between states and no absorption occurs. For single-crystal silicon, the absorption lengths varies between 0.4 μm for blue light (450 nm), 1.5 μm for green light (550 nm), and 10 μm for NIR light (800–1000 nm). There is no absorption beyond 1100 nm because the photon energy is less than the silicon bandgap (1.1 eV). The data demonstrate the possible range of spectral response for a silicon device between 200 and 1000 nm. Only a low NIR response is possible for shallow devices.

The competing process is the recombination of the carriers before detection. The recombination probability depends strongly on the number of interface states and defects or impurities (Auth *et al.* 1977). Therefore, recombination occurs particularly near the surface and can be influenced by defects.

3.08.2.1.2 Reflection

The relatively high index of refraction η of silicon ($\eta = 4$) results for the pure uncovered surface in a reflection of nearly 30% of the visible light (**Figure 3**). Thin layers of dielectric material with lower refractive index serve as antireflection coatings (ARC). A reflection lower than 10% results for layers with a thickness of odd-numbered multiples of a quarter wavelength. This is caused by the interference of the reflected waves on both surfaces. The so-called quarter wavelength layers for antireflection

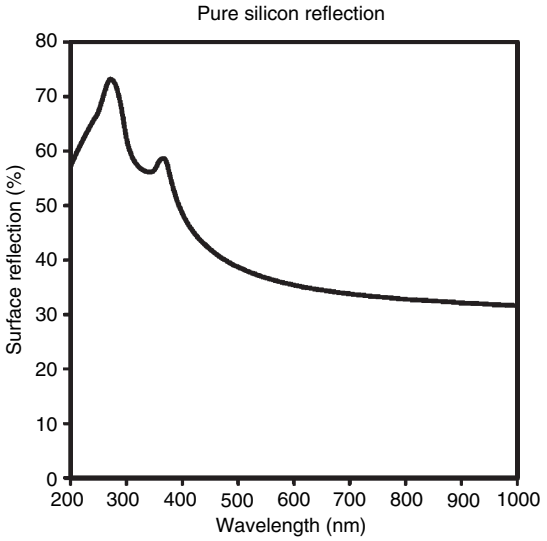


Figure 3 Reflection of light on silicon.

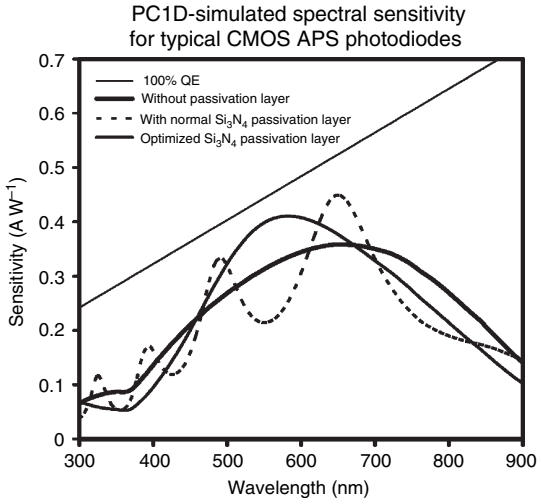


Figure 4 Spectral sensitivity of silicon.

are only effective when the layer is thin (<150 nm). With thicker layers, the interference is disturbed by thickness variations and different incident angles.

3.08.2.1.3 Spectral response

A measure for the spectral sensitivity of the diode is the quantum efficiency (QE), which is the ratio of the charge collected to the number of incident photons. Thus, for QE = 100%, every incident photon generates a detected charge. The QE depends on the transmission of the incident light through the layers covering the junction diode, on the absorption of the light in silicon, and on the probability of the detection of a generated electron–hole pair before recombination. This function describes the major parameters for the optimization of silicon photodetectors: adjustment of the covering layers of the diode to the optical requirements, design of the junction diode (e.g., optimization of the depth of the junction) for optimal absorption of the desired light, and optimization of charge collection, which implies the reduction of recombination probability. The spectral sensitivity is also an indicator of the photon energy, which increases with lower wavelength. In **Figure 4**, the spectral response of a silicon photodiode is shown. The QE = 100% is indicated by the straight line in **Figure 4**. The relatively low values for blue light (400 nm) are because of the lower transmission (i.e., light absorption in the cover layers) and the high recombination probability in the near-surface regions of the junction. Special efforts are needed to improve the near-UV light sensitivity. Infrared light has a low absorption probability due to the low energy of the photons. Consequently, the sensitivity to the IR light is reduced.

3.08.2.2 Image Sensor Devices

3.08.2.2.1 Photodetectors

There are three basic types of structures used for image sensing: the junction photodiode, the photocapacitor, and the photoconductor. While the junction photodiode is generally a single-crystal silicon device, the latter are mostly fabricated in amorphous silicon. In this section, the junction photodiode, the detector widely used in CMOS devices, and the photocapacitor, the typical CCD structure, are described. A detailed description of the devices and their performance is given in Sze (1981) and Tredwell (1995).

3.08.2.2.1.(i) The junction photodiode A pn junction photodiode is shown in **Figure 5**. The photodiode is formed by boron (p)- and phosphorus or arsenic (n)-doped areas in the single-crystal silicon. If a voltage is applied between the p and the n contact in reverse bias, typically between 1 and 5 V, a depletion layer is formed. Photon-generated minority carriers are accelerated in the resulting electric field and collected, if they do not recombine before collection. The depletion layer width depends on voltage and dopant concentration in the junction diode. Typical widths of depletion layers for silicon photodiodes are 0.3–3 μm. Minority carriers generated outside the depletion layers contribute to the

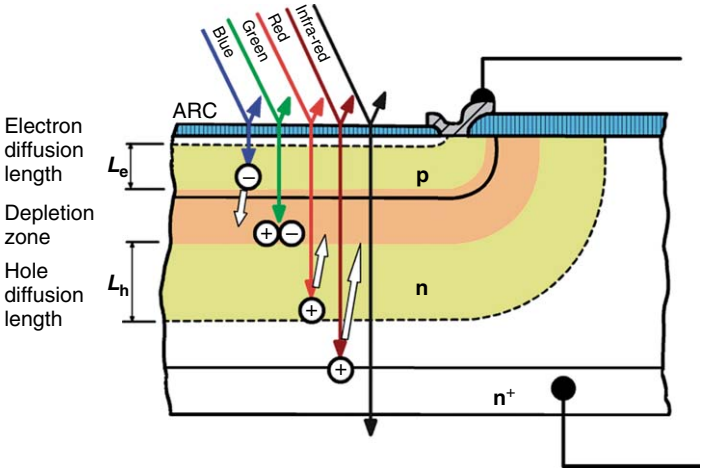


Figure 5 Schematic of a junction photodiode with ARC (antireflection coating).

photo-generated current, if they get into the depletion region. This carrier transport can occur by carrier diffusion or the drift in the small electric fields generated by different dopant concentrations inside the n or p region of the diode.

In many applications, the junction diode is also used for storage of the collected carriers until they can be read out. As a consequence, the reverse voltage on the junction decreases depending on the amount of charge collected until the charge is removed in the readout process. In the case of high light intensities, many charges are collected and the voltage drops to forward bias. Then additional light is compensated by the diode's forward current. No additional response occurs and the sensor is saturated.

3.08.2.2.1.(ii) CMOS image sensors The CMOS device, the major technology used for consumer semiconductor products such as integrated circuits (IC), microprocessors, and memories, is based on transistor structures fabricated in single-crystal silicon. Thus, the integration of a pn junction photodiode into a CMOS technology does not require additional effort.

The first image detector using MOS technology was the passive MOS array, an addressable array of photosensitive junction diodes (**Figure 6**). An individual photodiode is selected by one or two transfer transistors. After selection, the photodiode is recharged to a fixed reset voltage using an integrating amplifier. The charge needed for a given integration time is a measure of the photocurrent.

The rather poor performance of these devices compared to that of the already established CCDs was mainly due to the nonuniformities and noise. Noisy output signals were caused by low signals levels, which were connected to the read line with the large capacitance of many transistor drains.

Consequently, the integration of a gain transistor, which enhances the signal and improves the image intensity and uniformity, was the next step. The device, the active pixel sensor (APS), is the most widely used CMOS imager type today. The schematic of passive (MOS) and active (APS) pixels is shown in **Figure 6**. The APS sensor has an amplifying transistor integrated into each pixel.

The photodiode is manufactured using the diffusions necessary for the transistor. Each diode is connected to the transistors needed for pixel addressing, amplification, and reset. The CMOS APS enables the implementation of desired functions directly at the pixel and allows APSs with flexible readout schemes such as access to subframes or single pixels and many other features. These advantages, however, are limited by the feasible structure size in CMOS technology and the surface area needed for the devices. Both the photodiode and the transistors cover a defined area of the silicon surface, so the part of the surface used by the photosensitive element, the fill factor, is typically 30–60%. The high complexity of the circuitry lowers the fill factor. The minimal transistor size is determined by the minimal structure size, which can be reproduced in the photolithographic processes used in CMOS technology. If this structure size (i.e., gate width) is 0.5 μm, which is the technological limit that was reached in 1993, one

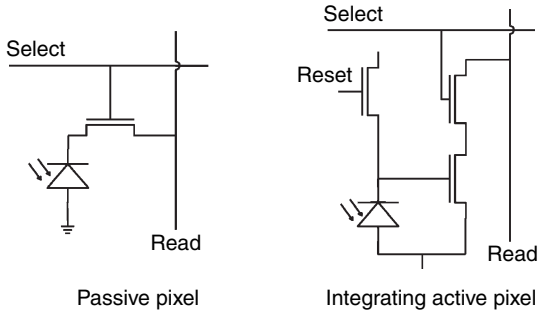


Figure 6 Schematic of metal oxide semiconductor (MOS) and active pixel sensor (APS) pixels.

transistor covers an area of about 2 μm × 3 μm. The resulting pixel consists of three to seven transistors and the photodiode needs at least 20 μm × 20 μm, which gives a 500 × 500 pixel array with an image field of 1 cm × 1 cm. This illustrates the fact that CMOS image sensors could only participate significantly in the image sensor market since submicron CMOS technologies with structure sizes of 0.25 μm and below are available.

One major goal when designing image sensors for microcameras is the achievable pixel size. Thus, in most cases, only three transistor cells are used. Advanced pixel designs share transistors (1.5 transistor cell) in order to reduce the pixel pitch.

In CMOS technology, the photosensitive element can be combined with a number of devices (transistors, resistors, and capacitors). Thus, different pixel types and readout schemes are possible. The most widely used type is the integrating linear sensor.

The photodiode is precharged to a fixed voltage using the reset transistor. After the integration time used for collecting photo-induced currents, the output voltage change on the photodiode is measured using the fixed voltage or a stored dark pixel output as reference. The capacitance of the noise-limiting node is build by the photodiode, reset transistor, and amplifying readout transistor only. But each pixel readout path forms an individual assembly of circuit elements resulting in variations of gain and offset.

This approach reveals good results for medium-contrast scenes and enables, by an increase of the integration time, a high sensitivity for low light conditions. Typically, the dynamic range of these sensors is 60–80 dB, the adaptation to the different illumination levels is done by mechanical apertures. For high-contrast scenes, saturation in the bright spots cause a loss of image information.

Another concept, the pixel with logarithmic response, uses a nonlinear shunt for continuous reset. The photon-induced voltage on a subthreshold transistor, which has an exponential rising current versus the linear rising source/drain voltage, is detected (**Figure 7**). The resulting response for different illuminations shows a logarithmic increase of the output signal with the incident light, rather similar to the sensitivity of the human eye. The dynamic range of this type of image sensor (high dynamic range CMOS (HDRC)) is >120 dB, allowing the saturation-free detection of scenes with a variation of illumination over six and more decades (spot welding, light bulbs, near-field illumination, and reflection of light sources) (HDRC, 1996, 1997, Seger *et al.* 1993, 1999). Compared to linear sensors, the lower sensitivity for low light conditions and a higher fixed pattern noise (see Section 3.08.2.2.1.(v)) are the disadvantages of the logarithmic approach. The LinLog sensor (www.photonfocus.com) combines both sensor types and the advantages of logarithmic and linear sensors. But an increased complexity of the device restricts its performance for many applications.

CMOS image sensors differ in the pixel design; also, there are a number of different addressing and readout schemes. In principle, each individual pixel could be addressed and read out separately. In practice, the readout is triggered by a clock signal and the charge measured at each pixel is transferred by a multiplexer to the output. The time needed for the complete readout of an image determines the number of images, which can be transferred per second: the frame rate.

A typical addressing scheme is the line-synchronous sampling. It is simply the sequential readout of one pixel row after the other, often called rolling shutter. This principle is widely used in consumer applications. Additionally, variations of this method are possible, i.e., the readout of subframes or single lines.

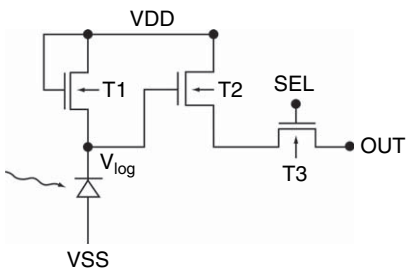


Figure 7 Schematic of log-compressing pixel with VDD (positive supply voltage), VSS (ground voltage), SEL (select pulse), V_{log} (illumination dependent voltage with logarithmic response), OUT (output voltage), T1, T2, and T3 (transistors).

For some applications, the rolling shutter proves disadvantageous, for example, for observing fast-moving objects. As the imager detects images during readout, the rolling shutter image changes within the process with the consequence that the images show smeared figures. An alternative readout scheme is field-synchronous sampling. This approach, also known as global shutter, freezes the information at each pixel at the same time. The readout itself is done sequentially, but the image information is stored once, resulting in good image quality independent of the movements. This scheme requires an additional transistor at each pixel, thus requiring a larger pixel size.

In order to obtain the desired functionality, each image sensor has a number of circuit blocks (clock generator, addressing logic, etc.), which are located outside the photosensitive area. Other camera functions such as analog (A)/digital (D) conversion, image data processing (fixed pattern noise (FPN) correction), or color conversion are in many cases integrated in the CMOS imager too. This is one advantage of the CMOS technology and helps to reduce the number of additional boards, interconnects, package efforts, and cost. In microcamera applications such as endoscopy, where the chip size correlates directly to the system size, the integration of additional functionality and the chip size optimization are a trade-off.

3.08.2.2.1.(iii) The photocapacitor The photosensitive element used in CCD imagers is the photocapacitor. As shown in **Figure 8**, an optically transparent conductive electrode forms one plate, the silicon substrate is the other plate, and a thin oxide layer is the insulator of the photocapacitor. A voltage applied to the transparent electrode results in the accumulation of the majority carriers in one direction and a depletion layer in the opposite direction. This depletion layer is formed by the fixed (nonmobile) ionized dopants. A large depletion layer thickness and thus a low capacitance is observed immediately after applying the voltage. After that, the photon-generated minority carriers are formed in the depletion layer or enter it by diffusion. These carriers are transferred and collected near the silicon–oxide interface (*cf.* **Figure 9**). As a consequence, the depletion layer shrinks and the total capacitance rises. The total collected minority charge in this depletion, the so-called bucket, is the charge Q, which depends on the voltage U and the difference between the oxide capacitance C_{ox} and the minimal capacity of the

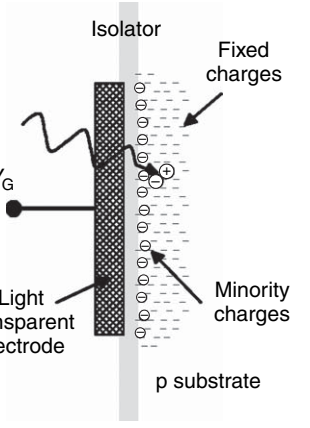


Figure 8 Schematic of a photocapacitor.

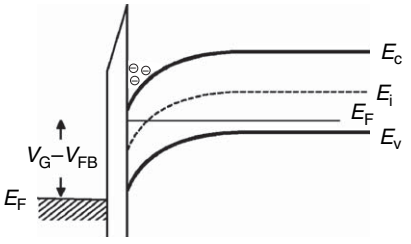


Figure 9 Bandmodel of the semiconductor–insulator interface. The energy levels of the valence band (E_V), the conduction band (E_C), the intrinsic energy (E_i), and the Fermi level (E_F) are shown as well as the gate voltage V_G and the so-called flatband voltage V_{FB}, which depends on the material.

device C_{min}. This is the storage capability of the device:

$$Q = (C_{ox} - C_{min}) \times U$$

Only a very slow recombination of minority carriers in this bucket occurs because of the absence of recombination partners. This type of capacitor is also an ideal device for the collection and storage of a limited number of photon-generated minority charges. Excessive charges are injected in the non-depleted substrate area or are collected by neighboring pixels (blooming).

3.08.2.2.1.(iv) CCD image sensors In a row of connected photocapacitors (**Figure 10**), the collected charge can be transferred from one capacitor to its neighbor. By applying the depleting voltage to both electrodes, a larger depletion bucket is formed,

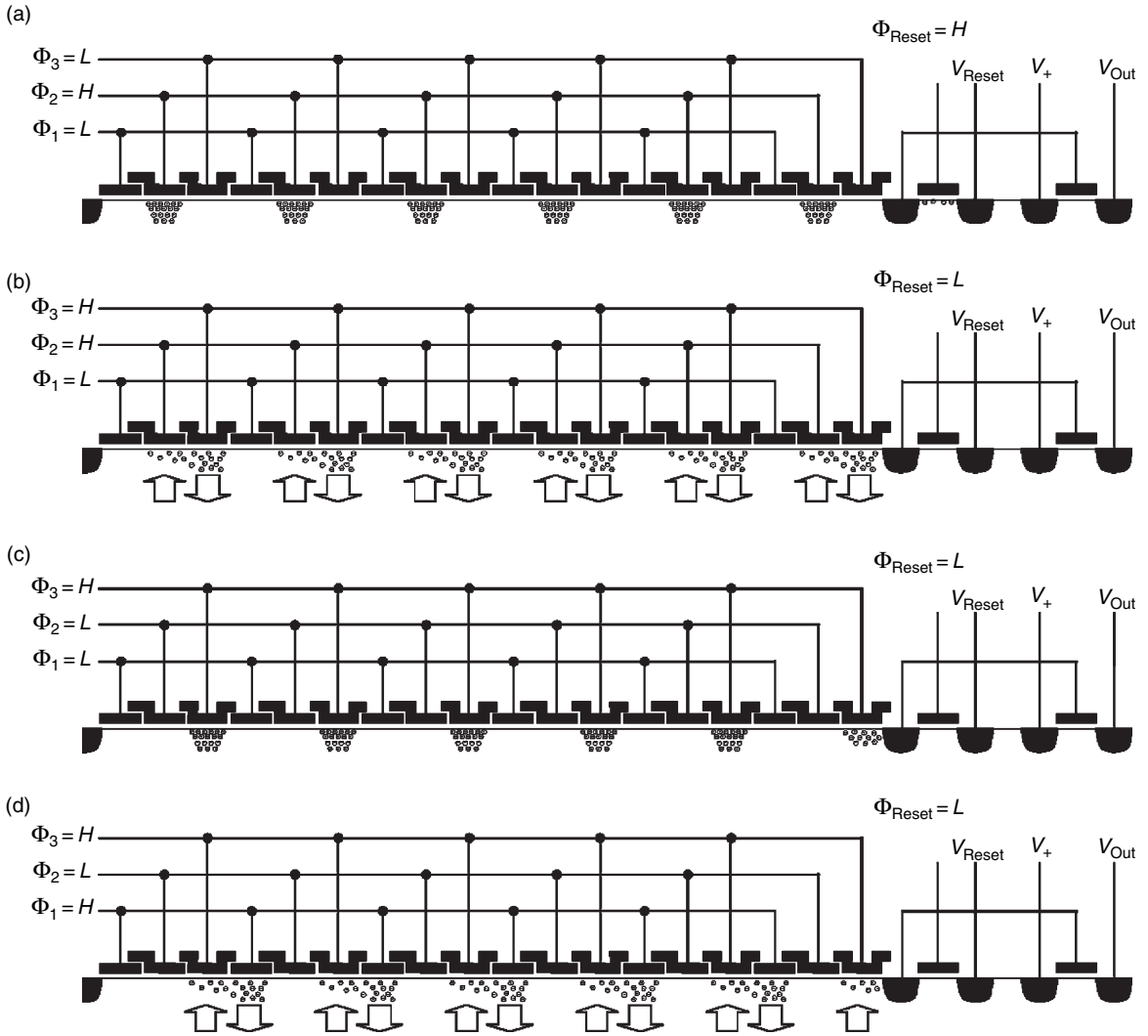


Figure 10 Schematic of charge-coupled device (CCD) readout principle.

holding the collected charge from the first capacitor. By the subsequent reduction of the depletion voltage on the first electrode, the collected charge is transferred to the neighboring capacitor. This is known as the CCD principle.

By means of three phase-shifted overlapping voltages, a number of N charges can be collected and transferred in a line of $3N$ capacitors. A diffused diode at the end of the line works as a contact to the amplifying transistor. This diode is periodically precharged to a fixed voltage by the reset transistor. After arrival of the transferred charge, the gate voltage changes on the amplifying transistor. This results in modulations of the transistors source/drain current.

For 2D arrays with M rows and N columns, arrangements of one horizontally oriented CCD line with N vertically arranged CCD lines are used. One bucket from each vertical CCD is transferred to the horizontally arranged CCD before this CCD is read out from the whole line. After $3N$ transfers reaching the end of the line, the next bucket of the vertically arranged CCDs is transferred to the horizontal CCD. This way the whole image is sequentially read out row by row.

Different types of arrangement of photosensitive and nonsensitive buckets are used for the storage of integrated charges during readout.

In an interline transfer CCD, all charges integrated in the photosensitive buckets are simultaneously

transferred into a metal-shielded nonphotosensitive CCD (Figure 11).

In the frame transfer CCD, the vertically arranged CCDs double their length. The first half of this CCD is photosensitive, the second half is covered with metal. All collected charges from the photosensitive part are transferred vertically to the metal-covered storage part by $3M$ short transfers. After this first step, the photosensitive part is set to charge integrating conditions. Only the second half transfers the image content to the horizontally arranged CCD (Figure 12).

CCDs provide a very sensitive picture cell with the lowest possible capacitance and the best noise performance. All image contents are transferred to the same output, resulting in a high homogeneity of the image. The photosensitive electrodes, however, show strong oscillations in the spectral sensitivity. The high sensitivity of the device is correlated to a limited charge transfer. Overexposure can cause charge detection in neighboring pixels. The signal-to-noise (S/N) ratio and the limited dynamic range lead to image artifacts such as saturation, blur, or smearing.

A major advantage is the CCD process technology. Because to its lower complexity, it needs less lithographic steps. Additionally, a smaller pixel size and a higher fill factor can be achieved especially for the frame transfer organization. Therefore, image sensors of this type have been used successfully in most of the image sensor applications since the early 1990s and still dominate the markets for camcorders and digital cameras.

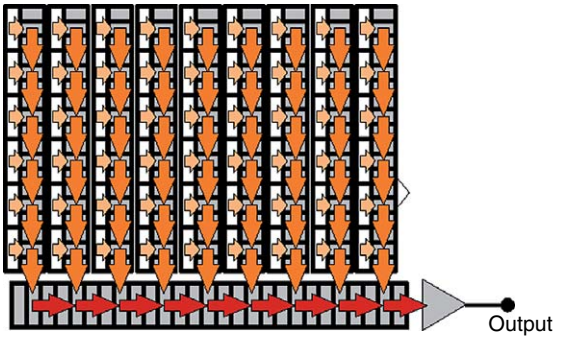


Figure 11 Interline transfer charge-coupled devices (CCDs). Small horizontal arrows (orange) indicate the charge transfer from the photosensitive cells to the column CCDs. The next step is the transfer of the first charge packet of every column CCD into the row CCD. This is shown by the vertical arrows (light red). Then all charges from the first row are transferred to the output (horizontal arrows, dark red). This sequence is repeated until all charges are read out.

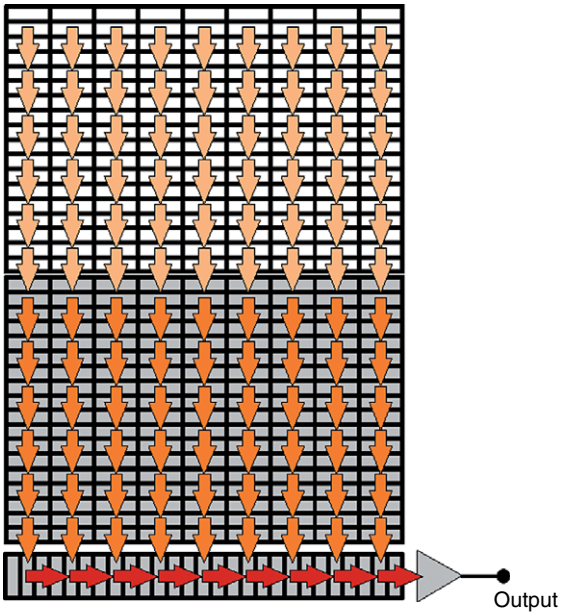


Figure 12 Frame transfer charge-coupled device (CCD). The integrated charges are transferred simultaneously into the light-shielded storage area (vertical arrows, orange). For readout, the stored charges are transferred to the horizontal row CCD (vertical arrows, light red). Then all charges from the first row are transferred to the output (horizontal arrows, dark red).

For some applications, very large and highly sensitive backside illuminated sensors can be fabricated and assembled on glass or ceramic substrates. Operating on low temperatures and with multisampling, they offer a nearly unlimited low light sensitivity and a good dynamic range for long-time exposure.

3.08.2.2.1.(v) Limitations of image quality For both device types, there are a number of parameters limiting the image quality obtainable. First, different sources of noise cause signals, added to the photo-induced signals. The major sources of temporal noise are photon noise, reset and readout noise, and the dark current. The photon noise is caused by the statistical variation of the incident photons. This noise depends on the light intensity and cannot be affected by the performance of the imager.

Reset and readout noise are caused by noise generated in the circuitry. Strategies in design optimization for reduced noise are applied in image sensor design using low-input node capacitance and

high gain. In CCDs and some CMOS APS, the reset noise can be compensated by the difference between the reset voltage and output voltage of the each pixel.

The dark current affects the low light sensitivity, which is the major task in image sensor technology optimization. Even in the absence of light in every photosensitive element, a number of electron–hole pairs are generated resulting in the dark current. It restricts the minimal illumination detectable by a photosensor and is an important criterion for the device quality. The dark current limits the applicable integration time. Even if a dark current compensation is applied, the dark current shot noise reduces the S/N value of the devices.

Major sources of these charges are defects and impurities in the silicon and interface states at the $Si–SiO_2$ interface. Typical values of dark currents for silicon junction diodes are 10–100 pA cm^{-2}.

The dark current depends significantly on the temperature: it doubles approximately at every 10°C rise in temperature. This is the reason for using cooled devices for high-performance imaging. On the other hand, this indicates the difficulties to obtain good image quality for applications with higher ambient temperature (human body, sensors, and robotics in the production environment). Self-heating of the device is another concern, which might cause an unintended rise in temperature and thus in increased dark currents.

Another source of noise, the static noise, is caused by nonuniformities of the device parameter. A dark image of a pixel array shows a variation of dark signal intensity known as fixed pattern noise (FPN). This FPN pattern is constant for every device. Normally, a subtraction of the dark values is implemented in the image processing (FPN correction).

For a more detailed discussion, the FPN can by split into different parts: the dark current value of each photosensitive element depends on the overall material properties of the semiconductor; it also shows slight variations from pixel to pixel caused by technological imperfection in structure size and defect density. A measure for the dark current variations is the dark signal nonuniformity (DSNU) value of the image sensing device.

The same applies for the values of the photoresponse and gain of each pixel. This is measured in the photoresponse nonuniformity (PRNU) given for a signal near saturation level and the photoresponse nonlinearity (PRNL).

The effects influencing the signals in neighboring pixels are known as crosstalk. This can be caused by carrier diffusion in CMOS sensors or by charge overflow in CCDs (blooming). During the carrier transport, crosstalk along the readout path causes a bright line (smearing). Image lag is caused by incomplete discharge of the pixels at reset. The resulting image of a moved bright object shows a trace.

For both device types, improvements of image quality and the noise reduction strategies are major tasks in the development of devices and technology. The optimization of the QE and the reduction of the dark current require efforts to improve the properties of the material (minimization of defects, optimization of top layers). Additionally, some efforts in device design and correction circuitry are needed for miniaturized image sensors.

3.08.2.2.1.(vi) CMOS vs. CCD While the CCD imagers dominated almost all solid-state imaging applications from the beginning in 1972 until a few years ago, CMOS imagers have been acquiring a growing market share since the late 1990s (Fossum 1997, Mendis *et al.* 1994). This is due to the fact that only with the advanced submicron CMOS technologies (minimal lithographic structure size <0.5 μm), acceptable pixel sizes were achievable. The CMOS technology improvements pushed by the mainstream products (processors and memories) enabled the development of CMOS imagers. Major advantages are certainly the flexibility of the approach, allowing random access readout, active pixel sensors, and the integration of additional features. The CMOS process as the mainstream technology enables high-volume and thus low-cost production. Additionally, the various devices and design blocks available in libraries facilitate the implementation of new features. The CMOS technology itself, however, has a number of disadvantages for photosensitive devices.

The reduction in pixel size is going on for both device types. CMOS image sensors with 3.3 μm pixel pitch are available (www.tsmc.com). For 2007, a value of 2.2 μm (Nummela *et al.* 2006, www.tsmc.com) is expected. Thus, the achievable pixel size even with the most advanced CMOS technology is bigger than that for CCDs, which is an important issue especially for microcameras. The fill factor, normally <50% for small pixel sizes, limits the sensitivity of the sensor. Additionally, standard CMOS processing result in devices with low performance of the photosensor in terms of dark current, light absorption, and sensitivity. In order to reach the requirements of the photodiode, CMOS processes have to be optimized

for image sensors (see Section 3.08.2.3). For advanced CMOS technologies, illumination levels of 5 mlux to 5 lux can be detected at normal videoframe rates.

CCD technology, on the other hand, is still the champion in pixel size and achievable fill factor. A CCD with 1.56-μm pixel (Oda *et al.* 2005) has been reported. The QE in the visible range, especially for blue light, is better and the special and less complex CCD process technology has more degrees of freedom for the optimization of optical parameters (i.e., the top layers). An experience of more than 20 years in device optimization and image quality improvements is another advantage of the technology.

The main restriction is the fixed readout scheme. The process incompatibility with CMOS technology anticipates the integration of additional features. For microsystem approaches, the higher power consumption and the relatively high number of inputs/outputs (I/O) are drawbacks.

The performance of the CCD depends mainly on the charge capacity and the charge transfer efficiency of the device. Typical values for the charge handling capacity of microcamera CCDs are 10 000 electrons per pixel. This allows the detection of 1-mlux signal on the device surface. The major limitations are saturation and blooming for high light intensities.

3.08.2.3 Technology Aspects

3.08.2.3.1 CMOS technology

As stated above, the manufacturing of pn junction diodes is integrated in the CMOS process. There are two diffusion processes available for the formation of the diode: the relatively deep diffusion of the well of the transistor and the shallow diffusion of the source and drain contact. Depending on the junction depth, the spectral response of the diode is modified (*cf.* Section 3.08.2.1). Generally, when developing submicron CMOS processes, the depth of the well has to be reduced to meet the lateral size requirements. It is not possible to manufacture photodiodes having deep junctions (>3 μm) by advanced standard processes. Extra process steps are needed to obtain the full spectral sensitivity.

The performance of junction diodes fabricated by standard CMOS technology is in many cases low, because reduction of dark current and low defect densities in the diode are not the major optimization criteria for process engineers. The reduction of the dark current, the optimization of the spectral response, and the reduction of the recombination probability are the tasks for a good imager process technology. Thus, the reduction of defect densities and the adaptation of junction depth and optical performance of cover layers are the issues in process engineering. With the growing market for CMOS imagers, special imager processes with these optimized parameters for photodiodes have been developed and are now provided by most foundry services.

3.08.2.3.2 High fill factor processes

The performance of the photosensitive element depends also on the optical parameters of the layers covering the sensor surface. In advanced CMOS technologies, up to five metallization layers are used for wiring. These layers are separated by dielectric layers, mainly deposited oxides, which are planarized by chemomechanical polishing in order to reduce the topography. Thus, the silicon diode surface is covered by a stack of dielectric layers that are several microns in thickness, which might vary over the wafer due to polishing process variation. As shown in **Figure 13**, this is a problem for the transmission of light; it also reduces the field of view. Additionally, reflected light might be transmitted in the horizontal layers resulting in crosstalk. The other major limitation of CMOS technology is the relatively low fill factor, which reduces light sensitivity and device performance.

One approach to enhance the photosensitivity is the concept proposed in the fill factory patent. An additional diffusion collects the photon-induced electrons generated below the photodiode and in silicon areas covered with transistors.

Another concept, the vertical integration of the photodiode and the CMOS readout circuitry, has been proposed and several approaches have been shown (Fischer *et al.* 1992, Harendt *et al.* 1998, Neidlinger *et al.* 1999). One straightforward process is the low-temperature deposition of amorphous silicon photodiodes on top of CMOS image sensors as shown in **Figure 13**.

These Thin-Film-on-CMOS (TFC) diodes have a p–i–n layer structure of amorphous silicon-based photodiodes. It mainly consists of an undoped (i.e., intrinsic i) absorber layer sandwiched between very thin doped layers generating the required electric field across the intrinsic layer.

Because of the short diffusion length of charge carriers in the amorphous Si:H-based semiconductors of typically 100–150 nm, no pixel separation is

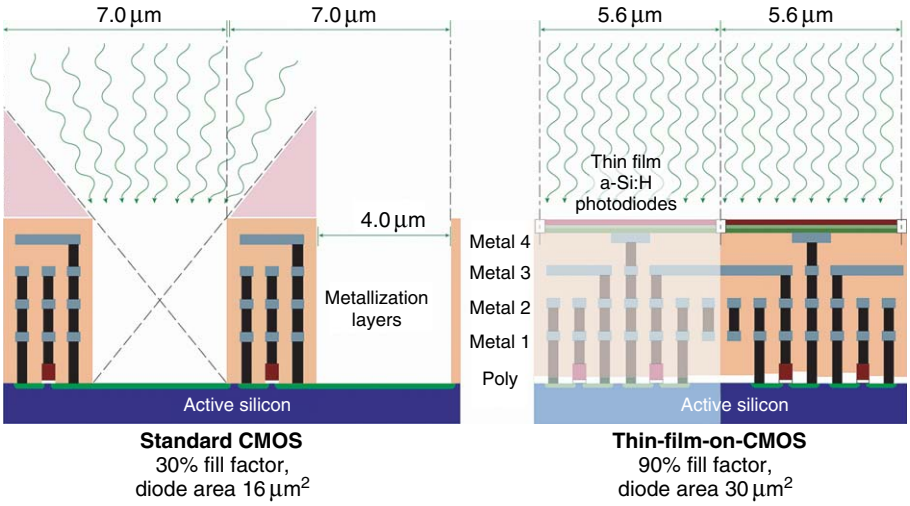


Figure 13 Cross-section view of a pixel in bulk silicon and the Thin-Film-on-CMOS (TFC) photodiodes.

needed in the upper photodiode layers. Down to a back contact spacing of 1 µm, the crosstalk between adjacent pixels outperforms CCDs by a factor of 5–10. Moreover, the small diffusion length determines the layer structure of the thin film photodiodes. In contrast to crystalline semiconductors, diffusion-based carrier collection is almost impossible in these amorphous materials. Therefore, a built-in electric field must serve the purpose of separating the photon-generated charge carriers. The TFC process is an add-on to the standard CMOS production process. The wafers are fabricated in a 0.25-µm CMOS process by using a wafer foundry service. Before depositing the TFC diodes, an additional layer with metal contacts is patterned for each pixel. The manufacture of the amorphous p–i–n diodes is done via plasma-enhanced chemical vapor deposition (PEVCD) at low temperatures of 100–150°C. Subsequently, a layer of zinc oxide is deposited, forming a transparent conductive top electrode, which is contacted via an alumina layer that is patterned at the edge of the image field. Deposition of an antireflection coating and opening of the bond pads completes the wafer-level processing.

3.08.2.3.3 Color technology

For many applications color imaging is preferable. The majority of color image devices in use today use color filters on top of the sensor. The filter material is normally an organic resist containing color pigments. The wafers are spin coated with the color resist and the color filters are patterned using lithographic processes. The pixels are coated with the three color filters (red, green and blue or yellow, magenta, and cyan) in a regular pattern, i.e., alternately red and green or blue and green. For some applications also, black filters for pixel grid and circuitry covers are used. These absorbing grid structures help to avoid the detection of scattered light in neighboring pixels and thus reduce crosstalk.

Often color processes include planarization layers to reduce surface topography. The color filter material is typically stable up to temperatures slightly above 200°C. Color degradations after sterilization processes in medical applications have been observed. The color process is often combined with the fabrication of microlenses on top of the coating, requiring additional planarization layers. Finally, antireflection coating is necessary to obtain good performance. In total, the color process consists of three to five lithography steps and a similar number of layers resulting in a significant increase of processing costs. So far, this technology with organic color filters has been used in most microcameras. For very small pixel sizes, however, the resist thickness required for the color separation and the size of the pigments limit the performance, mainly the device sensitivity.

Recently, amorphous silicon has been tested as a color filter material (Kasano *et al.* 2006). The color separation is determined by the different absorption length of the light in the amorphous silicon, which is

used as a filter material with a different thickness on top of the CCD or CMOS device. The deposited amorphous silicon layers can be easily patterned lithographically and thinned by dry etching processes. These layers are only 40–80 nm thick and reveal a better transmission than do organic filters.

The color information obtained from each pixel is processed in the software resulting in a color picture. Although software processes such as interpolation between the pixels improve the quality, the color image has significantly lower resolution due to the combination of three pixels. Additionally, the color filters absorb light and thus increase the minimal detectable illumination. As a consequence, images may show moiré pattern, increased noise, or color seams in low light conditions.

The restrictions of the color filter technology led to a number of different approaches. The sequential illumination with colored light (using mechanical or electrical shutters) is one of the earliest techniques. The resulting color images have the full image resolution, but suffer from reduced image quality for moved scenes (color seams).

Another approach is the separate color detection in one pixel presented by Foveon (Lyon and Hubel 2002, www.foveon.com). This concept uses different absorption lengths for photons of different wavelengths (*cf.* Section 3.08.2.2.1) for separate detection. The pixel contains three vertically stacked subpixels formed by diffused p and n wells. While blue light is detected in the upper junction diode, green and red light is absorbed in the lower parts of the stack. Compared to color filters, however, the color separation is less optimized.

In the past, a three-color pixel technology has been demonstrated using amorphous silicon photodiodes (Eberhardt *et al.* 1995, Neidlinger *et al.* 1996, Zhu *et al.* 1994). Again, a different absorption length is used for the color separation, but within the amorphous silicon photodiode stack on top of the CMOS device. The drawback of this approach is the time constant for the charge collection in each section, limiting the frame rate.

A promising technology for the future might be a detection principle somewhat similar to that of organic light-emitting electrodes (OLEDs) for pixel image sensors. Recently, in a press release from Fuji, a new color image sensor technology has been announced using organic pigments as photodetectors (Tanaguchi and Takada 2005). Three different layers of organic material are forming a color sensor. Products in this technology are expected in 2010.

3.08.2.3.4 Microlenses

In the sections above, the fill factor problem has been addressed. With the reduction of the pixel pitch and the size reduction in technology (both CCD and CMOS) the photosensitive detector is located at the bottom of a stack of transparent layers (**Figure 14**). For small pixel sizes, the aspect ratio could be 8:1 or more. By positioning a lens on top of each pixel, the amount of light detectable at the sensor increases. A schematic of such a device is shown in **Figure 14**. Typically, these on-chip microlenses are fabricated using a layer of organic material with the desired optical performance, which is patterned with a grid. The microlenses are formed in a thermal molding step. In process optimization, the focus of the lens has to be adjusted to the photodetector, e.g., by transparent planarization layers. The combination of microlenses with the optics and the camera, however, can reduce the aperture angle. An offset between microlenses in the center of the sensor and at edge positions reduces the effect.

3.08.2.4 Scanning Sensors

While CMOS and CCD sensors are self-scanning systems, scanning sensors consist of a row of photosensitive elements and a scanning system. The advantage of this principle, which is used in many volume products such as fax machines, is the reduced complexity of the image sensor. It requires, however, a moving part for the scan. This can be either the sensor itself or a movable mirror. Microscanners use movable mirrors fabricated using microelectromechanical (MEMS) technology, which can be precisely controlled by electrostatics. Single-crystal silicon, often covered with a metal coating in order to improve the reflectivity, is a common material for mirrors.

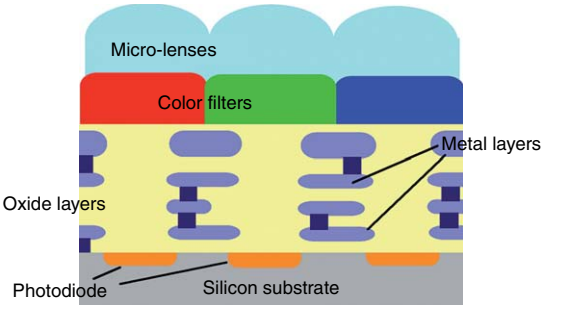


Figure 14 Schematic of an image sensor with color filters and microlenses.

For scanning systems with a control for x- and y-movement, the detector can be a single photodiode or a photomultiplier. The detector determines the sensitivity and can be optimized for the application. The resolution obtainable depends on the scanning system.

One application in microscopy is a confocal laser scanner (Knittel *et al.* 2001, Albrecht *et al.* 2006). Confocal laser scan is used in microscopy for biological samples, because this method provides much better images than does conventional microscopy.

3.08.3 Packaging and Integration

3.08.3.1 Packaging for Image Sensors

The integration of the image sensor chip with other necessary components into a camera or a more complex system such as a cellular phone or an endoscope requires packaging. Although the focus seems to be on improvement of pixel size and performance of the sensor, packaging is one of the most important issues for microcamera development. It affects the system size, the reliability, and the performance of the sensor and is a critical task for cost reduction.

There are various packaging technologies for silicon chips ranging from relatively high-price ceramic package for advanced products to plastic package used for volume production (Menz *et al.* 2001, Tummala 2001). For image sensors, however, many of the standard technologies are not applicable because the photosensitive part of the chip surface has to be exposed to light. Additionally, the image field of the chip must not be damaged or contaminated with particles and, for organic color filters, not treated at temperatures above a limit of approximately 230°C. Thus, many of the standard packaging technologies, mainly the plastic packages common for volume production, cannot be used.

One standard packaging approach is the ceramic package. After thinning the silicon wafer and sawing, the good tested bare dies are placed in a ceramic housing. Depending on the need for the device, a conducting or insulating glue is used to fix the die. After a heat treatment for the bake of the adhesive, the pad connections on the chip are wire bonded to the pads on the ceramic package (**Figure 15**). Instead of the metal cover, which is soldered on top of the ceramic housing in the standard process, a transparent glass cover is used for image sensors. This glass lid is glued to the ceramic. Often the glass is replaced by a special filter (e.g., an IR filter to reduce the NIR response).

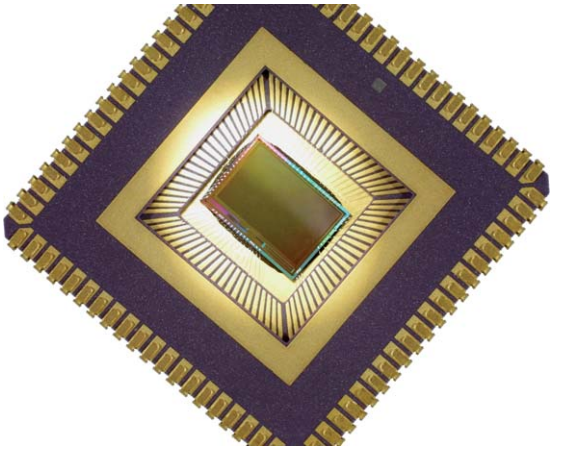


Figure 15 Image sensor in a ceramic package.

The technology is used for many image sensors, mainly in the low-volume and advanced product market. There are, however, a number of drawbacks limiting this approach. First, ceramic packaging is not a low-cost technology. The price for ceramic packages is in the range of US$10–30 per device, which excludes volume markets such as cellular phone cameras. Additionally, many processing steps are critical in terms of particle generation or performance reduction. The processes used in wafer packaging such as wafer thinning, which is done by grinding, generate many particles. In the sawing process, the silicon particles are directly washed onto the chip surface and remain to some extent on the photosensitive area. There are organic protective covers and cleaning procedures, but the removal of both the protective covers and the particles remains a challenge. This is especially difficult for imagers covered with delicate layers of organic color filters or microlenses. These organic top layers can also degrade by thermal treatments such as glue bakes. Particle contamination or even damage of the top layers of the sensor might occur in the pick-and-place process, when the die is placed into the cavity using a front-side tool.

The cover glass lid must be free of particles too. Additionally, the optical parameters of the glass are important for the image quality. Reflections on the glass can cause ghost images, absorption will reduce the sensor sensitivity. Therefore, a strategy to avoid the cover glass is the application of an epoxy resin to protect the wire bonds. The glob top is a semifluid

epoxy material, which is applied on the wire bonds and hardened thermally. Because of the protected bond wires, these packages can be handled without an additional cover. The optimization of the glob top process is not easy and requires special tools and experienced staff.

In a similar way, the image sensor can be mounted as a bare die on a circuit board or ceramic plate. As demonstrated for the miniaturized endoscope sensor, the chip on the ceramic plate is wire bonded to pad connectors on the holder. The glob top cover avoids contacts and damage of the bond wires during further assembly.

For volume applications, the costs of ceramic packaging are not acceptable. One approach is to adapt the plastic package technology to the requirements of the image sensor. Open or semiopen plastic packages have been developed. Elmos Advanced Packaging (www.elmos.de) offers plastic packages with glass openings for optical applications. Another packaging service describes a technology, where a film is used for the cover of the photosensitive surface during molding (www.boschman.nl). The film functions as a seal and protects the image sensor surface. As described for the ceramic packaging, the requirements of image sensors in terms of temperature sensitivity and particle contamination differ from the standard IC processes. For plastic packaging, the molding process has to be adapted to some extent to the needs of the special device, i.e., temperature limitations. Developments in this field are driven by volume market products such as image sensors for cellular phones. Thus improvements in terms of size and cost are to be expected.

For a next generation Nokia cell phone camera there is another approach. In order to meet the size requirements for this mobile phone application, the mounted camera module should have a similar height of about 2.5 mm as the other components on the board (e.g., the display). Chip packages used at present are more than twice the size. If the chip is mounted directly on the board, the height can be reduced to the desired dimensions. That implies, however, the application of the reflow process with high temperatures above 250°C, a risk for the temperature-sensitive organic components (color filter, microlenses) (Nummela *et al.* 2006).

A technology using a glass–silicon–glass sandwich for the package of image sensors has been presented by Shellcase in 2000 (www.shellcase.co.il). The wafers with image sensors are mounted upside down onto a glass substrate and subsequently thinned and separated. Metal interconnects from the front-side pads to the back are fabricated in several steps. The resulting device has a size almost identical to the dimension of the bare die. Additionally, the delicate image sensor surface is protected after the first processing step, which certainly reduces the particle contamination. The complexity of the process, however, seems to limit this approach to applications in advanced products.

Generally, in image sensor packaging front-side wire bonds are disadvantageous for the assembly process. They increase the package size and obstruct simple assembly techniques for the optics. The straightforward approach of flip chip mounting instead of wire bonding is impossible because the photosensitive surface has to be exposed to the light. Thus, a front-to-backside via would be the desired technology for microimager packaging. With a plain front side, the optics can be directly positioned on the image sensor chip. Cost reduction in assembly with plastic or glass optics and a gain in reliability is expected.

Some approaches for front-to-backside interconnects have been described (Rasmussen *et al.* 2003) but the high aspect ratio of the interconnects, which have to be hundreds of microns deep and only 5–50 µm wide, is a challenge. The improvements in MEMS technologies will certainly support process developments in this field.

A wafer-level packaging (WLP) concept seems to be successful in the high-volume market of image sensors (Leib and Töpper 2004). The processed silicon wafers are bonded to a glass wafer and thinned to about 100 µm. Via interconnects to the front-side metal layer and backside bumps are fabricated after wafer thinning (**Figure 16**). The resulting chip package has a similar size as the chip and meets the height requirements for videophone applications.

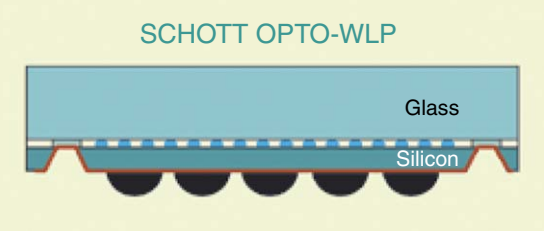


Figure 16 Schematic of wafer-level package for optical devices (Courtesy: Schott).

3.08.3.2 Integration of Micro-optics

Miniaturized cameras frequently consist of small-sized components similar to those used in larger devices. Thus, lenses assembled in lens holders are mounted on the image sensor package located on a circuit board. In the assembly process, the optical system has to be adjusted in order to reach the desired optical performance (i.e., focus and position of the optics). For advanced products such as endoscopes, these optimization processes are done separately for each device.

In a microimage sensor, the optics can be integrated in the cover of the device, e.g., in the metal cap of an endoscope or in a separate optics socket such as the dome in the capsule (**Figure 17**). Examples of integrated optics in capsule endoscopy are shown in Section 3.08.4.4.

For microcameras provided with illumination, special attention has to be paid to the position of the light source, because improper shielding of the sensor reduces the image quality. In volume production, cost and size are the major optimization criteria. Plastic lenses are mainly used. The reduction of the height of the lens stack is a major issue. For cost reasons, automated processes are used for focus and position control. For better manufacturability, wafer-level fabrication technologies for optics have been reported (WLP Optic patent).

One approach used in consumer microcameras today is a combination of digital image data processing (Nummela *et al.* 2006). These opto-digital systems enable a wider focus range without mechanical components (Mait *et al.* 2003). Recently, compound-eye optics have been proposed for image sensors (Duparre *et al.* 2004). These optics are very thin (0.2 mm) compared to the lens stack system, but the performance needs to be improved.

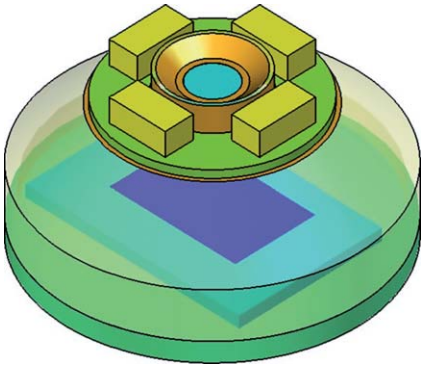


Figure 17 Optical dome with a lens and a circuit board with light-emitting diode (LEDs), a component of a videocapsule.

3.08.3.3 Integration with Micromechanical Components and Electrical Systems

The trend to further reduction of size of the devices will certainly lead to the integration of more functions in the camera chip. The electrical functions can be subdivided into four parts:

(1) Digital circuits
(2) Wireless transceivers
(3) Power systems
(4) Illumination

For CMOS image sensors, the integration of additional digital functions such as signal processing is straightforward. The functional blocks of many of the circuits are available in the CMOS libraries. The increase in chip size and complexity is compensated by the reduction of interconnects and the size of the system. Generally, the reduction of chip-to-chip interconnects increases the reliability of the system because of the higher failure probability of wire bonds. The reduction in size of a one-chip approach compared to a solution with several chips on a board is another argument for the integration. In many CMOS imagers, functions located earlier on the camera board are now integrated on-chip. For applications such as endoscopy, where the size of the camera chip is an important factor for the system size, two or more chip solutions may be advantageous. For CCD devices fabricated using a different technology (see Section 3.08.2.2.1.(iv)), the integration of other functions is not possible.

The benefits of a combination of cameras provided with wireless image data transmission were demonstrated in camera capsules. The integration of these functions into the camera chip has been prevented to date because of the difference in technologies.

Self-powered microcamera systems would be a major step toward autonomous microsystems. The examples in capsule endoscopy (see Section 3.08.4.4) show the technologies available at present. There is a gap between the power consumption of the devices and the energy provided by miniaturized power supplies, but the activities in low-power CMOS process technology and the research in power systems indicate, however, the progress in this field. Some of the power systems such as thermogenerators in silicon technology have the potential to be integrated into the silicon device. Again, the CCD

imagers are unlikely to participate in these developments because of their power consumption.

Only some of the microcameras such as endoscopes need illumination, typically LEDs. These diodes are fabricated in special technologies and mounted on the board. The integration of these devices is unlikely because of the differences in technology. To reduce size, on-chip mounting of bare LED dies would be the best solution.

The major issue for the integration of mechanical components in microimaging systems is the focus and the zoom for the optics. Different approaches have been shown ranging from magnetic drive for zoom lenses (Uehara and Hoshina 2003) to liquid zoom lenses (Berge and Peseux 2000). These components are located on top of the image sensor. Therefore, the integration of the driver circuitry for such a device in the imager is more likely than the integration of MEMS devices.

3.08.4 Microcameras

3.08.4.1 Overview

Digital still cameras and camcorders became products in the 1980s. At that time, the consumer market was still dominated by photographic film and videotube cameras. Miniaturized cameras were exclusively developed in the field of military applications.

Within the last three decades, the feature size reduction in semiconductor process technology was driven by major product memories and ICs. This also enabled developments in market segments of smaller volume such as image sensors. Because of the lower complexity of the process, CCD image sensors were the first to become small enough to be integrated in small-sized cameras. Examples were small systems (so-called finger tip cameras) for various applications such as inspection, technical endoscopy, toys.

The major field of development was medical endoscopy, except for military applications, where microcameras have been used in combination with other devices, but access to state-of-the-art results is impossible. Developments and results in endoscopy are described in Sections 3.08.4.3 and 3.08.4.4.

3.08.4.2 Consumer Applications

In digital photography, the major trends were performance improvements in resolution and sensitivity; the size of the device was a minor topic. For both digital still imagers and videocameras the market has roughly two segments: products for consumer applications and high-performance cameras for professionals. While developments of the latter were concentrated on performance optimization (more and smaller pixels, better sensitivity, color quality, and low noise levels), the cost reduction was a major issue in consumer products, resulting in improved device yield and lower packaging costs. With the technical improvements, devices and systems became smaller and the reduction of size and weight was an important factor in the market. Nevertheless, cameras were far from being microsystems and miniaturization was important but not essential.

The situation changed when cameras where integrated into cellular phones. The commercial success of a mobile phone offering the possibility to make snapshots and transmit the images via a phone to friends started the volume production of microimagers. While the first were popular among Japanese youngsters, today about 500 million camera phones are used (Nummela *et al.* 2006).

As a consequence, miniaturization of devices and components became an issue not only for the semiconductor industry, but also for companies specializing in packaging and optics. Because camera phones are the fastest-growing market for consumer products, their needs determine the tasks for the process development, and the major tasks apart from image quality are size reduction and cost reduction.

The cost per device is determined by the price of the image sensor with integrated electronics, the package with the connector and the optics. In 2006, this price was reported to be about US$5 per camera module (Nummela *et al.* 2006), and further reduction is expected. This demonstrates the need for high-volume and low-cost processes such as automated assembly with plastic packages and lenses. Also processes so far done manually or semiautomatically like quality control or focusing have to be automated. Cost reduction is also the driver for the definition of packaging and interface standards for mobile phones.

The other important parameter, the size of the camera module, does not depend only on the image chip size. Package and optics determine in a similar way the final size of the product. A typical module size for a 1 mega pixel image sensor is 10 mm × 10 mm × 7 mm (Bosiers *et al.* 2006). The package thickness for the next generation of small cameras in the common intermediate format (CIF) or in the videographics array (VGA) format will be only 2.5 mm (Nummela *et al.* 2006), which demonstrates the challenges for size reduction.

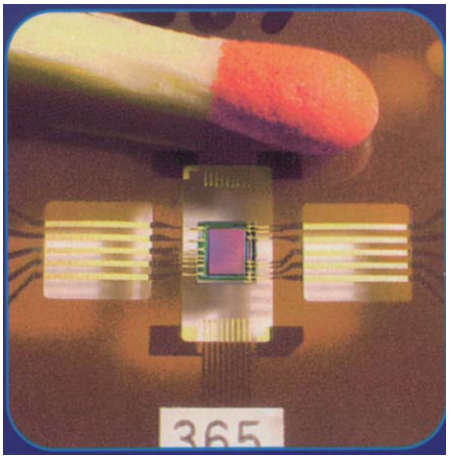


Figure 18 Miniaturized charge-coupled device (CCD) image sensor.

An example of a small-size image sensor in comparison to a match is shown in Figure 18.

3.08.4.3 Videoprobes in Endoscopy

Visualization of the status of an organ's health is one major task in medical diagnosis and therapy. Endoscopy and minimal invasive surgery are techniques for this purpose in medical applications. Endoscopy emerged in the beginning of the 20th century as a method in medical diagnostics and became increasingly important with the development of techniques for minimally invasive surgery.

Most endoscopes use optical lens systems (rigid endoscopes) or fiber bundles (flexible endoscopes) to transmit the images to the viewer's eye or a camera located at the proximal end of the system (Buess 1990). At the distal end, wide-angle optics are used usually in order to obtain images with a reasonable depth of field. Additionally, endoscopes comprehend an illumination source, normally a fiber bundle connected to an external light source, and a number of additional functions (support channels for air or water, biopsy, or medication). There is a special endoscope type for each medical application ranging from colonoscopes for inspection and treatment of the colon (10- to 15-mm diameter) to bronchoscopes for the inspection of the lung with a diameter <5 mm. Fibroscopes used in neurosurgery have diameters in the range of 2 mm and lower. Details on medical endoscopy are given in Buess (1990) and Vilardell (2006), the available instruments and technical details are described by the major endoscope manufacturers (e.g., www.olympus-europa.com/endoscopy/427.htm, www.karlstorz.com, www.richard-wolf.com, www.pentaxmedical.com).

For applications in minimal invasive surgery, more than one endoscope is needed. A typical setup consists of an endoscope with wide-angle optics to have a complete view of the scene and one or more endoscopes for a detailed view. Often stereoendoscopes are used. In some cases, overview and detailed view endoscopes are combined in one instrument, which has a diameter of not more than 15 mm.

Developments in endoscopy have focused on improvements of image quality, on size reduction, and better flexibility of the distal part (Irion 2005). For flexible endoscopes, the resolution is limited by the number of optical fibers used. Endoscopes with diameters of about 4 mm consist of fiber bundles with 30 000 fibers. Thus, there is a trade-off between image resolution and the diameter of the endoscope. On the other hand, thick fiberglass bundles limit the bending property of the system, a drawback for many applications. A typical value for the bending radius of a fiber is 25 times its diameter. Additionally, the loss of image intensity over the distance of up to 2.5 m between optics and camera restricts the sensitivity of the endoscope.

Because of the development of small-size image sensors, the integration of the camera chip into the tip became feasible. Such a videoendoscope or videoprobe contains a microsystem with an image sensor chip, optics and illumination, and electrical wiring at its head. Image data are transmitted via cables, providing the power for the system. The requirements for the medical application, i.e., the secure encapsulation of all microparts and the stability of the whole system in sterilization procedures, are additional challenges for the technical realization.

The first videoprobes demonstrated for this application contained CCD image sensors, which are still used in the majority of products today. In the mid-1980s, the first prototype of a videoendoscope was presented by the American company Welsh Allyn. It had a black and white CCD image sensor with 36 000 pixels. Color images were obtained by sequential color illumination.

Typical CCD sensors in use for videoendoscopes have 480 000 pixels (chip size 4 mm × 4 mm) or 250 000 pixels (chip 2 mm × 2 mm). A drawback, however, is the need for at least 10 external connections (*cf.* Section 3.08.2.2), which makes packaging, wiring, and interconnects more complicated.

Today the complexity of device fabrication, packaging, and system integration limits the

Figure 19 Tip of a laparoscope (10 mm) with a charge-coupled device (CCD) image sensor (chip-on-a-tip, www.richard-wolf.com).

application of these videoprobes to the field of advanced and high-price products (**Figure 19**).

The integration of CMOS sensors into small-size videoprobes has been demonstrated recently (Arena *et al.* 2005, http://ivp.ims-chips.de/start.html). General aspects of the development of a CMOS imager for this application are:

- The minimization of the pixel size to allow acceptable dimensions for the sensor.
- The reduction of the number of pad connections, because bonding pads with its area and the required sizes for pad circuitry cover a significant area of the chip. Additionally, each bond also requires size on the package.
- The minimization of the chip size by area optimization and reduction of functionality.

The reduction of the pixel size is always a trade-off between the complexity of the circuitry and the area reserved for the photosensitive element, the fill factor (*cf.* Section 3.08.2.3). The image sensor developed in the European joint research project IVP (intracorporeal videoprobe) (http://ivp.ims-chips.de/start.html) could be designed with a pixel pitch of 4.6 µm in a 0.25 µm CMOS technology. Although the number of transistors per cell was reduced to the minimum necessary for addressing and readout, the fill factor remains below 30%. The fill factor can be improved by using microlenses (*cf.* Section 3.08.2.3) or the TFC technology.

The number of pad connections is a crucial issue for both the chip and the package size: Each bond pad with a size of 50 µm × 50 µm to 100 µm × 100 µm requires a significant area of the chip with pad driver circuitry, circuits for electrostatic discharge (ESD) protection, and minimal spacing between structures defined by design rules. This area can easily have a size of 150 µm × 300 µm per pad. Besides, the number and position of the bond pads increases the size requirements in the package (*cf.* Section 3.08.3.1). For the IVP image sensor, the number of I/Os could be reduced to four pads (power and signal lines).

The minimization of the overall chip size requires again a trade-off between functionality and size. Although the optimization of the configuration of functional blocks can result in a significant area reduction of the complete chip, the integration of additional functions such as A/D conversion or FPN correction is difficult for these applications. The on-chip A/D conversion is advantageous, because digital signals are less noise sensitive. Chip size restrictions, however, make a two-chip solution favorable in this case. For the IVP sensor the signal output is analog, A/D conversion and other features (FPN correction) are executed in a circuit board located in the hand piece of the videoprobe. This approach enables simple solutions for the micro-camera head; it implies, however, the risk of noise superimposition to the analog signal. Additional screening is necessary to reduce the problem.

The CMOS image sensor developed in the IVP project is an HDRC sensor with a logarithmic response function (*cf.* Section 3.08.2.1). The resulting sensor has an image field of 200 × 180 pixels with a pixel size of 4.6 µm. The overall chip size is 1.7 mm × 1.3 mm and the device has four connections. The image sensor is coated with color filters (red, green, blue). **Figure 20** shows a color image of a human eye obtained with an IVP image sensor. Although the bright reflection in the center seems to be saturated,

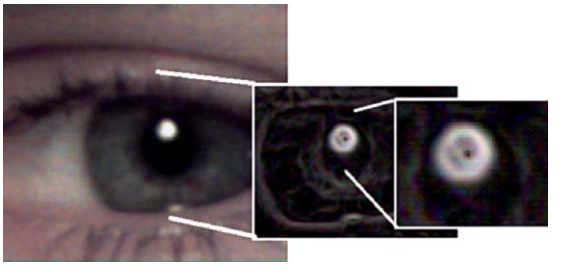


Figure 20 Color image obtained with a miniaturized High Dynamic Range-CMOS (HDRC) image sensor.

more details are visible in the enlarged images, being processed with an edge detection software. The details are the shape of the light bulb used for the illumination during the shot. The image demonstrates the obtainable quality and resolution of the sensor. The complete system is shown in **Figure 21**; details of the image sensor package are shown in **Figure 22**.

Colonoscopy is a very important part of endoscopy due to the high number of patients dying of colon cancer (in Germany about 30 000 patients per year). Early diagnosis could reduce this number significantly since efficient treatment is available. Thus, colonoscopy is a widely used diagnosis, and there are about 25 million colonoscopies worldwide per year. The unpleasant and in some cases painful colonoscopy procedure is one drawback for a wide acceptance among the patients.

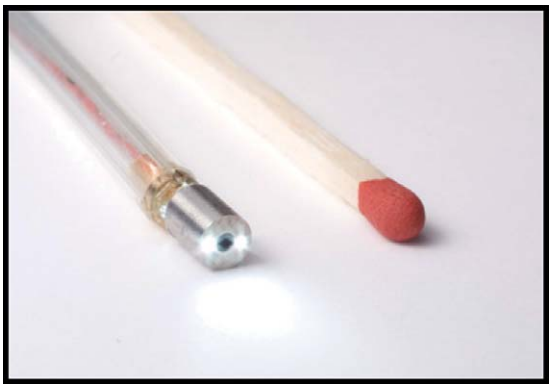


Figure 21 Videoprobe IVP1 (outer diameter 3.5 mm).

Different approaches for advanced colonoscopes using miniaturized cameras have been shown. An instrument intended for single use has been developed by STM Medizintechnik (www.stm-medizintechnik.de). In order to realize this product, the cost reduction of components has been a main objective. The imager is a CMOS sensor with 512×512 pixels and a pixel size of 6 µm. The chip size is 9 mm^2. The sensor is integrated into plastic optics. In the colonoscope, it covers only a part of the system, which has a diameter of 15 mm. It is also equipped with LEDs for illumination and channels for biopsy necessary for the diagnosis. The advantage of the system is not mainly the reduction of size but the single use. The absence of sterilization costs and the higher acceptance by the patients are the major benefits.

The colonoscope robot is another more advanced concept. The focus here is the improvement of the locomotion of the distal part of the colonoscope, one major source of pain for the patient. Colonoscope robots using the inchworm movement principle have been evaluated and first products are entering the market (Dario *et al.* 2004, Kim *et al.* 2006). The camera integrated is a CMOS sensor similar to the ones used in capsule endoscopy.

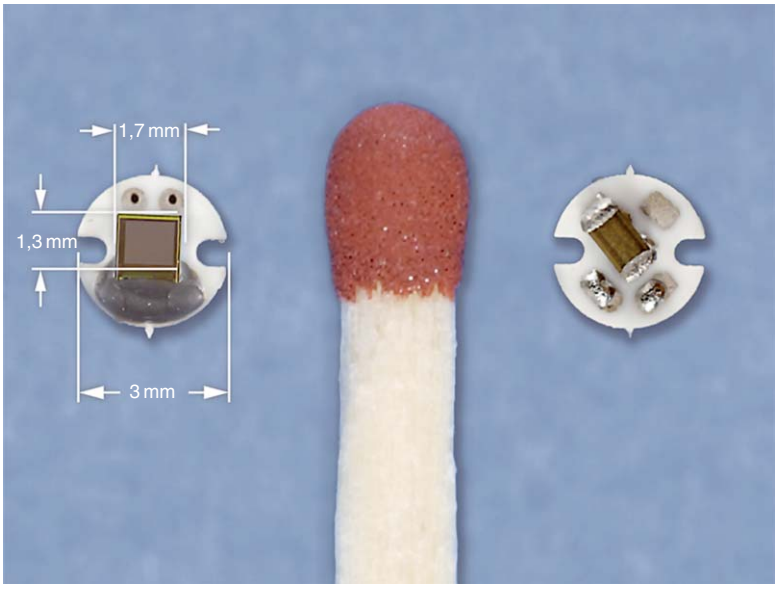


Figure 22 Miniaturized image sensor mounted on a ceramic plate with openings for the fiber optics (left). On the right the backside with four connectors and a capacitor.

The examples of the applications in endoscopy show again the competition between the CCD and CMOS technology. In comparison, the CMOS approach suffers from the bigger pixel size, thus reducing the resolution that can be obtained for small-size imagers. However, the integration of CMOS image sensors in videoprobes has a number of advantages compared to CCDs, which are today certainly the best solution for high image quality. For CMOS devices, the possibility to minimize the number of pad connections and the relatively low cost per device together with improved packaging technologies will make low cost, possibly even disposable videoprobes feasible. The trend to integrate sensors in instruments and the desire to avoid sterilization procedures will be the reasons for proceeding in developments of miniaturized CMOS sensors in videoprobes. Additionally, with the flexibility of the CMOS technology, the possibility to design sensors with special features, and the structure size reduction driven by the mainstream CMOS developments will make these types of microcameras the future devices for videoendoscopy.

3.08.4.4 Videocapsules

A much more sophisticated approach in the field of medical videoprobes is the capsule endoscopy. An autonomous system equipped with a microcamera, a battery or powering system and a transmitter moves through the body and sends image data wireless from inside the body to the outside world. Early examples of wireless capsules for capsules for medical diagnosis were shown in the 1950s (the radio pill) for pressure measurements and in the mid-1960s for pH detection.

The integration of cameras and the transmission of image data were presented by Iddan *et al.* (2000) and Meron (2000). The product based on the invention, the M2A capsule fabricated by the Israeli Company Given Imaging Ltd., started clinical trials in 2000 (www.givenimaging.com). After approval in various European countries, the US (FDA approval in 2001) and Australia, the technology attracted attention worldwide. Already by the year 2004, over 122 000 patients were treated with the capsule.

The M2A capsule has a diameter of 11 mm and is 23 mm long. Its contents are a color CMOS image sensor with optics and camera boards, four LEDs for illumination, an antenna, a transmitter, and two batteries. The CMOS image sensor has a resolution of 256×256 pixels with a pixel size of $10\,\mu m \times 10\,\mu m$. The sensor with an on-chip A/D converter has a power consumption of 2 mW (Krymski 2002). The principle of the capsule is shown in **Figure 23**, a schematic of the new capsule PillCam™. As shown in the figure, image sensor packaging and system integration have been challenges in the development.

During an operation two images per second are transmitted. The overall operating time of the capsule is limited by the battery power; it has been reported to be 7–8 h. The electronic devices are sealed in a biocompatible capsule cover. The capsule is designed for single use; a recovery after use is necessary because of the batteries. The receiver is located in a belt, which also contains a storage system. Image sequences can be analyzed by the physician after the patient's data have been transferred to a computer processing the images.

The major application of capsule endoscopy is diagnosis of diseases in the small intestine. The movement of the videoprobe is caused by the natural

INSIDE THE PILLCAM CAPSULE

1. Optical dome
2. Lens holder
3. Lens
4. Illuminating LEDs (Light Emitting Diode)
5. CMOS (Complementary Metal Oxide Semiconductor) imager
6. Battery
7. ASIC (Application Specific Integrated Circuit) transmitter
8. Antenna

Figure 23 Given Imaging PillCam™ capsule. (Given Imaging GmbH, Hamburg, Germany; reproduced with permission.).

contraction of the intestine. No locomotion is provided. Thus, the direction and the orientation of the capsule during movement are not controlled. Recently, Given Imaging presented a capsule for the diagnosis in the esophagus. This PillCam ESO capsule is equipped with two image sensors, one at each end of the pill. For the inspection of the esophagus the patient is in a supine position; this reduces the velocity of the capsule and enables, in combination with the possibility to catch two images, the full inspection (www.givenimaging.com). Results for diagnosis in the large intestine have not been shown yet using capsule endoscopy. This might be due to complexity of the organ topography; the battery lifetime is another restriction for this application. Microrobots (*cf.* Section 3.08.4.3) with videocameras and steering options are the better choice for colonoscopy.

Since the first product release in 2000, technical improvements of both the capsule and the peripheral instruments have led to a number of new features. Today the next-generation capsule, the PillCam, with a new generation of CMOS imager provides a higher resolution (**Figure 23**). A smaller videocapsule for children is offered (Arguelles *et al.* 2003) and many additional features and accessories are provided. Because of the success of this approach, capsule endoscopy is an increasingly accepted technique in medical diagnosis now.

There are, however, other developments of videocapsules. The Korean Intelligent Microsystem Center (IMC) demonstrated in 2003 a capsule MIRO #1, also with a CMOS image sensor (Park *et al.* 2003). The major difference to the Given Imaging concept is data transmission. The MIRO system monitors the localization of the capsule and transmits images in real time to a PDA. The IMC, which works on robot colonoscopy using inchworm locomotion (Kim *et al.* 2006) (see Section 3.08.4.3), is developing various multifunctional endoscopic microcapsules on the basis of the MIRO #1 system with diagnostic and therapeutic functions and locomotion.

In 2001, the Japanese Company RF Norika presented on the web a capsule with different features (www.rfnorika.com); the details are described in Uehara and Hoshina (2003). A miniature CCD camera has been integrated in a videocapsule based on the developments for space applications. The energy of the capsule is provided by microwave power transmission. The rather high-energy consumption of the CCD camera required special efforts in the power system optimization. The camera system itself consists of the CCD imager with 410 000 pixels, a magnetic focus system for the optics, and four LEDs for illumination. For the CCD, the driver circuitry within the capsule has been reduced to the minimum necessary; functions such as digital signal processing (DSP) are done outside the capsule. In order to increase the field of view, the camera chip is positioned in an angle to the front. The whole inner part of the videopill can be rotated around the axis of the capsule. For this movement miniaturized permanent magnets are located in the pill. The rotation is controlled by three coils located in a vest the patient wears during treatment. Many additional features of the capsule such as integrated tanks for drug delivery or variation of the illuminating light (e.g., by integration of IR LEDs) are reported. Up to now, the product has not been available on the market. In 2006 a new model, the Sayaka capsule, was presented (**Figure 24**).

A capsule development in Europe, the IVP2 system, uses a HDRC camera for image detection (http://ivp.ims-chips.de/start.html). The results obtained in a joint research and development project in the European IST program between 2001 and 2005 (Arena *et al.* 2005) demonstrate the feasibility of this approach. The capsule contains a CMOS image sensor with camera, optics and illumination, a transceiver, a system control with image data compression unit, and a power supply using radio frequency inductive powering (**Figure 25**). The optical part is located on a tiltable plate driven by a wobble motor. The camera chip is a color image sensor with a pixel pitch of 4.6 μm, an address decoder logic, and an integrated A/D converter. The chip dimensions are 4.4 mm × 3.5 mm; it has VGA resolution and provides about 40 frames per second. For illumination four LEDs are used and the wide-angle optics in combination with the movable plate guarantees a field of view of 120°. During the operation the transmission of at least 10 color images per second is needed to secure sufficient information. With about 3 Mbits per image a data compression was necessary to reach the desired frame rate. A special compression chip (ENDOCOM, Turgis and Puers 2004) was developed for this purpose. The system is shown in **Figure 25**.

Recent publications and announcements demonstrate the interest in the field of capsule endoscopy. The Japanese endoscope manufacturer Olympus announced in 2005 a new endoscope pill with many features such as localization and control of the capsule (www.olympus.co.jp/en/news/2004b/

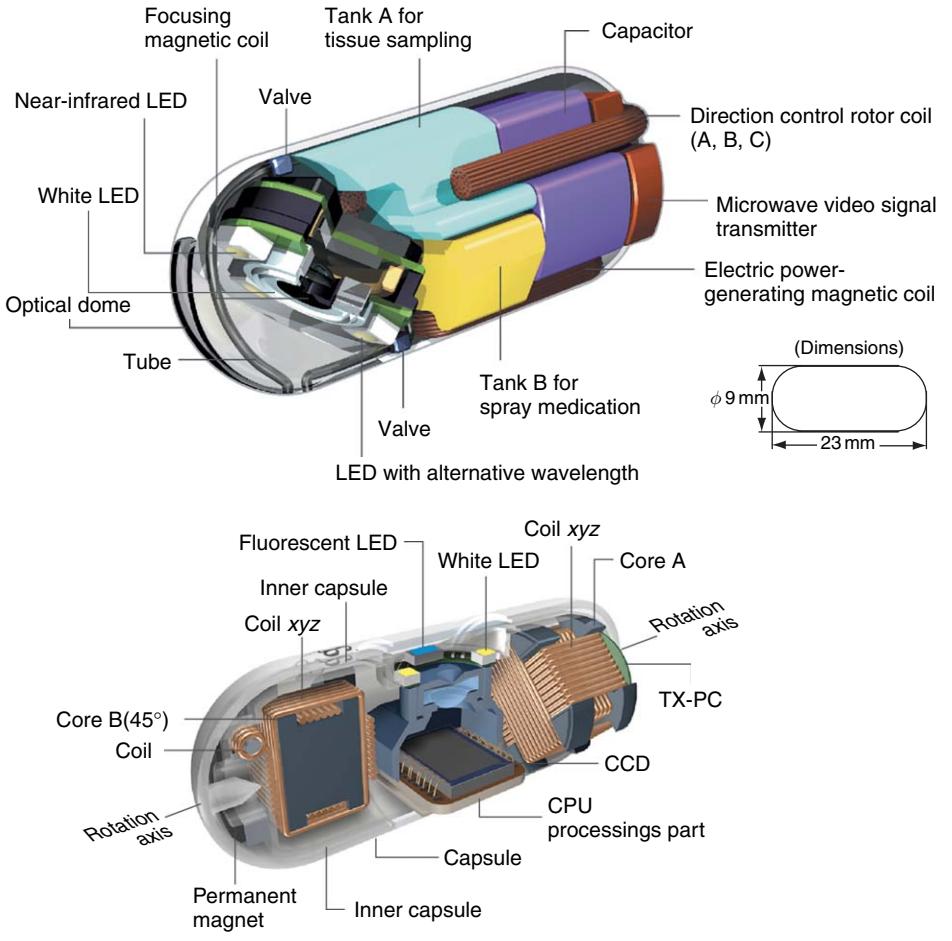


Figure 24 Schematic of RF-NORIKA capsule (top) and the new SAYAKA capsule. (Courtesy: RF System lab).

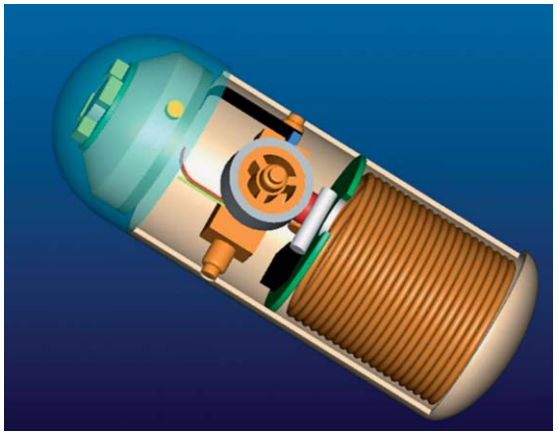


Figure 25 Videocapsule IVP2.

nr041130capsle.cfm). A capsule camera development in China was reported in 2004 (Meng *et al.* 2004).

3.08.5 Summary and Outlook

The development of miniaturized cameras started from products for special applications, such as instruments for medical diagnosis, but the fabrication of products for volume markets was the driver for the fast optimization and size reduction. The input from progress in other technologies, mainly the IC industry and telecommunications, was essential for miniaturization.

The rapid development of microcameras and the trend to integration and miniaturization will continue in future. As demonstrated for the growth of the camera phone market, the combination of technologies generates new types of products and changes the behavior and desires of customers.

The next step for new fields of applications will be the development of autonomous systems with wireless communications. So far the power demand of

circuitry limits the technology, but the progress in low-power CMOS devices and the developments in power generation for small-sized systems indicate that solutions are to be expected in the near future. The integration with MEMS systems and different types of sensors such as chemical or biochemical sensors will lead to many new products.

Wireless systems for health and care with functions for control of medical parameters as well as support for elderly or handicapped people will be a major application. Viewing functions, i.e., cameras, have to be integrated in many of these devices. The same is true for the large market of robotics ranging from production, where cameras are already integrated, to consumer and household applications.

In summary, with the further integration of functions, mainly electrical circuitry, a number of new products are feasible. The future imager will be integrated in smart autonomous systems, which communicate wireless with the outside world. These sensors will most probably be CMOS devices.

References

Albrecht H G, Papageorgas P G, Maroulis D, Theofanous N, Karkanis S, Wagner B, Schurr M, Depeursinge C, Menciassi A 2006 Structure and characterization of an enhanced laser-scanning endoscope based on microelectromechanical mirrors. *J. Electron. Imag.* **15(2)**, 023003

Arena A, Boulougoura M, Chowdrey H S, Dario P, Harendt C, Irion K M, Kodogiannis V, Lenaerts B, Menciassi A, Puers R, Scherjon C, Turgis D 2005 Intracorporeal videoprobe (IVP). In: Bos L, Laxminarayan S, and Marsh A (eds.) *Medical and Care Compunetics* 2. IOS Press, London, pp. 167–74

Arguelles Arias F, Arguelles Martin F, Caunedo Alvarez A, Rodriguez-Tellez M, Herrerias Gutierrez J M 2003 Utility of capsule endoscopy in pediatric gastroenterology. *Ann. Pediatr. (Barc.)* **59(6)**, 586–9

Auth J, Genzow D, Herrmann K H 1977 *Photoelektrische Erscheinungen*. Vieweg, Braunschweig

Berge B, Peseux J 2000 Variable focal lens controlled by an external voltage: An application of electrowetting. *Eur. Phys. J. E* **3**, 159–63

Bosiers J, Dillen B, Draijer C, Manoury E J, Meessen L, Peters I 2006 Imagers for digital still photography. In: Tervonen A, Kujawinska M, IJzerman W, and De Smet H (eds.) *Proceedings of SPIE*. SPIE Bellingham. MA, Vol. 6196, p. 619696–1

Boyle W, Smith G 1970 Charge coupled devices. *Bell Syst. Tech. J.* **49**, 587–93

Buess G, 1990 *Endoskopie-von der Diagnostik bis zur neuen Chirurgie.* Deutscher Ärzteverlag, Köln

Dario P, Ciarletta P, Menciassi A 2004 Modeling and experimental validation of the locomotion of endoscopic robots in the colon. *Int. J. Robotics Res.* **23(4–5)**, 549–56

Duparre J, Dannberg P, Schreiber P, Bräuer A, Tünnermann A 2004 Artificial apposition compound eye-fabricated by micro-optics technology. *Appl. Opt.* **43(22)**, 4303–10

Eberhardt K, Neidlinger T, Schubert M 1995 Three color sensor based amorphous n-i-p-i-n layer sequence. *IEEE-ED* **42**, 1763

Esser L 1972 Peristaltic charge coupled device: A new type of charge transfer device. *Electron. Lett.* **8**, 620–1

Fill factory patents *Eur. Pat.* EP 97870084 (1997), *US Pat.* 6 225 670 (2001)

Fischer H, Schulte J, Giehl J, Böhm M, Schmitt J P M 1992 Thin film on ASIC, a novel concept for intelligent image sensors. *Mater. Res. Soc. Symp. Proc.* **258**, 1139–45

Fossum E 1997 CMOS image sensors: Electronic camera-on-a-chip. *IEEE Trans. Electron Devices* **44**, 1689

Harendt C, Beintner J, Schuhbauer A, Apel U, Dudek V, Graf H G, Höfflinger B, Seger U 1998 Photodiodes in bonded SOI films. *ECS Proc.* **97-36**, 568–75

HDRC 1996, 1997 Deutsches Bundespatent Nr. 42 09 536 filed September 5, 1996, *US Pat.* 5 608 204 filed March 4, 1997

Iddan G, Meron G, Glukhovsky A, Swain P 2000 Wireless capsule endoscopy. *Nature* 405–17

Irion K M 2005 Endoscopes – Potential for microsystem technology. *MST News*, Special Issue 2005 Microsystems Made in Germany, p. 26

Janesik J R 2001 *Scientific Charge-Coupled Devices.* SPIE Press, Bellingham, MA, p. 7

Kasano M, Inaba Y, Mori M, Kasuga S, Murata T, Yamaguchi T 2006 A 2.0-μm pixel pitch MOS image sensor with 1.5 transistor/pixel and an amorphous Si color filter. *IEEE Trans. Electron Devices* **53**, 611–17

Kim B, Lim H Y, Park J H, Park J-O 2006 Inchworm-like colonoscopic robot with hollow body and steering device. *JSME Int. J. Ser. C* **49(1)**, 205–12

Knittel J, Schnieder L, Buess G, Messerschmidt B, Possner T 2001 Endoscope-compatible confocal microscope using a gradient index-lens system. *Opt. Commun.* **188(5–6)**, 267–73

Krymski M 2002 From technology to application. *1st Fraunhofer Workshop on CMOS-Imaging*, Duisburg, Germany

Leib J, Töpper M 2004 New wafer-level packaging technology using silicon-via-contacts for optical and other sensor applications. *Proc. 2004 Electronic Components and Technology Conference*, *IEEE*, pp. 843–7

Lyon R F, Hubel P M 2002 Eyeing the camera: Into the next century. *IS&T/TSID 10th Color Imaging Conf. Proc*. pp. 349–55

Mait J N, Athale R, van der Gracht J 2003 Evolutionary paths in imaging and recent trends. *Opt. Express* **11**, 2093–101

Mendis S, Kemeny S E, Fossum E R 1994 CMOS active pixel image sensor. *IEEE Trans. Electron Devices* **41(3)**, 452–3

Meng M Q H, Tao Mei, Jiexin Pu, Chao Hu, Xiaona Wang, Yawen Chan 2004 Wireless robotic capsule endoscopy: State-of-the-art and challenges. *Intelligent Control and Automation*, WCICA, Vol. 6, pp. 556-555a

Menz W, Mohr J, Paul O 2001 *Microsystem Technology*. *Wiley-VCH*, Weinheim

Meron G 2000 The development of the swallowable video capsule (M2A). *Gastrointest. Endosc.* **6**, 817–19

Neidlinger T, Harendt C, Glöckner J, Schubert M B 1999 Novel device concept for voltage-bias controlled color detection in amorphous silicon based CMOS cameras. *Mater. Res. Soc. Symp. Proc.* **558**, 285–90

Neidlinger T, Schubert M B, Schmid G, Brummack H 1996 Fast color detection with two-terminal P–I–I–N devices. *Mater. Res. Soc. Symp. Proc.* **420**, 147

Nummela V, Viinikanoja J, Alakarhu J 2006 Cameras in mobile phones. *Proc. SPIE* **6196**, 61960B-1–10

Oda M, Kaida T, Izawa S, Ogo T, Itsumi K, Okada Y, Sasada K 2005 A 1/4.5″ 3.1M pixel FT-CCD with 1.56 μm pixel size for mobile applications. *ISSCC Tech. Digest 2005* **19.1**

Pankove J I 1971 *Optical Process in Semiconductors*. Dover Publications, New York

Park Hee Joon, Il-Yong Park, Jeong-Woo Lee, Byung-Seop Song, Chul-Ho Won, Jin-Ho Cho 2003 Design of miniaturized telemetry module for bi-directional wireless endoscopes. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **E86-A(6)**, 1487–91

Rasmussen F E, Frech J, Heschel M, Hansen O 2003 Fabrication of high aspect ratio through-wafer vias in CMOS wafers for 3-D packaging applications. *Transducers '03*, *the 12th Int. Conf. Solid State Sensors*, *Actuators and Microsystems*, Boston, MA, USA

Seger U, Graf H G, Landgraf M E 1993 Vision assistance in scenes with extreme contrast. *IEEE Micro* **13(1)**, 50–6

Seger U, Apel U, Hoefflinger B 1999 HDRC-imagers for natural visual perception. *Handbook of Computer Vision and Application, Vol. 1: Sensors and Imaging*. Academic Press, New York, pp. 223–35

Sze S M 1981 *Physics of Semiconductor Devices*. 2nd edn. A Wiley-Interscience Publication, John Wiley and Sons, New York

Taniguchi M, Takada S 2005 Fuji patent. *US Pat.* 5 0205958

Tredwell T J 1995 Visible array detectors. *Handbook of Optics*. McGraw-Hill, New York, Chap. 22, pp. 22.1–22.37

Tummala R 2001 *Fundamentals of Microsystems Packaging*. McGraw-Hill, New York

Turgis D, Puers R 2004 Image compression in video transmission for capsule endoscopy. *Eurosensors XVIII*, Rome, Italy

Uehara A, Hoshina K 2003 Capsule endoscope NORIKA system. *Min. Invas. Ther. Allied Technol.* **1(1)**, 227–34

Vilardell F 2006 *Digestive Endoscopy in the Second Millennium, from the Lichtleiter to Echoendoscopy*. Thieme Medical Publishers, New York

Weckler G P 1967 Operation of p-n junction photodetectors in a photon flux integration mode. *IEEE J. Solid-State Circuits* **SC-2**, 65

Yadid-Pecht O, Fossum E R 1997 Wide intrascene dynamic range CMOS APS using dual sampling. *IEEE Trans. Electron Devices* **44**, 1721

Zhu Q, Stiebig H, Böhm M 1994 A novel a-SiC:H color sensor array. *Material Research Society Spring Meeting 1994,* Vol. 336, pp. 843–8

Relevant Websites

www.boschman.nl – Boschman.

www.elmos.de – Elmos.

www.foveon.com – Foveon.

www.givenimaging.com – Given Imaging.

http://ivp.ims-chips.de/start.html – IVP.

www.karlstorz.com – Karl Storz.

www.karlstorzindustrial.com – Karl Storz Industrial.

www.photonfocus.com – Photonfocus.

www.rfnorika.com – RF Norika.

www.olympus-europa.com/endoscopy/427.htm – Olympus

www.olympus.co.jp/en/news/2004b/nr041130capsle.cfm – Olympus.

www.pentaxmedical.com – Pentax.

www.richard-wolf.com – Richard Wolf.

www.shellcase.co.il – Shellcase.

www.stm-medizintechnik.de – STM Medizintechnik.

www.tsmc.com – TMSC.

Biographies

Christine Harendt heads the department Add-On Processes at the Institut für Mikroelektronik Stuttgart, Germany (IMS CHIPS). She is involved in the development and application of new technologies in combination with CMOS processes. In 1988, she received her Ph.D. in Physical Chemistry from Freie Universität, Berlin. The same year she joined IMS CHIPS. She has participated in several national and international research programmes (IST, PROMETHEUS, German Research programs). From 2001 to 2005, she coordinated the European Research Project (IVP) dealing with the development of miniaturized videoendoscopes.

Heinz-Gerd Graf heads the Microsystems Department of the Institut für Mikroelektronik Stuttgart, Germany. He was a member of the research staff of the Department of Electrical Engineering at the University of Dortmund and a researcher in the Advanced Physics Group of Messerschmitt-Boelkow-Blohm GmbH. He received the diploma in physics from the University of Dortmund. CMOS Image Sensors, IR Detectors, IC Packaging, and SOI-CMOS technologies are his main research activities.

3.09 Biophotonics

Wolfgang Mönch and Hans Zappe, Department of Microsystems Engineering – IMTEK, University of Freiburg, Freiburg, Germany

Glossary

AFM Atomic Force Microscope
ASIC Application-Specific Integrated Circuit
ATR Attenuated Total Reflection
CCD Charge-Coupled Device
CMOS Complementary Metal Oxide Semiconductor
COC Cyclo-Olefine Copolymers
DFB Distributed Feedback (laser)
DOE Diffractive Optical Element
FRET Förster Resonance Energy Transfer
GRIN Gradient Index
MEMS Microelectromechanical Systems
NA Numerical Aperture
OCT Optical Coherence Tomography
PCR Polymerase Chain Reaction
PDMS Poly(dimethyl siloxane)
PMMA Poly(methyl methacrylate)
PMT Photomultiplier Tube
SLM Spatial Light Modulator
SPR Surface Plasmon Resonance
TIR Total Internal Reflection
TIRF Total Internal Reflection Fluorescence
VCL Vertical Cavity Laser

3.09.1 Introduction

Biophotonics is concerned with the interaction of light and biological material. While the use of light for analysis, manipulation, and modification of biological specimens, from cells to organisms, has a long history, developments in microsystems engineering have made new types of optical components available, and also allow strong miniaturization of the optical systems involved (Motamedi 2005). The fields of photonics (Iizuka 2002), frequently defined as that in which lasers and semiconductor-based components, such as laser diodes, photodetectors, modulators, and waveguides, play an essential role, along with that of micro-optics (Herzig 1997, Sinzinger and Jahns 2003), have thus become of great importance in biology. As a result, the term biophotonics is used to refer to optical systems for biological analysis and modification, which employ laser technology, with particular emphasis on miniaturization and incorporation of microelectromechanical systems (MEMS) technologies.

This chapter provides an overview of microsystems technologies as they are applied to biophotonics (Prasad 2003). In correspondence to the focus of *Comprehensive Microsystems*, the coverage is restricted to those aspects of biophotonics in which microsystems play a decisive role; macroscopic imaging or laser-based systems, while of great significance to the field, are not considered here. Other chapters in *Comprehensive Microsystems* that deal with biophotonics include Chapters 3.06 and 3.14.

Beginning with an overview of the nature of the interaction of light with biologically relevant materials, ranging from individual molecules and cells to tissue and entire organs, a brief summary of the optical techniques employed by the biologist or the physiologist for analysis and alteration is provided. Following this introductory summary, the focus is on four areas in which biophotonics has seen particularly rapid and fruitful development:

1. Microscopic imaging systems, which take advantage of advances in micro-optics to allow the fabrication of strongly miniaturized microscopes and interferometers.
2. Optical biosensors, which frequently employ integrated or fiber optical structures for the measurement of biologically relevant parameters

3. The combination of microfluidics and photonics technology for enabling flow cytometry, essential for highly parallel measurement of individual cells.
4. Microarrays, the workhorses for genomic or proteomic analysis, in which highly parallel measurement of fluorescence is accomplished by means of the so-called biochips, in which the complete system has benefited from miniaturization of its components.

3.09.2 Photobiology

This section gives a brief description of the interaction of light with biologically relevant matter. In the context of biophotonics, light means electromagnetic waves from the near-ultraviolet, over the visible and near infrared, to the far-infrared, or Terahertz, frequency range. The interaction of light with biochemical and biological species of interest is described as the interaction of photons with a complex molecular energy level structure, including electronic, vibrational, and rotational levels. Virtually all these molecular energy levels play a role in biophotonics and are used for qualitative and quantitative analysis of biochemical species. For a description of the interaction of a system with discrete energy levels with a light field, only two energy levels directly involved are conveniently considered. Einstein identified three fundamental processes, i.e., absorption, spontaneous emission, and stimulated emission, in the interaction of a photon with a two-level system, which occur when the energy $h\nu$ of the photon equals the difference in the energies of the upper (E_2) and lower (E_1) electron state. Of these processes, only absorption and spontaneous emission are considered in this section.

3.09.2.1 Absorption

In the absorption process of a single molecule, the energy of an incoming photon is transferred to the molecule, leaving the molecule in its electronically excited state. A light wave traveling through a sample of thickness x containing an ensemble of absorbing molecules with a concentration c experiences attenuation of its initial intensity I_0. The intensity after passing through the sample, $I(x)$, is given by Beer–Lambert's law:

$$\frac{I(x)}{I_0} = \exp\{-\alpha(\lambda)x\} = \exp\{-\zeta(\lambda)cx\} \qquad [1]$$

Here, α is the absorption coefficient (cm^{-1}). In analytical chemistry, the molar extinction coefficient ζ (in units of $\mathrm{l\,mol}^{-1}\ \mathrm{cm}^{-1}$) is generally preferred over the absorption coefficient α, because the concentration of the molecules c can be determined from an absorption spectrum when ζ is known. Both the absorption coefficient and the molar extinction coefficient show a strong dependence on the wavelength λ. For typical biological materials, such as tissue and blood, their values differ to a great extent among individuals and with environmental conditions. Absorption spectra of water, oxygenated, and deoxygenated hemoglobin are given in Figure 1.

3.09.2.2 Fluorescence Emission

When a molecule is excited, a number of decay processes of the excited state decay due to the complex energy diagram of molecules. In all these mechanisms, conservation of energy, momentum, and angular momentum have to hold. This implies that not all transitions are necessarily accompanied with photon emission. Fluorescent dyes are frequently used in biophotonics to label molecular species of interest; examples of typical absorption and emission (fluorescence) spectra of two rhodamine dyes are shown in Figure 2.

Fluorescence is a radiative transition from a vibrational level of an excited electronic state (usually the first) into a vibrational level of the electronic ground state of a molecule (Figure 3). The emission spectrum of a fluorescence transition is usually shifted to longer wavelengths compared to the absorption spectrum,

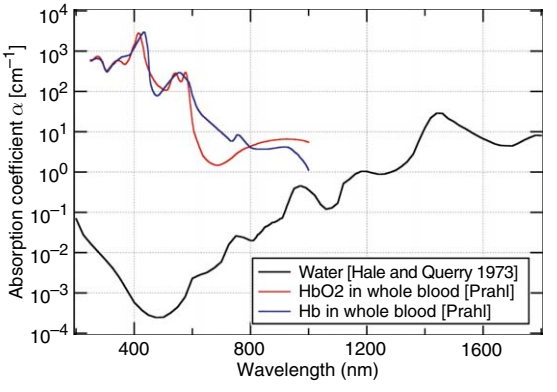


Figure 1 Absorption spectra of water (Hale and Querry 1973), oxygenated and de-oxygenated hemoglobin (Prahl 2006) in the visible to near-infrared wavelength ranges.

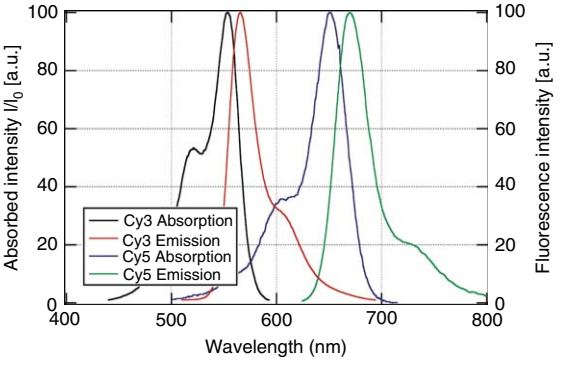


Figure 2 Absorption and emission (fluorescence) spectra of two rhodamine dyes.

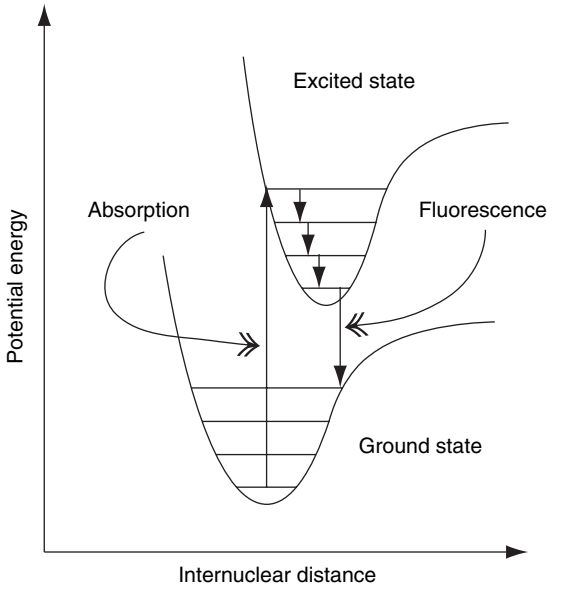


Figure 3 Absorption and fluorescence due to absorption and emission from the energy levels of a typical molecule.

because some vibrational energy in the excited electronic state is dissipated to the environment. Hence, the upper level in fluorescence emission is usually the lowest vibrational level in the excited electronic state (Atkins and de Paula 2006). This so-called Stokes shift between the maxima of the absorption and the emission spectra is typically some tens of nanometers in wavelength. The excitation energy of a fluorescent molecule can be provided not only by direct irradiation but also by coupling an evanescent light field to the fluorescent molecule; this mechanism is discussed in Section 3.09.2.4.

Practically every material shows endogenous fluorescence in the UV spectral range. Therefore,

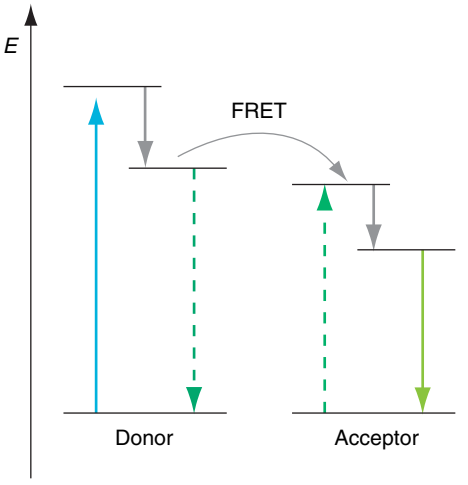


Figure 4 The FRET process: the energy from an excited donor molecule is transferred to a second fluorescent molecule in close physical proximity.

materials for fluorescence-based systems have to be chosen carefully to reduce background fluorescence. In that respect, COC have proven to be superior to glass as substrate materials for many biophotonics applications. On the other hand, endogenous fluorescence can be advantageously used for label-free fluorescence analysis.

Collision of an excited molecule with molecules in its surroundings can lead to dissipation of the excitation energy without photon emission. This effect is known as fluorescence quenching and is usually considered detrimental in biophotonic analysis.

Organic fluorescence dyes generally suffer from a limited lifetime due to photobleaching, i.e., a photon-induced irreversible damage of the molecule. Therefore, inorganic dyes such as semiconductor nanoparticles (quantum dots) are currently much in favor. Besides their stability, semiconductor nanoparticles as fluorophores offer the advantage of tailoring the emission and the absorption spectrum by their size.

Nonradiating transfer of excitation energy between molecules has a prominent application in Förster (or fluorescence) resonant energy transfer (FRET). In this technique, a primarily excited donor molecule transfers its excitation energy to an acceptor molecule in its vicinity. In a simple classical model, illustrated in **Figure 4**, both the donor and the acceptor molecules are described as interacting oscillating dipoles with characteristic r^{-6} dependence of the interaction energy on the distance r between both molecules. The efficiency of the FRET process depends furthermore on the relative orientation of the donor emission

dipole moment and the acceptor absorption dipole moment and on the spectral overlap of the donor emission spectrum and the acceptor absorption spectrum. The latter condition affects the appropriate selection of donor and acceptor molecules for a particular application. Because of the strong dependence on distance and orientation, the FRET process is efficient only for very close distances in the range of a few nanometers. FRET has turned out as a powerful tool for spectral imaging in biomedical research. Its main utility is in the visualization of interactions between proteins and nucleic acids in fluorescence microscopy and in the quantitative measurement of the concentration of nucleic acids.

3.09.2.3 Scattering

From a physical point of view, scattering is a very complex problem and is dealt with in standard textbooks on optics (Born and Wolf 1999) and a number of monographs (Bohren and Huffman 1998, van de Hulst 1982). We restrict ourselves here to a few aspects of relevance to biophotonics. Scattering is a frequently encountered effect in biophotonics; many natural materials (of which blood and tissue are the only two prominent examples) contain soft particles (micelles, vesicles, or cells) with a typical size in the micrometer range and with dielectric properties different from the surrounding host material; these act as scattering centers when light is transmitted through such materials.

Scattering generally occurs when an incident electromagnetic wave interacts with a particle of a size comparable to the wavelength λ. In the simplest case, a scattering particle can be modeled as an infinitely small oscillating dipole driven by the electromagnetic wave. This is a reasonable model for scattering particles of a size much smaller than the incident wavelength (known as Rayleigh scattering). The scattered field then has a typical dipole radiation characteristic of scattered intensity proportional to λ^{-4}. For scattering particles of a size comparable to or larger than the incident wavelength, full Mie scattering theory is taken into account. In Mie theory, the scattered intensity and polarization depends in a complicated way on the ratio of particle size and wavelength and on the scattering angle. For both these scattering mechanisms and nonfluorescent materials, the scattered light is at the same frequency as the incident light.

The intensity and polarization of the scattered wave depend strongly on the optical material properties, size and shape of the scattering object, and on the direction and angle of observation; possible absorption mechanisms can further complicate the problem. A rigorous treatment of the scattering problem of metal spheres is presented in Born and Wolf (1999). Unfortunately, scattering objects such as erythrocytes exhibit an even more complicated scattering behavior due to their irregular shape. If only the extinction of a light beam by an absorbing and scattering sample is considered, the effect of scattering can be summarized in a scattering coefficient $\sigma\,(\mathrm{cm}^{-1})$, which is added to the absorption coefficient α in eqn [1].

The frequency of scattered radiation can be different from that of the incident wave due to a number of inelastic scattering processes. An important example is Raman scattering: here, a molecule is excited to an unspecified (virtual) electronic state with a lifetime of zero (Atkins and de Paula 2006). From this excited level, radiating transitions occur to a vibrational or a rotational level of the electronic ground state. The energy difference of the scattered radiation is provided by or absorbed by the energy present in the molecular rotations or vibrations. The Raman effect is frequently used in spectroscopy and is as such a means for discrimination and identification of molecules; Raman microscopes that provide optical imaging and spectral information are commercially available.

3.09.2.4 Reflection

A frequently encountered situation in biophotonics is that of a propagating electromagnetic wave impinging onto the interface between two media, e.g., tissue layers as imaged by optical coherence tomography (OCT). The simplest way to model this is by a layer sequence of homogeneous and isotropic dielectrics with a certain refractive index; the interfaces between these layers then give rise to reflection.

At every interface between two media, with refractive indices n_i (from which the wave is incident) and n_t (into which the wave is transmitted), part of the incident light is reflected and the reflectance R defines the intensity of the reflected light as a fraction of the intensity of the incident light. In the case of normal incidence onto the interface, i.e., at an incidence angle θ_i of zero, R is given by

$$R = \left\{\frac{n_\mathrm{t} - n_\mathrm{i}}{n_\mathrm{t} + n_\mathrm{i}}\right\}^2 \qquad [2]$$

and for arbitrary incidence angles, the Fresnel equations define the resultant reflectance (Hecht 2003).

In addition, in the case of non-normal incidence, i.e., for $0 < \theta_i \leq 90°$, the angle of the transmitted beam with respect to the interface normal is described by Snell's law:

$$n_t \sin\theta_t = n_i \sin\theta_i \quad [3]$$

for which the special case, when $n_t < n_i$, is of particular interest. In that case, for a range of input angles, θ_i, greater than a critical angle, defined as

$$\sin\theta_c = \frac{n_t}{n_i} \quad [4]$$

the transmitted angle is predicted to be $\theta_t > 90°$. This situation is known as total internal reflection (TIR) implying that the incident wave is completely reflected at the interface. Nevertheless, there is an electric field behind the interface, the so-called evanescent field, which decreases exponentially with increasing distance normally from the interface (Born and Wolf 1999, Hecht 2003). Associated with the exponential decrease is a characteristic length or penetration depth, which is of the order of the wavelength, typically several hundred nanometers.

In the evanescent field, no net energy is transferred across the boundary to the less dense medium. However, if fluorescence dye molecules are present in the vicinity of the interface (i.e., within the penetration depth of the evanescent field), energy is transferred from the incident beam to the fluorophores by the evanescent field, as is shown schematically in Figure 5. The fluorophores are excited and as a consequence, the intensity of the reflected beam is smaller than that of the incident beam. This situation is called frustrated or attenuated total reflection (ATR). Total internal reflection fluorescence (TIRF) microscopy is based on this mechanism. It is a powerful technique for fluorescence imaging of processes in biophotonics with a pronounced surface sensitivity due to the exponential dependence of the evanescent field on the distance from the interface.

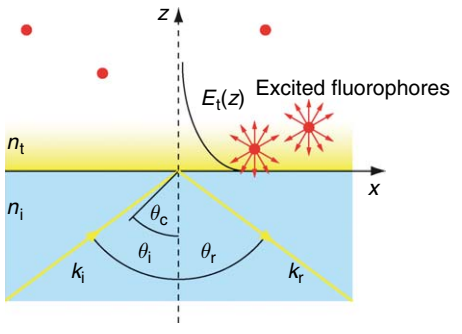


Figure 5 Total internal reflection (TIR) at an interface between two dielectrics; the material with the higher refractive index is at the bottom. The evanescent field extends beyond the interface and can be used to excite fluorescent molecules.

3.09.2.5 Optical Forces

In many applications in biophotonics, molecules, cells, small organisms, and tiny amounts of liquids have to be manipulated. Well-established standard tools for the exertion of forces to such soft or biological micro-objects are optical traps and optical tweezers based on dipole forces and radiation pressure (Ashkin 1992, Visscher and Brakenhoff 1991, 1992). Since these are macroscopic systems and well known for two decades they are not discussed here in more detail. Instead, the focus of this section is on the recently developed techniques based on wetting and surface forces.

In all the techniques for transportation of liquids involving wetting, the interfacial energies of the involved surfaces play a decisive role. In thermal equilibrium, the contact angle θ of a liquid drop on a planar surface is given by the balance of the interfacial energies γ_{sg} (substrate–gas interface), γ_{sl} (substrate–liquid interface), and γ_{lg} (liquid–gas interface; this is the surface tension of the liquid) by Young's equation

$$\gamma_{sg} - \gamma_{sl} = \gamma_{lg}\cos\theta \quad [5]$$

This balance of interfacial energies can be influenced by electric fields, as shown in Figure 6. In the electrowetting on dielectrics (EWOD) setup (in Figure 6, *left*), a conductive (or polar) liquid rests

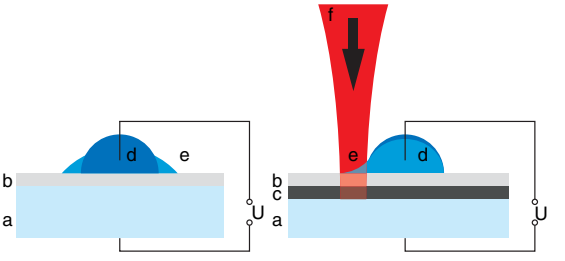


Figure 6 (*Left*) Principle of electrowetting and (*right*) optoelectrowetting; a, conductive substrate; b, insulating coating; c, photoconductive layer; d, liquid at zero voltage; e, actuated liquid droplet; f, light beam; U, driving voltage.

on an insulating layer (thickness d and dielectric constant ε) deposited on a conductive substrate. If a voltage V is applied, a term accounting for the electrostatic energy stored in the system must be added to γ_{sl}, namely

$$\gamma_{sl} \rightarrow \gamma_{sl} - \frac{1}{2}\varepsilon\varepsilon_0 \frac{V^2}{d} \quad [6]$$

By inserting this expression into Young's equation (eqn [5]), the Lippmann equation is obtained, describing the contact angle as a function of the applied voltage V as

$$\cos\theta(V) = \cos\theta(V=0) + \frac{\varepsilon\varepsilon_0}{2\gamma_{lg}} \frac{V^2}{d} \quad [7]$$

Here, $\theta\,(V=0)$ is the zero-voltage contact angle as given by Young's equation. Eqn [7] shows that the contact angle can be reduced by the applied voltage. Since a lower contact angle always means an energetically more favorable state for a droplet, a droplet always move toward those regions on the substrate on which its contact angle is smaller. For a recent review on electrowetting, see Mugele and Baret (2005).

A local reduction of the contact angle can be accomplished by using an array of individually addressable electrodes under the insulating layer and can be used for moving droplets across a surface along a path defined by the electrode array by wetting forces. Another possibility of moving droplets across a surface but without an array of electrodes is by optoelectrowetting (OEW). The experimental setup of OEW (see **Figure 6**, *right*) is a modification of the EWOD configuration with an additional photoconductive layer (e.g., hydrogenated amorphous silicon) deposited between the substrate and the insulating layer. By irradiation with light, the conductivity layer of the photoconductive is increased by several orders of magnitude such that the effective thickness of the insulator is that of the insulating layer and the contact angle of the liquid is reduced by the applied voltage. In contrast, the photoconductive layer acts as an additional insulator in the dark state such that the electrical field in the capacitor is small and does not significantly affect the balance of interfacial energies. Therefore, the contact angle is not changed in the dark state.

In summary, OEW can be used for transportation of liquid droplets by light-induced surface forces on a substrate. Since micromachined spatial light modulators (SLMs) are becoming commercially available, OEW is an attractive effect for use in MEMS-based biophotonics. Experimental demonstrations of electrowetting approaches for droplet manipulation are discussed in Section 3.09.6.3.

3.09.3 Macroscopic Biophotonic Systems

Photonic systems employ light for analysis or for modification of biological material. The use of laser technology for biological applications is virtually as old as the laser itself, and many of the applications that employ these components are decidedly macroscopic. A number of areas in biophotonics have profited from developments in microsystems and MEMS technology. In particular, miniaturized imaging systems and optical sensors have seen significant developments in this regard, as have new microsystem approaches for flow cytometry and microfluidics-based microarrays. The following sections provide an overview of some of these microbiophotonic systems.

3.09.3.1 Imaging

One of the first uses of the microscope developed by Anton van Leeuwenhoek (1632–1723) was to image the complex biological system found in a drop of pond water. Microscopy remains one of the most essential tools in biology, as it allows 2D and 3D imaging of biologically relevant samples. Illumination is by incoherent or coherent sources and the image is formed by reflection or transmission. Limitations in feature resolution are due to the illumination wavelengths employed, possible lack of contrast between the feature and its background, and scattering from the surrounding material, the latter especially if the feature of interest is below the surface.

Microscopy techniques have been developed to partially offset these inadequacies; dark field, phase contrast, or interference contrast microscopes employ scattered light to reveal features with too little contrast using conventional imaging. The confocal microscope (Semwogerere and Weeks 2004) uses a pinhole in the optical path to limit the depth of the field of the image, thereby strongly reducing scattering from layers above and below the feature to be imaged. A variation of this approach is 4Pi imaging, in which confocal approaches are used for illumination and imaging (Hell and Stelzer 1992) (the technique is called 4Pi microscopy because the

full solid angle ($= 4\pi$) of the wavefront generated from a fluorescent object is collected by the setup. The setup includes at least two microscope objective lenses). Miniaturized confocal microscopes are discussed in Section 3.09.4.3.

Fluorescence imaging restricts the image to that of a fluorescent feature, so that either the autofluorescence of cellular components or tissue may be selectively imaged or regions with certain functionality may be stained using fluorophores and thus made visible. The excitation illumination is filtered out and the result is strongly increased in contrast. Imaging using FRET takes advantage of the energy transfer between an optically excited molecule and the one at a lower energy level, thereby permitting an analysis of molecular functionality based on the proximity to known fluorophores. Multiphoton microscopy results in images only from regions in which two or more photons need to be absorbed to generate the detected fluorescence. Two-photon absorption is strongly intensity-dependent, so that optics with a narrow depth of field can be used to image features embedded in the scattering tissue; however, high-power pulsed laser sources are generally required. Both these latter techniques provide enhanced contrast and may reveal features not otherwise visible; they may also allow functional imaging.

A further means to image features below the surface of a scattering medium is through OCT in which a low-coherence light source is used in a Michelson interferometer in which the sample is placed in the measurement arm (Bouma and Tearne 2002). The reference arm contains a scanning mirror such that a detector signal is given only when the arms are balanced and is thus restricted to emission from a single point; the scan then allows slice-by-slice imaging of several millimeters below the surface with micrometer resolution. Optical coherence tomography systems have benefited greatly from advances in microsystems.

Optical near-field techniques use wavelength-sized apertures to circumvent the diffraction limit and image features smaller than the illumination wavelength (Betzig and Trautman 1992). When either the illumination or the image is situated close (within a wavelength) to the aperture, typically a lensed and apertured fiber with an opening in the 50 nm range, imaging is accomplished in the optical near field; resolution better than 100 nm is achievable.

3.09.3.2 Spectroscopy

Measurement of the absorption, transmission, or fluorescence spectra of biological materials (as discussed in Section 3.09.2) is often a direct approach to determine their composition. Whereas standard imaging microscopy provides structural information, spectroscopy can yield compositional data and provides clues as to the functional characteristics of the sample.

Two approaches for spectroscopic analysis can be defined. The first approach uses natural absorption or emission features, inherent to the molecules being measured. All molecular species have absorption spectra corresponding to their structure; most of these are in the near- to mid-infrared frequency range, and manifest themselves as distinct absorption lines at certain wavelengths. Alternatively, some materials show natural fluorescence emission, excited at higher energies and the characteristic emission lines can be used to analyze the sample.

The second approach uses fluorescence from specialized marker molecules or dyes that are selectively hybridized with the biological compound of interest. The fluorescent markers are typically attached to molecules that selectively bind to components in the analyte, thus providing an optical signal proportional to its concentration. This method is widely used for materials that have no easily excitable natural fluorescence.

Spectroscopic analysis can be imaging or nonimaging. The former case corresponds to microscopy approaches in which fluorescent emission is mapped two- or three-dimensionally using either steady-state or time-dependent analysis; a precise spatial distribution of a particular analyte in a larger sample (such as tissue) can be determined. Nonimaging spectroscopy relies on absorption in or emission from the bulk of a sample, e.g., oxygen in a gaseous medium (Zappe *et al.* 2000). Scanning probes allow an analysis of spatial distribution, and interferometric approaches are often used to improve sensitivity. For materials with multiple emission lines, spectroscopic analysis relies on dispersive or wavelength-selective elements, such as bandpass filters, gratings, or prisms, in the optical system to allow spectral separation of various fluorescent signals.

3.09.3.3 Microarrays

At present, one of the most industrially and scientifically significant areas in biophotonics is that of genomic or proteomic analysis. Significant progress in DNA sequencing and in the analysis of the

genomic structure of humans and other species has been made using microarrays. As discussed in Section 3.09.7 in greater detail, microarrays are 2D arrays of immobilized DNA sequences or proteins that selectively bind components of a fluorescence-marked analyte. Upon illumination, optical imaging or scanning measurements are used to determine the positions of fluorescence.

A microarray, sometimes referred to as a biochip, consists of a glass slide on which a 2D array of DNA fragments is immobilized; the spots, formed by variants of ink-jet printing (Chapter 3.11), are typically 50–100 μm in diameter and an array typically consists of 1000–10 000 spots. The sample to be analyzed consists of mRNA or cDNA fragments, labeled with a fluorescence marker; following hybridization of the sample fragments with the spots, the entire chip is imaged and the distribution of fluorescence peaks is determined, thus allowing conclusions to be drawn concerning the structure of the analyte protein or DNA.

A wide range of variations on this approach has been developed, including the photolithographic assembly of oligonucleotides on the microarray itself or hybridization that takes place in on-chip cavities using advances in microfluidics (Whitesides 2006) and lab-on-a-chip technology as discussed in Chapter 2.13. In addition to the analysis of DNA sequences, antibody or protein sequences can be determined, and highly parallel microarray characterization of cells and tissues that express certain DNA sequences is under development.

3.09.3.4 Sensors

Optical sensors form one of the mainstays of biophotonics. Determination of the presence of and the concentration of biologically relevant compounds using optical techniques also has a long history and miniaturization has allowed implementation of optical biosensors in applications not previously feasible. A more general discussion of microsystems-based sensing can be found in Chapters 2.11 and 2.12.

With the notable exception of gas sensing systems based on absorption spectroscopy as discussed in Section 3.09.3.2, which are frequently based on free space optics, most miniaturized optical sensor systems are based on guided wave optics; this technology is the subject of Chapter 3.05. Fiber- or waveguide-based optical fields can be used for excitation and detection of fluorescence, and evanescent wave-based configurations (recall Section 3.09.2.4), employing unclad waveguides or sections of stripped fiber, are employed for the measurement of refractive index changes. Attenuated total reflection (ATR) or surface plasmon resonance (SPR) concepts also employ an evanescent optical field, but they are generally macroscopic approaches with limited prospects for miniaturization using MEMS technologies.

Integrated optical or fiber-based guided wave systems can benefit greatly from interferometric measurement approaches. Michelson or Mach–Zehnder type interferometers can be miniaturized and fabricated in arrays, and the sensitive phase measurements possible with these can be used for high-resolution optical sensing of biological or chemical parameters. These systems allow on-chip referencing and highly integrated approaches provide the means for temperature compensation and exclusion of nonspecific changes in phase.

Transduction of optical biosensors is accomplished using a wide variety of methods; the simplest method is by the measurement of an unspecific change in refractive index, a measurement which is of limited value without sensor calibration. Concentration measurements of specific compounds are generally accomplished by selective binding of analytes to molecules immobilized on the waveguide core. The change in effective waveguide index due to adhesion of a surface layer can be measured with high sensitivity using an interferometric approach. Likewise, fluorescence intensity, either autofluorescence or that of dyes, or change in optical absorption or polarization can be used for optical biosensing. The details of microsystems-based optical biosensors are discussed in Section 3.09.5.

3.09.3.5 Optical Tweezers

Of increasing importance in macroscopic biophotonics systems is the use of light fields for manipulation of microscopic particles, including cells and organelles (Rohrbach *et al.* 2003). Using relatively high-power laser light sources, generally Nd:YAG emitting at 1064 nm or diode lasers with emission in the 780 nm range, tightly focused light fields can be used to capture, move, and analyze cells or their components; the effects considered in Section 3.09.2.5 can also be employed. Using advanced optics and data evaluation techniques, the elastic properties of cells can be measured directly.

Chiou *et al.* (2005) have shown that directly or holographically generated light field distributions can be used to guide, manipulate, and ultimately

sort particles such as cells, suggesting future directions for optical flow cytometry. Optical tweezers and related techniques are undoubtedly promising for a variety of biophotonics applications, but the prospects of benefit from microsystems technologies are limited.

3.09.3.6 Tissue Engineering

A final area in which biophotonics has become firmly established is in tissue engineering, a collective term implying slicing, drilling, dissecting, cauterizing, coagulating, or welding tissue using intense light fields. The relevant technologies, generally employing truly macroscopic laser systems and frequently requiring femtosecond laser pulses, have become standard in various surgical specialities, predominantly ophthalmology.

Related to radical surgical approaches, photodynamic therapy uses considerably lower optical powers for destruction of undesirable tissue, generally tumors. These selectively absorb the illumination, low-power lasers or high-power lamps, due to the presence of a sensitizer, which is optically activated. Activation yields a high concentration of elemental oxygen leading to destruction of the tumor.

As high optical powers and macroscopically sized delivery systems are required, tissue engineering also remains a facet of biophotonics just beginning to benefit from MEMS technology in the near term.

3.09.4 Microimaging

Optical imaging is one of the most significant areas in biophotonics: the use of light to create an image of a biological specimen ranging in size from a molecule to an entire animal. It has a long history and is exceptionally well established. Advances in microsystems have allowed extreme miniaturization of many of these optical methods and thereby permit imaging (particularly *in vivo*) in locations not previously accessible. Imaging is in direct contrast with sensing (discussed in Section 3.09.5), which is typically a point measurement; an image is usually 2D or 3D.

Microimaging is used in the development of tools for minimally invasive surgery and diagnostics (Haga and Esashi 2004), and it has benefited greatly from advances in microcameras (Chapter 3.08). Closely related, but outside the focus of this chapter, is the development of biomimetic systems, which use microsystems technologies to fabricate structures inspired by nature; an advanced example is an artificial compound eye (Jeong *et al.* 2006).

3.09.4.1 MEMS Components for Imaging

In many cases the incorporation of microsystems technologies into biophotonic imaging systems has occurred stepwise; individual miniaturized components frequently become essential parts of imaging concepts that still use classical (macroscopic) technologies. The most useful MEMS components in this regard have been microlenses, micromirrors, and optofluidics.

3.09.4.1.1 PDMS lenses

Making smaller microscopes allows their implementation in areas for which a classical apparatus is not suitable; using microlenses, particularly tunable microlenses, allows fabrication of novel miniaturized microscopes. Using poly(dimethyl siloxane) (PDMS), a microstructurable elastomer, a rubber microscope has been fabricated (Gambin *et al.* 2006). A solid immersion lens, with which numerical aperture (NA) > 1 is possible, was manufactured from PDMS and combined with spherical lenses, a light emitting diode (LED) light source, and a charge-coupled device (CCD) imager yielding a system that is only 10 cm long. This novel application of low-cost MEMS technology yielded a microscope with which fluorescent cellular imaging could be performed. Further integrating a microfluidic channel into the device has shown that a handheld flow cytometer is feasible.

Elastomer materials such as PDMS can also be structured as membranes, and they open opportunities for a new class of tunable micro-optical components suitable, for example, for incorporation into new generations of endoscopes (Werber and Zappe 2005). A 50-μm-thick membrane is suspended over a silicon micromachined fluidic cavity; filled with an optically transparent liquid with a suitably high refractive index, the distensible membrane forms a convex lens and controlled changes in pressure allow tuning of the radius and thus the focal length. An example of two distended lenses on a silicon substrate is shown in **Figure 7**.

Lenses fabricated using this technology, with typical diameters of 400 μm, are tunable in the range $0.4\,\text{mm} \leq f \leq \infty$. Concave lens structures can be made by applying negative pressure relative to the ambient pressure. Pressure-actuated tunable microlenses such as this are of particular value in endoscopic zoom optics, since classical optical

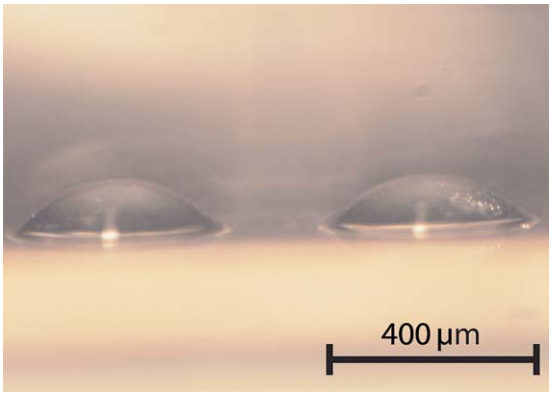


Figure 7 A pair of PDMS membrane lenses extending from a silicon surface; pressure variations in the liquid filled cavity below allow dynamic tuning of the focal length.

systems at the required size scales are impractical and pressure actuation implies that no high voltages are required at the distal end of the endoscope. Tunable microlenses are also of interest in OCT systems, allowing a more effective depth scan.

3.09.4.1.2 Micromirrors

Micromachined mirrors, such as the one shown in Figure 8, not only are well established in communications technology but are also finding an increasing number of application domains in biophotonics. As discussed in Chapter 3.01, OCT systems rely on 1D and 2D scanning mirrors to enable complete scanning of tissue sections. Micromirrors have also expanded the utility of endoscopes, allowing a larger field of view for a compact optical system. Advanced applications (Fu *et al.* 2006) include an invasive two-photon fluorescence microscope, which employs a compact imager based on a silicon micromirror capable of 6° angular motion with only 5 V required for electrostatic actuation. A gradient index (GRIN) lens was used for imaging and a photonic crystal fiber was used to separately guide excitation and fluorescence light to and from the imaging area.

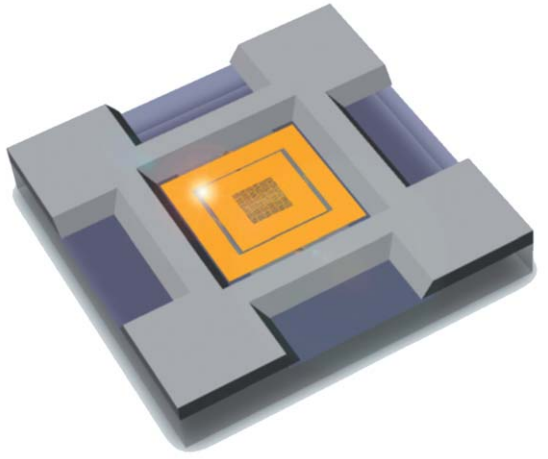


Figure 8 Sketch of a dual-axis silicon-based micromirror developed for endoscopic scanning; the size of the central mirror region is 500 µm × 500 µm. The mirror rotates about two perpendicular axes parallel to the mirror edges.

An alternative approach to two-photon microscopy features two-axis scanning using Si-based micromirrors with an area of 750 µm × 750 µm and 16° angular motion (Piyawattanametha *et al.* 2006). With an acquisition rate of 3.5 kHz, this and the previous example represent technologies useful for both endoscopy and portable imaging systems.

3.09.4.1.3 Optofluidic imaging

Ultracompact imaging can also be accomplished by using a completely different approach without a macroscopic analogue. Microfluidics can do more than transporting liquids or particles suspended therein; the field of optofluidics uses microscopic fluidic structures or flows to perform optical functions. Tunable liquid lenses (Krogmann *et al.* 2006) represent one such application; another application is the use of microfluidics for high-resolution imaging directly (Psaltis *et al.* 2006). Using a fluid system to transport a particle of interest, for example, a cell, over an arrangement of micromachined holes through which the sample is imaged, flood illumination leads to a sequence of partial images of the object, which can be reassembled after it flows through the system.

The micrometer-sized holes are arranged such that the entire sample is scanned as it moves past, and illumination in the visible wavelength range leads to image resolution better than 100 nm. The structure is very simple and compact, requires no classical optical components such as objectives or other lenses, and can conceivably be manufactured using replication such that low-cost, disposable microimagers for cellular analysis are feasible.

3.09.4.2 Adaptive Optics

Adaptive optical systems are well established in astronomical applications, in which they are used to compensate for atmospheric distortion of the incoming optical wavefronts from stars. The concept, illustrated schematically in Figure 9, relies on the measurement of wavefront distortion using a wavefront sensor,

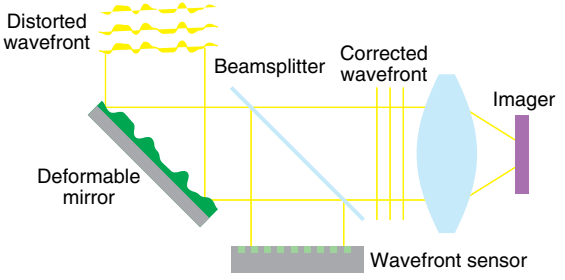


Figure 9 The principle behind adaptive optics: a distorted wavefront is incident on a deformable mirror, which compensates the wavefront distortion. Part of the signal is directed to a wavefront sensor, the output from which defines the mirror shape dynamically. The image is then of the corrected wavefront.

frequently a Shack–Hartmann sensor. Output from the sensor provides an input to a dynamically deformable mirror, whose reflective surface is modulated to compensate for the distortion. When placed in a feedback loop with the appropriate control electronics, the result is a corrected wavefront and thus a higher resolution image (Tyson 1997).

Macroscopic adaptive optical mirrors, typically with apertures ranging from 50 to 150 mm, driven by piezoactuators and with up to 4000 segments, are industrial products, but of a size and cost that makes them unsuitable for biophotonic applications. In contrast, MEMS-fabricated adaptive optical mirrors have recently become available (Doble *et al.* 2002), using electrostatically deformed silicon membranes (Vdovin *et al.* 1999) or highly resilient photostructurable polymer membranes (Friese *et al.* 2004) as mirror surfaces. These microfabricated mirrors, an example of which is shown in **Figure 10**, are not only considerably smaller than their macroscopic counterparts, but also require lower operating

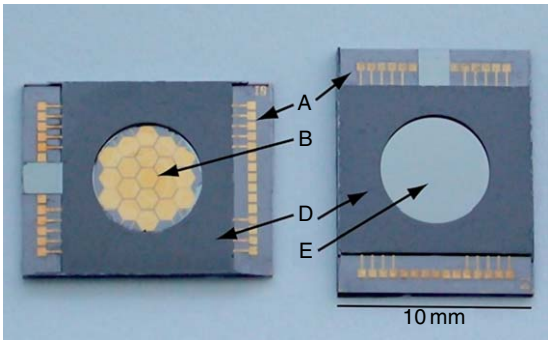


Figure 10 A polymer-based adaptive optical mirror with a diameter of 5 mm and 39 actuation electrodes. A, Power input; B, electrodes; D, silicon frame; E, reflecting membrane.

powers, switch faster, and are usually significantly less expensive.

As a result, microfabricated adaptive optical mirrors are increasingly being used for medical imaging, particularly in ophthalmology (Liang *et al.* 1997). Adaptive optical wavefront correction has recently been applied to laser scanning ophthalmology, confocal microscopes, OCT systems, and two-photon fluorescence microscopy. In ophthalmological examinations, in particular of the retina, the inhomogeneous optical structure of the lens, tear fluid, the vitreous humor, and the cornea make high-resolution imaging of the eye difficult. Adaptive optics has improved the imaging capabilities to the point where individual rods and cones can be discerned on the retina of a live eye (Roorda and Williams 1999).

One example of a typical MEMS-based adaptive optical system for ophthalmology has a 4.4 mm × 4.4 mm aperture and consists of 144 individually movable mirror segments (Zhang *et al.* 2006b). With 3.5 µm of stroke for each mirror, the wavefront can be corrected to 0.1 µm, which, in a confocal arrangement, led to an axial resolution better than 100 µm with laser diode illumination at 655 or 840 nm. When combined with a method for retinal tracking, which compensates for the movement of the eye using a low-power, long-wavelength tracking laser, an image resolution of 10–15 µm has been obtained (Hammer *et al.* 2006).

MEMS-based adaptive optical mirrors have also been used to improve the resolution of standard microscopy techniques (Potsaid *et al.* 2005); when added to a scanning microscope, adaptive optics resulted in higher resolution with an increased field of view when compared to a standard configuration.

3.09.4.3 Confocal Microscopy

As briefly described in Section 3.09.3.1, a confocal microscope uses a pinhole aperture to limit the depth of the field of light reflected and scattered from the tissue and thus improves signal contrast. Most confocal microscopes use a laser beam for illumination and this is scanned over the surface to be imaged.

A microsystems-based implementation of a confocal microscope consisting of a micromachined scanning head fabricated from silicon and fused silica has been demonstrated in Dickensheets and Kino (1998). Input and output of light is through optical fibers, which are passively aligned to a scanning mirror structure using physical alignment features. The light is directed through a transmission phase

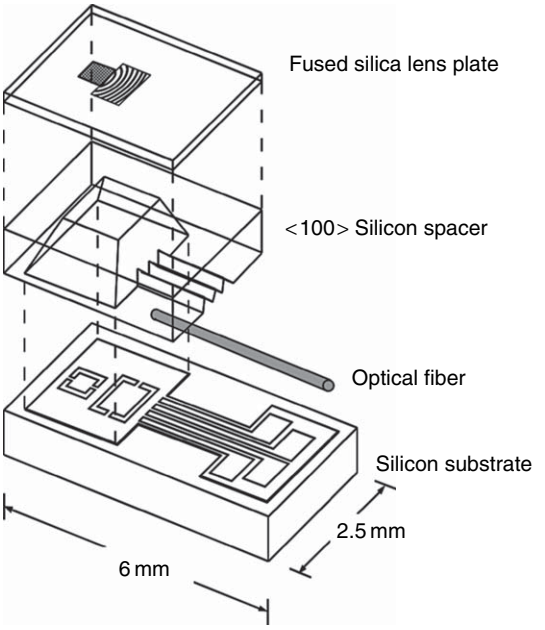


Figure 11 Diagram showing the structure of the silicon-based microfabricated confocal microscope. (Source: Dickensheets D L, Kino G S 1998 Silicon-micromachined scanning confocal optical microscope. *IEEE/ASME J. Microelectromech. Syst.* **7(1)**, 38–47; Diagram courtesy: David Dickensheets, Montana State University.)

grating etched into silica, which focuses the beam off-axis onto the sample, by using a reflecting silicon-based cavity. The reflected light is directed through the same system back to an external photodetector through an external fiber coupler.

As shown in **Figure 11**, the entire confocal sensor head has external dimensions of 6.5 mm × 2.5 mm × 1.2 mm; all light sources, detectors, couplers, and electronics are external such that the fiber-connected measurement unit can be used endoscopically. Using illumination at 633 nm, submicrometer resolution was achieved at a 20-Hz frame rate over a 90 µm × 90 µm field of view. Similar approaches, using a forward-looking optical configuration with a diameter of 3.4 mm and employing a scanning MEMS mirror and standard off-the-shelf lenses, have also been demonstrated in Furihata (2001).

3.09.4.4 Optical Coherence Tomography

Optical coherence tomography is one of the most advanced and useful forms of structural bioimaging, and microsystems-based miniaturized OCT imagers are the subject of Chapter 3.01. OCT allows the generation of 3D images up to several millimeters below

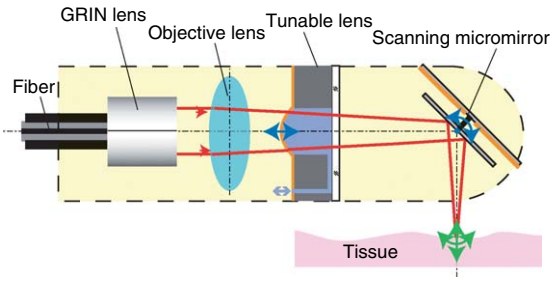


Figure 12 Schematic of an endoscopic OCT system; a tunable lens is used to dynamically change the focus with varying spacing of the tissue sample and the two-axis micromirror is employed to scan across the sample surface.

the surface of the tissue and has therefore become an essential tool for nondestructive diagnostics. Endoscopic OCT systems, which benefit from MEMS technologies such as scanning mirrors (Xie *et al.* 2003), ultra-miniaturized motor-rotated scanners (Tran *et al.* 2004), or tunable lenses (Divetia *et al.* 2005), now allow this type of imaging to be accomplished *in vivo* with a minimum patient discomfort; a schematic view of such a system is shown in **Figure 12**.

3.09.4.5 Imaging Microrobots

An increasingly important task in biophotonic imaging is to provide visual feedback prior to and during surgical procedures, and to combine diagnostics with high-resolution real-time imaging; microsystems technology is providing the means to accomplish extensive intracorporal diagnostics with highly functional systems with increasing autonomy (Dario *et al.* 2000a, b, Haga and Esashi 2004).

Imaging microrobots represent a semiautonomous system; it is used inside the body but attached to the outside world through wires or fibers. For colonoscopy, for example, a pneumatically actuated microrobot has been developed for imaging inside the colon with minimum patient discomfort (Dario *et al.* 1999). Illumination is through a fiber optic bundle and imaging is accomplished using a CCD-based microcamera of the type described in Chapter 3.08. Whereas this system is still controlled externally, current work is focusing on the use of MEMS technologies for autonomous actuation (Menciassi *et al.* 2005), power scavenging, and improved diagnostics tools, which permits the development of more highly functional microrobots of this type. All aspects of microsystems engineering are likely to play a role in these systems.

3.09.4.6 Capsule Endoscopy

It has long been a dream of many surgeons, and not a few film directors, to have a truly autonomous, ultraminiaturized system for diagnostics and therapeutics inside the body; of primary interest is the gastrointestinal tract. Moving from fiction to science, microsystems technologies have provided the methods to realize wireless systems, which transmit diagnostics data as they move through a patient. The idea is not new; first trials were performed in the 1950s (Zworkin 1957) and electronic capsules, which move through the gastrointestinal tract and monitor parameters such as temperature, pH, and oxygen concentration, have been clinically tested (Johannessen *et al.* 2004).

Optical imaging in the gastrointestinal tract using a wireless video capsule has been demonstrated in Iddan *et al.* (2000) and Meron (2000); the 11 mm × 26-mm capsule has no external wires or fibers, is swallowed by the patient, and moves by peristalsis through the bowel. The short focal length of the optics implies that no inflation of the intestine is necessary. The capsule is fully self-contained, and includes a complementary metal oxide semiconductor (CMOS) imager, application-specific integrated circuits (ASICs) for data processing, and white LEDs for illumination; data is transferred to an external receiver via UHF telemetry (Glukhovsky 2004). The capsule can operate for up to 6 h, and it images the regions of the tract that it faces; external manipulation of position is not yet possible. FDA approval for the video capsule, marketed by the Israeli company Given Imaging, was granted in 2001 and numerous other laboratories, including Olympus and Norika in Japan, are developing prototypes of a similar nature, incorporating increasing amounts of microsystems technology into the imager.

3.09.5 Optical Biosensors

Diagnostics of biological samples are most often performed with biosensors, whereby optical sensing techniques offer a wide range of measurement capabilities unmatched by electrical or other means. Hence, biosensors have been developed to measure a wide variety of analytes, often in array form to permit highly parallel or multimodal analysis (Ligler and Rowe-Taitt 2002, Vo-Dinh *et al.* 2001). Much of the work in this area has benefited from and promoted developments in micro total analysis (μTAS) systems (Janasek *et al.* 2006), cell chips (El-Ali *et al.* 2006), and the long history of fiber optical sensors (Mehrvar *et al.* 2000) as well as recent developments in semiconductor processing technology (Zappe 1999).

3.09.5.1 Integrated Optical Approaches

As mentioned in Section 3.09.3.4, a large fraction of optical biosensors use an integrated optical approach, confining the optical field to waveguides (see, e.g., Zappe 1995 and Chapter 3.05). Most implementations use a section of waveguide with the core exposed (analogous to a section of stripped optical fiber) for which a change in the external refractive index, Δn_{ext}, or the adhesion to or change in thickness of a surface layer, Δt, gives rise to a change in the effective index of the waveguide, ΔN (the effective waveguide index, N, should not be confused with the refractive index of a material, n). The task of most optical biosensor systems is to sensitively measure this change.

3.09.5.1.1 Mach–Zehnder interferometer

One of the most sensitive methods to measure a change in waveguide effective index is by means of an interferometer. ΔN gives rise to a phase shift $\Delta\phi$ of the field in the waveguide and this latter parameter is best determined interferometrically. Integrated optics technology allows strong miniaturization of interferometers, so that compact, highly functional sensor microsystems can be produced using this approach.

The Mach–Zehnder interferometer, illustrated in **Figure 13**, translates the phase shift in one arm into an intensity change. The sensor region is a section of waveguide of length L with the core exposed; a change in the effective index of the waveguide ΔN in this region results in $\Delta\phi$ with respect to the reference arm (which has a phase shifter to allow adjustment of the optimal operating point), given by

$$\Delta\phi = \frac{2\pi}{\lambda} L \Delta N \qquad [8]$$

for the operating wavelength λ. The Mach–Zehnder interferometer translates this phase shift into an intensity change in the optical output intensity, I, by the well-known interferometer relationship

$$I = 2I_0 + 2\sqrt{I_0}\cos\Delta\phi \qquad [9]$$

where I_0 represents the intensity of the optical input.

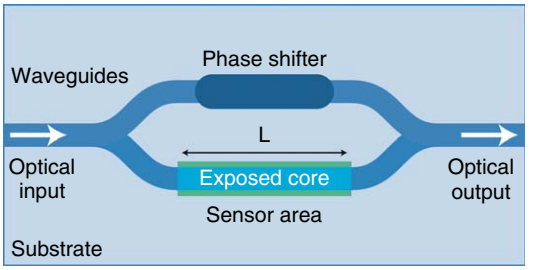


Figure 13 A waveguide-based Mach–Zehnder interferometer. In the sensor region, the waveguide core is exposed and variations in the external environment give rise to a phase shift with respect to the other arm, in which a phase shifter fixes the operating point.

A number of factors must be taken into account to optimize the performance of this optical sensor. A long sensor region of length L improves system sensitivity, usually given by a factor such as resolvable Δn per sensor length. The waveguide must be designed to have a large evanescent field strength outside the waveguide core such that the effective index N changes strongly with variations in external parameters; the resulting waveguide sensitivity is often given as

$$S = \frac{\partial N}{\partial n_{\text{ext}}} \quad \text{or} \quad S = \frac{\partial N}{\partial t} \qquad [10]$$

The phase shifter in the reference arm should be used to tune the operating point to the condition of highest sensitivity, which occurs when the reference and sensor arms have a phase offset of $\pi/4$. For most waveguide materials, changes in temperature will give rise to a change in N; the advantage of a compact integrated optical approach is that the temperature is generally uniform over the substrate and the two arms are typically separated by a few tens or hundreds of micrometers; thus the system is typically self-temperature compensating, an important advantage of microsystems approach.

Numerous implementations of this type of integrated optical interferometer have been fabricated. Using III–V technology (Chapter 1.02), a monolithically integrated system including a laser, a electro-optical phase shifter, and a photodetector has been fabricated (Maisenhölder *et al.* 1996, 1997a, b). The 5-mm-long chip had a sensor length of $L = 2$ mm and was combined with an external fluidic system to sensitively measure both external refractive index changes and selective adhesion of proteins onto the waveguide surface. A sensitivity of $\Delta n/L = 10^{-5}\text{mm}^{-1}$ was determined for the former and a concentration of $3\,\text{pg}\,\text{mm}^{-2}$ of surface analyte could be resolved.

Using silicon technology with an external light source, integrated Mach–Zehnder interferometer sensors with integrated microfluidics and CMOS-based evaluation electronics have been fabricated (Luff *et al.* 1998) and variations that use three mutually phase-shifted output waveguides (Sepúlveda *et al.* 2006) for enhanced phase shift resolution have also been shown. Significant recent progress in polymer-based waveguides suggests that hybrid microsystems that use separate substrates for light emission and detection, fluid handling, and sensing are feasible.

3.09.5.1.2 Young interferometer

The Mach–Zehnder interferometer translates a sensor phase shift into a single intensity change at the optical output. An alternative is to use a Young interferometer, as shown in **Figure 14**, which expands the output to a 1D line of interference maxima and minima, which can be imaged by, for example, an external CCD. As the sensor-based phase in one arm of the integrated optical Y-coupler shifts, the entire interference pattern shifts allowing a higher resolution of induced phase change. In addition, the ambiguity that results from the periodic nature of interference may be resolved such that a wider range of $\Delta\phi$ values can be measured.

Using dielectric-based (SiO_2 or Si_3N_4) waveguides on a silicon substrate, the Young interferometer has been demonstrated to be suitable for biosensing applications (Brandenburg *et al.* 2000). With a high sensitivity for external refractive index change ($\Delta n = 10^{-7}$) and for adhesion of surface analytes ($750\,\text{fg}\,\text{mm}^{-2}$) for a 1-mm-long sensor region, performance advantages over the Mach–Zehnder interferometer were noted. Miniaturization of this approach, perhaps using 1D photodetector arrays, is conceivable if not yet demonstrated.

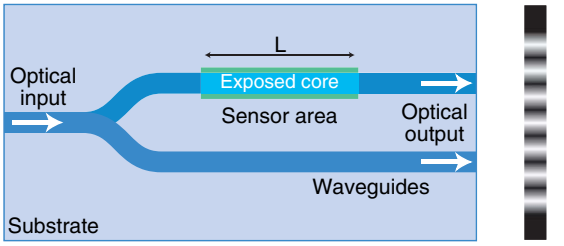


Figure 14 A waveguide-based Young interferometer in which the sensor region is found in one arm of a Y-coupler. The interference resulting from a phase shift in one arm results in a 1D interference pattern, which is imaged off-chip.

3.09.5.1.3 Directional couplers

Waveguide-based biosensor concepts need not be exclusively interferometric. An alternative concept, shown in **Figure 15**, uses an array of couplers to translate refractive index change into an intensity shift (Luff *et al.* 1996). In this structure, the optical input from the left is split into two waveguides using a Y-coupler, the lower of the two acting as a reference signal, unaffected by external index changes. The upper arm of the Y-coupler is fed into a directional coupler, formed by two waveguides in close proximity, in which part of the light is transferred to the uppermost waveguide (see e.g., Zappe 1995, Chap. 11.2.2).

The amount of light transferred into the output waveguide is a function of the spacing between the two waveguides, s, and the length of the overlap region, L, (both fixed) as well as the effective waveguide index; the latter parameter can be altered using an exposed waveguide core in the sensor region. Given an input power into a directional coupler I_{in} (in our example, the power delivered by the upper arm of the Y-coupler), the output power is given by the following equation:

$$I_{out} = I_{in} \frac{\sqrt{\sin^2(\kappa L)^2 + \left(\frac{\Delta\beta L}{2}\right)^2}}{1 + \left(\frac{\Delta\beta}{2\kappa}\right)^2} \quad [11]$$

in which

$$\Delta\beta = \frac{2\pi}{\lambda}\Delta N \quad [12]$$

where ΔN is the waveguide effective index and κ is the coupling coefficient between the two waveguides, determined from the waveguide dimensions and spacing. Thus, as in the interferometer structures, if external refractive index or surface layer changes occur, the resultant shift in ΔN leads to a change in the output intensity. The reference arm remains unchanged.

This type of coupler-based biosensor has been made using Ag^+ ion exchange waveguides (Luff *et al.* 1996). At an excitation wavelength of λ=633 nm, a waveguide spacing in the coupler $s = 7.2\,\mu m$ and an overlap length $L = 1$ mm, a refractive index resolution of $\Delta n = 5 \times 10^{-10}$ was achieved. Since an intensity comparison is required for a precise measurement, it is unlikely that this approach can reach the high resolutions achievable by interferometric techniques.

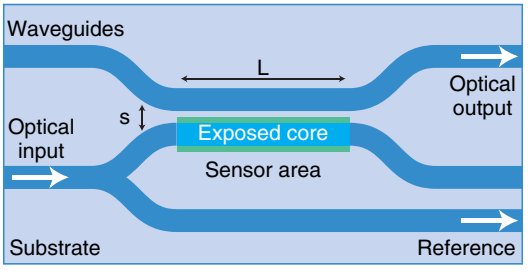


Figure 15 A combination of directional and Y-couplers used for biosensing. A change in the effective waveguide index in the sensor region leads to a change in coupling efficiency, and thus a measurable intensity shift at the optical output.

3.09.5.1.4 Waveguides as sensors

A surface-modified polymer waveguide biosensor using the waveguide itself as a sensor has been fabricated (Chang-Yen and Gale 2005, 2006). The waveguides were made from SU-8 photoresist and NOA61 optical adhesive. Both waveguides were surface modified with an oxygen-sensitive dye and the enzyme glucose oxidase, and were surrounded by a PDMS fluidic channel designed for restricting the diffusion process of the analyte into the enzyme in a single direction. The waveguide was coupled by optical fibers to an excitation LED operating at a wavelength of 450 nm and to a spectrometer for fluorescence detection. Except for the glass substrate, the biosensor itself consisted completely of soft and organic materials.

3.09.5.2 Surface Plasmon Resonance

Bulk optical systems do not use fibers or waveguides to guide light signals; instead, light is transmitted through free space and classical optical components (lenses, prisms, beam splitters, etc.) form the optical field.

Using this type of optical configuration, the SPR technique has become an established approach for high-resolution optical biosensing (Homola 2006). SPR relies on the excitation of surface plasmons (charge density oscillations) in thin metal films; this excitation occurs when the light field impinges on the surface at a precisely defined angle, which is in turn a function of the surface conditions. Thus adhesion of a biologically relevant substance onto the metal surface causes a shift in the excitation condition, which is then translated into a sensor signal.

A typical SPR measurement uses the Kretschmann configuration shown in **Figure 16**. Light is directed through a glass prism onto the thin metal layer, typically 40–50 nm of Au or Ag. At a particular incident

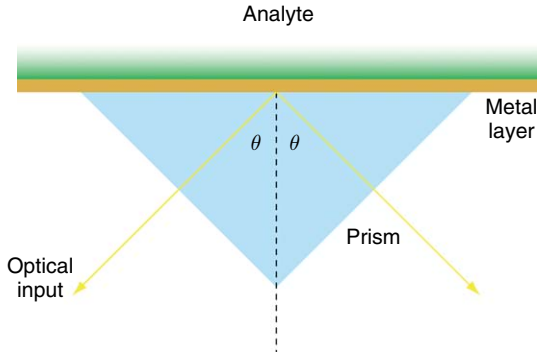


Figure 16 A surface plasmon resonance (SPR) measurement using the Kretschmann configuration. Surface plasmons are excited in the thin metal layer by an optical field incident at a precisely defined angle through the prism; this plasmon resonance leads to a dip in reflected intensity as a function of angle and is a sensitive function of surface conditions on the metal.

angle, known as the ATR angle, when the wavevectors of the excitation field and the plasmons overlap, resonance is achieved and the resultant transfer of energy to the plasmons results in a decrease in the reflection intensity. This angle is a sensitive function of the surface conditions on the metal such that external refractive index changes of a measurand or selective adhesion of an analyte results in a change in resonance angle. The SPR measurement works in TIR such that only an evanescent field exists in the metal and the analyte, similar to the waveguide-based approaches discussed in Section 3.09.5.1; the ATR angle is always larger than the critical angle for TIR (Section 3.09.2.4).

The SPR measurement is a spot measurement, which can be done in real time. Due to its high sensitivity and relatively simple optical setup, the approach has become a standard for high-resolution laboratory biosensing. Using variations in which optical heterodyning of two reflections was used (Kuo *et al.* 2003) or the excitation of two plasmon types was employed (Slavík *et al.* 2006), sensitivities to surface concentrations in the range of 5–250 ng l^{-1}, equivalent to 200 pM of analyte, were observed. A differential phase measurement (possible because SPR employs only the *p*-polarization) accomplished using an external Mach–Zehnder interferometer led to refractive index resolutions on the order of $\Delta n = 5 \times 10^{-10}$ (Wu *et al.* 2004).

Despite the high resolution and flexibility of the SPR approach, few attempts have been made to miniaturize it using microsystems technologies. MEMS-based micromirrors have been employed in highly parallel SPR systems, in which a single light source can be scanned over an array of sensor regions; microfluidics have been integrated onto the analyte side, allowing controlled flow in a miniaturized system. The need for coupling vertically onto a metal film with high angular accuracy and the need to vary this angle imply that a 3D structure is required so that complete microsystem integration requires further advances in micromachining capabilities.

3.09.5.3 Fabry–Perot Interferometer

Interferometric biophotonic measurements can be made using bulk free space optics in combination with MEMS-based structures, enabling increased integration. A Fabry–Perot interferometer, well-established in spectroscopy and the basis for the resonant cavity of most types of lasers, can be employed to sensitively measure changes in refractive index. For the structure shown in **Figure 17**, the ratio of output to input optical intensity is given by the well-known relation

$$\frac{I_{\text{out}}}{I_{\text{in}}} = \frac{1}{1 + \left(\frac{2F}{\pi}\right)^2 \sin^2\left(\frac{\phi}{2}\right)} \qquad [13]$$

for the case in which absorption in the mirrors is ignored. In eqn [13], for mirror reflectance R, the finesse F is given by

$$F = \frac{\pi\sqrt{R}}{1-R} \qquad [14]$$

the phase ϕ can be written as

$$\phi = \frac{4\pi nt}{\lambda_0} \qquad [15]$$

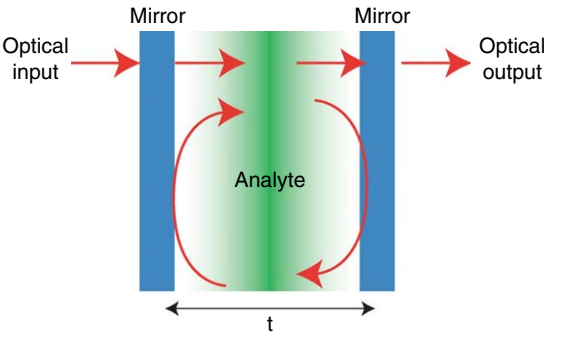


Figure 17 A Fabry–Perot resonant cavity consisting of two semitransparent mirrors with a spacing *t*. Changes in spacing or in analyte refractive index lead to a shift in the resonance wavelength.

where n is the refractive index in the cavity, λ_0 the free space (vacuum) wavelength, and t the cavity thickness.

From the above relationships, we see that the optical thickness (nt) can easily be varied to shift the resonance and thus the output intensity of the Fabry–Perot cavity; by allowing a change of n due to the presence of or variation in an analyte in the cavity, or possibly by changing the thickness t due to an external measurand, the output intensity I_{out} varies, thus providing a sensor signal.

The Fabry–Perot cavity can be fabricated using silicon micromachining technologies (Han 1999); the mirrors are formed using dielectric layers so that compact, intrinsically aligned high-finesse cavities can be manufactured in 2D arrays. The internal surfaces of the mirrors have been coated with chemically sensitive polymers, whose refractive index varies selectively as a function of the concentration of certain absorbed species. Interrogation can be done by using bulk optical, fiber-based or, in the future, micro-optical means allowing parallelization and miniaturization.

A deformable Fabry–Perot cavity has also been employed for blood pressure sensing (Haga and Esashi 2004, Katsumata *et al.* 2000, Tohyama *et al.* 1998). Using a silicon-based micromachined structure and a metal-coated SiO_2 membrane as a movable mirror, the resonant cavity with a diameter of 125 µm could be inserted into a blood vessel. Ambient pressure variations were sensed using light from an optical fiber, whereby pressure-induced movement of the diaphragm modulated the light reflected from the Fabry–Perot cavity; performance is as good or better than the established technologies (Reesink *et al.* 2002). Similar concepts using bendable micromachined cantilevers and white light interferometry or spectral modulation techniques for interrogation (Wolthuis *et al.* 1991) have also been employed for *in vivo* blood pressure sensing.

3.09.5.4 Promising Technologies

Numerous biosensing approaches have been developed, which show considerable promise for miniaturization and integration using the technologies of microsystems engineering; a few of these are discussed here. Of considerable interest for many is the availability of ultracompact, high-performance light sources, usually semiconductor-based diode lasers and LEDs (Zappe 2004), which have begun to replace many macroscopic laser systems. Optoelectronics technologies play a significant role in this area.

3.09.5.4.1 Laser diode technologies

For example, the integration of vertical cavity lasers (VCLs) with glass- or plastic-based microfluidics is reasonably well established; the 2D array capabilities coupled with the low power consumption of VCLs have made these attractive light sources for optical excitation and interrogation. Integration has been advanced further by including PIN photodiodes, microlenses, and filters, ultimately combining glass and GaAs substrates (Thrush *et al.* 2005); by employing an excitation wavelength of λ=773 nm, a detection limit of 40 nM (using dyes as a sample system) was reached using this complete optical biosensor microsystem. Silicon-based integrated microsystems, combining avalanche diodes as light emitters, PN photodiodes, and fiber mounting guides coupled with a microfluidic module, have been shown to have detection limits down to 20 fM using protein-based test systems (Misiakos *et al.* 2004); the concept uses ATR of waveguide modes in a silicon nitride-based fiber/waveguide.

A hybrid micro-optical sensor using the Doppler effect to continuously and noninvasively measure blood flow velocity *in vivo* has been developed using a distributed feedback (DFB) laser diode mounted along with a photodetector on a silicon-based substrate; as seen in **Figure 18**, the substrate incorporates a waveguide circuit and provides mechanical, thermal, and electric contact (Higurashi *et al.* 2002). The convex waveguide edge allowed horizontal focusing without the need for an external

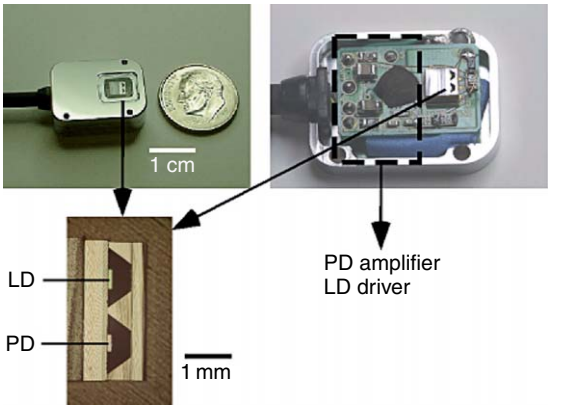


Figure 18 A micro-optical blood-flow sensor using the Doppler effect to sense flow velocity noninvasively. (Photo courtesy: Eiji Higurashi, University of Tokyo.)

lens, and the completely assembled optical sensor, with dimensions of 1.5 mm × 2 mm × 3 mm, was packaged together with a laser diode driver and photodiode amplifier. The performance of the sensor was sufficient to measure heartbeat and to determine the presence of occlusions, which lead to reduced blood flow; measurements could be accomplished by manually holding the sensor onto the skin.

3.09.5.4.2 Micro-optical technologies

The availability and manufacturability of microlenses (Chapter 3.02) together with microfluidic systems have led to numerous compact lab-on-a-chip concepts (Chapter 2.13) in which the optics is tightly integrated with the sensor system. In one example, PDMS-based microfluidics is combined with planar lens structures on a single, disposable substrate (Seo and Lee 2004); the optical excitation, an LED emitting at 466 nm, and the detection, a CCD camera or spectrometer, are part of a separate entity, which is passively aligned to the fluidic/optic chip. As a result, medically relevant fluorescence measurements can be made in which the orthogonal light paths of excitation and detection result in low cross talk and high signal-to-noise ratio. The concept allows fabrication of compound planar lenses, which improve measured fluorescence intensity markedly, in part due to reduced spherical aberration.

Microlenses have also been combined with microfluidic channels on a single chemical chip (Roulet *et al.* 2001); using cylindrical microlenses, separate light paths for excitation and fluorescence were generated, allowing, again using a dye test system, a resolution of 20 nM. The concept employed can be stacked to yield a highly parallel implementation capable of detecting a spectrum of compounds. Resonant microcavities (e.g., silica glass beads with a diameter of 300 μm) have been shown to have a resonance wavelength, which is a sensitive function of protein adhesion on the surface (Vollmer *et al.* 2002); these cavities have very high Q ($>10^6$), and using fiber input/output at $\lambda = 1340$ nm, molecular weights down to 50 g mol^{-1} can be detected, a limit below that of common SPR sensors.

Micro-optical sensing and interrogation concepts have been developed for continuous blood glucose monitoring, one of the most important routine medical diagnostic procedures. Whereas numerous transdermal measurement schemes, which use transmission of IR light through a section of skin, have been demonstrated, accuracy is insufficient to allow widespread application. An alternative is the measurement of glucose content in tear fluid, which has a concentration of about 10% of that in blood with a time lag of about 30 min. Using a soft contact lens doped with a boronic acid-based fluorescent probe, which binds glucose and as a result exhibits a shift in fluorescence intensity or wavelength, the user can monitor his or her blood glucose using a handheld micro-optical interrogation system (Badugu *et al.* 2004, Geddes and Lakowicz 2004). Excitation employs a white LED and the system including filters and detectors forms a hybridly assembled optical microsystem with which a sensitivity to glucose in the range of 50–500 μM is achieved. In the search for efficient, cheap, and adjustable light sources for fluorescence excitation in the blue and UV spectral ranges (where many proteins show intrinsic fluorescence such that fluorescence dyes are not required for their detection), microscopic discharge sources with a metal anode and a cathode cavity filled with an aqueous solution have been reported (Mitra *et al.* 2006).

Finally, a number of optical microdevices under development can provide technologies of importance for future biophotonic sensor systems. Of particular relevance may be atomic force microscope (AFM)-type cantilevers with optical readout, incorporating VCLs, photodiodes, electronics, and microfluidics in a compact system (Lechuga *et al.* 2006); the cantilevers themselves can be employed as waveguides and the efficiency of input or output coupling is a function of mass adhering to the surface. In addition, it is expected that further miniaturization, performance improvement, and integration of microspectrometers (Adams *et al.* 2003, Wolfenbuttel 2005) can provide a further micro-optical component integratable into compact biosensing systems.

3.09.5.4.3 Swelling polymers

Recently, swelling of polymer optical elements upon interaction of the polymer with organic compounds has been recognized as an effect, which allows the fabrication of tunable micro-optical components, and can be used as an optical sensing principle. Optical elements in which this effect was observed comprise Bragg mirrors and filters, and microlenses. It has been shown that the transmittance maximum of a Bragg mirror fabricated from cross-linked poly(methyl methacrylate)/poly(styrene) multilayers is shifted by 600 nm in a nitrogen atmosphere saturated with organic solvents (Mönch *et al.* 2006b). It was demonstrated using Bragg mirrors that polymer swelling with solvents is a robust and sensitive tuning principle for

such multilayer Bragg structures that can even be extended to 3D Bragg structures (Mönch *et al.* 2006a).

Microlenses have been fabricated from poly(*N*-isopropyl acrylamide) hydrogels and their tunability with pH and temperature of the surrounding aqueous solution has been demonstrated (Kim *et al.* 2004). Hydrogel particles have been functionalized with biotin molecules, and it was shown that microlenses fabricated from biotin-functionalized hydrogel respond specifically and reversibly to the target protein (Kim *et al.* 2005). This effect can be used for optical protein assays. Conversely, a tunable microlens system based on hydrogel response to various stimuli has been shown to be useful as a sensor for pH and temperature; sensitivity to light, electric field, or antigens was also shown (Dong *et al.* 2006).

3.09.6 Flow Cytometry

In this section, the focus is on microsystems that combine microfluidic elements such as microchannels with photonics technology and serve for detection and manipulation of biochemical or biological species. Since virtually all biological species are kept in liquid solution, MEMS technology-based optical manipulation strategies for liquids is also considered. Although flow cytometry in a narrow sense means measurement of cells in flowing liquids, here a wider view is adopted and also includes MEMS for optical detection of molecules in liquids in microfluidic channels.

3.09.6.1 Principles

Flow cytometry in its original sense means counting of cells or measuring their size while they flow past a detection point in a fluidic channel individually. For detection of the cells, a number of optical methods familiar from the field of biosensors can be applied, the most frequently used being fluorescence labeling and detection. In a wider sense, flow cytometry is a tool not only for measurement of cells in a liquid flow, but also for biochemical species such as molecular aggregates or colloidal particles. Colloidal particles tailored in their surface chemistry can be used for biochemical analysis at the surface of the particles. The measurement of cell size for subsequent mechanical, electrical, or optical sorting is indispensable for collecting cells with specific features.

3.09.6.2 Particle and Biochemical Detection

Flow cytometry microsystems require a microfluidic network and optical detection methods integrated on a chip. After briefly considering the materials required for the microfluidic part, different types of cytometry systems are discussed; these systems comprise fully integrated systems, systems fabricated in a hybrid approach, and microfluidic systems with external optical detection.

3.09.6.2.1 Materials

Many microfluidic channels and fluidic networks in optical microsystems have been fabricated from polymeric materials including not only those that can be processed by injection molding and hot embossing, including poly(methylmethacrylate), COC, and poly(carbonate), but also more specialized materials such as PDMS and SU-8 photoresist. Commercially available PDMS kits consist of liquid precursors (a base material and a curing agent), which are mixed and thermally cross-linked at moderately elevated temperature (≈60–100°C) to result in a nontoxic, optically transparent, and elastomeric material (Young's modulus ≈1 MPa).

PDMS is capable of being replicated using microstructures with nanometer precision and is frequently used as a mold material for different replication techniques. In an oxygen plasma, its surface is easily activated such that it adheres well to silicon dioxide surfaces. SU-8 photoresist is known for its exceptionally good adhesion to a variety of surfaces, its outstanding thermal and mechanical properties, and its capability of being processed into thick layers and structures (100–1000 μm) by conventional spin coating and photolithography. Both PDMS and SU-8 are compatible with silicon process technology.

3.09.6.2.2 The biodisk

One particular polymer-based low-cost microfluidic system for biological applications, such as colorimetric assays, is the so-called biodisk. The biodisk is a polymer (e.g., a cyclo-olefin copolymer) disc of the dimensions of a conventional compact disc with integrated microfluidic structures, including channels, reservoirs, and reaction chambers, depending on the particular application. Fluids in the microfluidic network are driven by capillary forces and, upon rotation of the disc, by the centrifugal and the Coriolis forces, leading to flow, mixing, and separation of the liquids, depending on the microfluidic structure and the

rotation speed and angular acceleration. After dispensing the liquids into the reservoirs of the biodisk, a rotation routine (including several steps of acceleration, shaking, and constant rotation) is run and the result is optically read out in a macroscopic device; the biodisk player may ultimately take the form shown in **Figure 19**.

Since it is fabricated from optical polymers, the biodisk allows an easy integration of optical structures by hot embossing or injection molding. The integration of two microprisms together with a detection channel into the biodisk, as shown in **Figure 20**, was recently demonstrated (Grumann *et al.* 2006, Steigert *et al.* 2005) using the example of a colorimetric glucose assay on whole blood. In these experiments, a light beam normally incident on the biodisk was totally reflected at the first 90° prism, propagated in the 10-mm detection path (within the plane of the biochip) containing the liquid under investigation, and was emitted out of the biodisk at the second prism. Compared to a beam directly transmitted through the biodisk, the interaction length was increased by a factor of 10 by this setup.

Figure 19 A biodisk installed in what a biodisk player might look like, in which the results of the microfluidic processes may be measured. The disk is of the size of a standard CD. (Photo courtesy: Jens Ducrée, University of Freiburg.)

3.09.6.2.3 *Flow cytometry systems*

Flow cytometry systems use integrated fluidics and optics to sort cells on a chip. As shown in **Figure 21**, the cells flow past an illumination and a detection system where the measured fluorescence is generally used as a criterion to sort a given cell into one of the several output fluidic channels.

A sandwiched multichip system consisting of microfluidic mixer chip on a silicon-based detector chip has been fabricated for quantitative measurement of the concentration of total protein in urine by spectrophotometry (Minas *et al.* 2006, Ribeiro *et al.* 2005). The microfluidic mixer consists of two fluid inlets, which are combined into a meandered mixing line, and a detection chamber followed by the outlet. The fluidic channels were 500 µm wide, and were defined by standard photolithography in a 600-µm-thick SU-8 layer on a glass substrate. A second glass substrate with drilled holes covered the fluidic channels. The bottom chip contained photodiodes and readout circuitry fabricated by a double-metal, single polysilicon, 1.6-µm *n*-well CMOS process. An optical Bragg filter consisting of nine layers (SiO_2 and TiO_2) is postprocessed onto the photodiode. The Bragg filter had a transmittance maximum at a wavelength

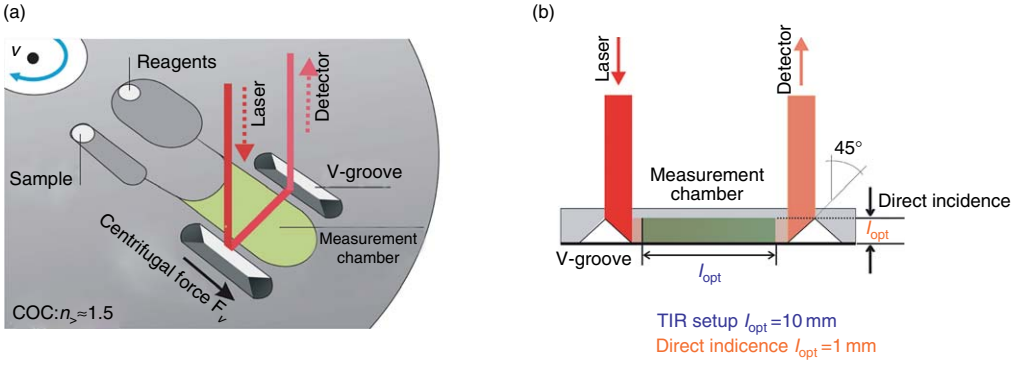


Figure 20 A biodisk with two integrated microprisms and an optical detection path length of 10 mm. (Source: Steigert J, Grumann M, Brenner T, Mittenbühler K, Nann T, Rühe J, Moser I, Haeberle S, Riegger L, Riegler J, Zengerle R, Ducrée J 2005 Integrated sample preparation, reaction, and detection on all high-frequency centrifugal microfluidic platform. *J. Assoc. Lab. Automat.* **10(5)**, 331–41.)

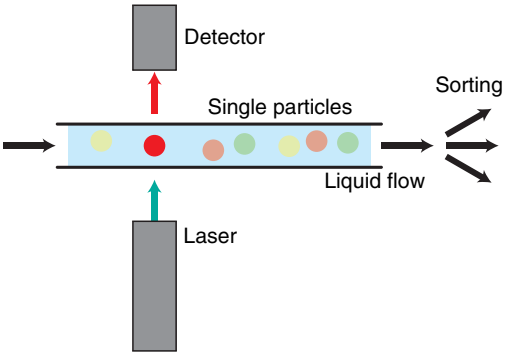


Figure 21 Flow cytometry: an illumination and a detection system use fluorescence of the cells, which flow past to sort these into one of several output fluidic channels.

of 600 nm and a passband FWHM of 10 nm. The fluidic chip is placed on top of the optoelectric chip, and a commercial passband optical filter is used for blocking IR and UV components of the excitation light spectrum. Excitation light was provided by an external tungsten halogen lamp and a monochromator.

It has also been shown that infrared particle image velocimetry may serve as a tool for measuring flow profiles in silicon-based MEMS without optical access through glass windows (Liu *et al.* 2005). This was demonstrated by Liu *et al.* (2005) by pumping a particle suspension thorough a fused silica capillary tube with a diameter of 255 µm; the optical absorption of a silicon microdevice was simulated in this work by placing a silicon wafer with a thickness of 500 µm in the optical path between the laser and the capillary tube. By that setup, the authors demonstrated that particle velocimetry in silicon-based microsystems is possible using infrared light.

A flow cytometry microsystem using mainly silicon technology was fabricated using a fluidic network in a deep-etched silicon wafer sealed at the top and the bottom by glass wafers (Baechi *et al.* 2002). The system was designed for manipulation of particles with a size ranging from a few hundred nanometers to a few microns (including cells and bacteria) and included photodiodes fabricated by ion implantation into the silicon and valves consisting of a thermopneumatically actuated silicone membrane with a thickness of 3 µm. The size of the valves was 300 µm × 300 µm and could be actuated at the rate of 10 Hz. In the experimental demonstrations, the system was illuminated with an external light source (a LED), and the integrated photodiodes were successfully used for particle detection.

3.09.6.3 Liquid Droplet Manipulation

Virtually all biological or biochemical species occur in a liquid solution. Consequently, manipulation and control of small volumes of liquids are major issues for all biochemical MEMS. Hence the electrowetting effect (Mugele and Baret 2005) has attracted renewed interest during the past decade. Apart from macroscopic optical traps and optical tweezers, several approaches for the manipulation of liquids on solid surfaces have been developed, including dielectrophoretic motion (Gunji and Washizu 2005, Jones *et al.* 2001) and digital microfluidic systems based on electrowetting with electrode arrays for creating, transporting, cutting, and merging of droplets (Cho *et al.* 2003). A review on integrated systems employing digital microfluidics is given by Fair and coworkers (Su *et al.* 2006).

Electrowetting has also been combined with photoconductive materials such as hydrogenated amorphous silicon to define active zones of the substrate (i.e., those zones of the substrate on which the electrowetting effect acts) by means of light (Chiou *et al.* 2003). By using such an additional photoconductive layer in the electrowetting substrate, the fixed arrangement of the electrodes limiting the applications of digital microfluidic devices is no longer needed. Instead, OEW allows the movement of liquids on a substrate in arbitrary paths by means of an SLM. The OEW effect was recently successfully applied for size sorting of microparticles (Chiou *et al.* 2005). Therefore, OEW has applications in light-directed cell sorting and highly parallel cell manipulation.

A photoconductive layer is also used in a recent experiment in which proteins in an electrolyte are moved by an electric field created between a photoconductive semiconductor electrode and a counter electrode (Hafeman *et al.* 2006). By applying a bias voltage between these electrodes, majority carriers in the semiconductor are depleted in the semiconductor–electrolyte interface such that no current flows across that interface. If the semiconductor is illuminated by a focused laser, carriers are generated and the depleted layer becomes conductive in the illuminated area. The resulting current leads to a high electric field over the illuminated area such that a trap is formed for the molecules. Molecules can therefore be moved by scanning the laser spot across the photoconductive layer. The authors call this method photoelectrophoretic localization and transport (PELT).

3.09.7 Microarrays

After the first experimental demonstration in the 1990s (Schena *et al.* 1995), microarray approaches to highly parallel biochemical analysis have become popular. Genomic and proteomic researches have profited enormously from the microarray approach, which allowed the detection of thousands of genes to the whole transcriptome of an organism. Consequently, microarray biochips are standard tools in genetic research today. The strong interest in fundamental genomics research and the need for fast and reliable biochemical analysis tools have triggered a considerable amount of research in microsystems technology in this field. In this section, an overview of optical microsystems devoted to particular steps in microarray biochip analysis is given and combined optical–fluidic microsystems for array-based biochemical analysis are presented.

3.09.7.1 Microarray Concepts

The first preparative steps in biochip analysis comprise amplification and fluorescence labeling of the analyte DNA. The biochip substrate consists of a glass or plastic (PMMA or COC) slide, which is coated with chemical coupling layers prior to spotting with probe DNA fragments. These coupling layers allow covalent linking of the probe DNA fragments to the surface of the biochip. Spotting is usually done using contact-free ink-jet or contact needle printing. A particular problem of microarray printing is that a standard format (pitch and spot size) of microarray biochips currently does not exist. While this fact impedes the conception of optical biochip excitation and readout laboratory equipment, it gives freedom to the microsystem engineer for designing integrated systems.

Current standard laboratory equipment for biochip fluorescence excitation and readout equipment are macroscopic systems, as shown in **Figure 22**. With regard to excitation, the simplest and most common method is flood illumination, but laser scanners (using Argon ion lasers) and several evanescent field approaches for excitation are also used (James *et al.* 2004, Lehr *et al.* 2003). For detection of the faint fluorescent light signals, most macroscopic biochip readers use photomultiplier tubes (PMTs). Photomultiplier tubes provide an excellent gain factor and linearity, but since typically only one PMT is used in a biochip reader, the chip must be read out sequentially. Consequently, only endpoint detection is possible, and the readout speed is low. Due to the availability of cooled (for increased signal-to-noise ratio) and/or electron multiplication CCD sensors, CCD cameras are becoming increasingly interesting for parallel and temporally resolved readout of whole arrays (Valentini and D'Andrea 2000).

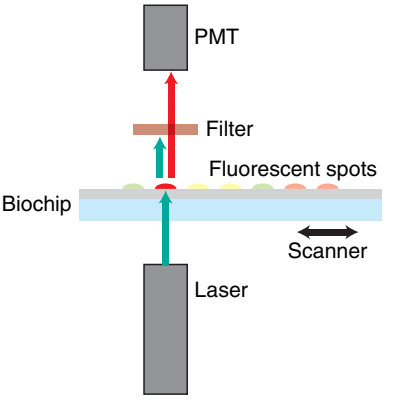


Figure 22 A typical setup of a biochip scanner. The fluorophores in a spot on the biochip are excited with a laser. The excitation light is filtered out in the optical path towards the detector (here a photomultiplier tube (PMT)). By mechanical scanning of the biochip, all spots are read out sequentially.

Current development in the field of biochip analysis instrumentation is driven by the need for fast and reliable point-of-care diagnostic tools for public and individual health care and for security. Though, until now, no fully integrated miniaturized DNA preparation device or biochip excitation/readout unit has been reported, many MEMS approaches have been realized for the preparation of the analyte or excitation or readout of biochips. Additionally, macroscopic biophotonics systems profit significantly from optical MEMS technology. In the following sections, an overview of biophotonic microsystems relying on an arrayed excitation or detection or for the generation of arrayed intensity patterns for use in macroscopic systems is given.

3.09.7.2 PCR Chips

Amplification of the analyte DNA fragment using polymerase chain reaction (PCR) is the first preparative step in genomic analysis. PCR is an enzyme-based method for synthesis of many copies of a given DNA molecule and requires careful temperature cycling. Macroscopic cyclers are standard equipment in biomolecular laboratories, but a

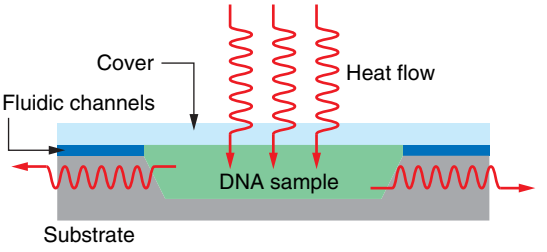


Figure 23 On-chip implementation of polymerase chain reaction (PCR), an enzyme-based method for synthesis of many copies of a given DNA molecule.

number of chip-based PCR approaches have been developed (Zhang *et al.* 2006a).

Several approaches for heating solutions in a micromachined chip, as shown schematically in **Figure 23**, have been developed; the promising approach is heating the analyte liquid in a volume of 5 nl by photon absorption (Slyadnev *et al.* 2001, Tanaka *et al.* 2000) using an external infrared laser diode (emission wavelength, 1472 nm; power, 150 mW). The liquid was in a simple, Y-shaped fluidic channel system in a chip and the system was operated under constant flow. Heating and cooling rates of 67 and 53°C s^{-1} were achieved; compared to conventional systems, this is faster by one order of magnitude.

Amplification of DNA by PCR in a silicon microreactor was also successfully demonstrated (Ke *et al.* 2004). Compared to a Peltier element-based temperature cycler, the amplification was achieved in a time reduced approximately by a factor of 2 (34 min instead of 62 min). A halogen lamp was used for noncontact heating of the liquid and a Pt100 resistor (a resistor with a very well defined temperature dependence of its resistance, hence a popular choice as a temperature sensor) for temperature monitoring and control.

Temperature control is an essential issue for successful PCR and many ultrasensitive biomolecular detection methods. It has been demonstrated that temperature can be measured in a cylindrical microchamber (8 µm diameter, 2.2 µm height) with an accuracy of 2°C (Arata *et al.* 2005). The proposed experimental method relies on the temperature dependence of the fluorescence intensity of a fluorescent dye, in this case Rhodamine B.

3.09.7.3 Integrated Photonic Detection

A further step toward system integration is a sandwiched silicon/pyrex PCR chip with pigtailed optical input and output fibers (Schabmueller *et al.* 2001). Two different systems have been developed: in the PCR chip, both fibers (one for fluorescence excitation, the other for detection of fluorescence light) were aligned perpendicularly. The other system employed only an input fiber, but instead of the output fiber an integrated photodiode was used at the bottom of the reaction chamber. The work demonstrates the feasibility of an integrated PCR system and fiber-based detection of fluorescence radiation.

A complete detection system for a pathogenic bacterium (*Listeria monocytogenes*) with external dimensions of 36 cm × 28 cm × 15 cm has been reported (Cady *et al.* 2005). The microfluidic system consisted of a connected DNA purification chamber and a second PCR amplification chamber, and was fabricated in a combination of silicon and PDMS using an SU-8 photoresist mold. However, the optical excitation and detection system was macroscopic and quite conventional, using a blue LED with an emission wavelength of 480 nm and 80 mW power for excitation and a PMT for detection of fluorescence signals at a wavelength of 520 nm. The excitation wavelength from the PMT was blocked by a notch filter.

Microfabrication techniques offer the advantage of the integration of a silicon photodetector and a blocking filter for the excitation light. Efficient blocking of the excitation light from the photodetector is indispensable in fluorescence experiments because of the low intensity of the fluorescence light, which is smaller by several orders of magnitude compared to the excitation intensity.

Such an integrated chip consisting of a glass chip with microfluidic channels which is bonded to a silicon-based chip with integrated photodiode and interference filter, has been fabricated (Namasivayam *et al.* 2004) and achieved a detection sensitivity of 0.9 ng μl^{-1}.

Finally, the development of arrayed single-photon detectors with a high temporal resolution (below 100 ps) for biophotonic applications is currently a hot topic in applied and industrial research (Pauchard and Rochas 2005).

3.09.7.4 Fluorescence Excitation

For excitation of an ordered array of fluorescence spots, the most straightforward approach (apart from flood illumination) is the arrangement of individual light sources, such as laser diodes, in an array with a pitch to match the spot pattern. However, this approach requires a completely new hardware design for every new spot array design. By several

microsystem-based approaches, this problem can be avoided in an economic and elegant manner.

Relying on the arrayed arrangement of the spots, several approaches use static or dynamic diffractive optical elements (DOEs) for providing an intensity distribution, which is matched against the spot pattern on the biochip. Diffractive optical elements have been used for multipoint holographic fluorescence excitation, and it has been demonstrated that this approach outperforms conventional methods with respect to excitation efficiency, signal-to-noise ratio and sample throughput (Blom *et al.* 2002, Mehta *et al.* 2001). For a new biochip design, only the DOE has to be exchanged on the hardware side.

Spatial light modulators, such as micromirror arrays and liquid crystal displays, can be used as dynamic diffractive optical elements, which allow flexible biochip excitation without exchanging hardware in the excitation unit. This was recently demonstrated using liquid crystal displays, as shown in **Figure 24** (Glasenapp *et al.* 2006). In this work, the holograms were calculated in advance to match the spot patterns and were loaded onto the liquid crystal display. The maximum hologram refresh rate using these displays was 100 Hz.

All holographic techniques for biochip fluorescence excitation require an accurately defined position of the spots on the chip. While accurate spotting of biochips is, in principle, not a problem in the current state of the art in microsystem technology, the reality is still different in many microbiological laboratories. Currently, coarse needle printers are frequently used for array spotting.

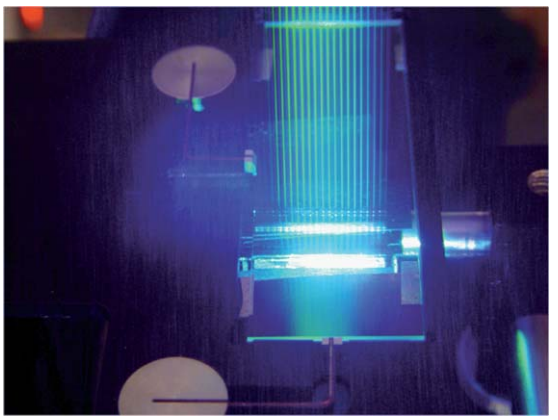


Figure 24 Photograph of fluorescence excitation using a holographically defined intensity pattern. The blue lines correspond to a linear array of excitation spots, which impinge on a biochip.

A completely different approach for multipoint excitation is the integration of low-cost dye microlasers into a microfluidic network (Kobayashi *et al.* 2005). The surface-emitting dye laser consisted of a dye-doped PMMA layer on a patterned glass substrate and was pumped optically by an external Nd:YAG-laser. Substrate emission occurs only over the patterned regions of the glass. By an appropriate choice of the periodicity of the pattern, the emission wavelength of the laser can also be selected. Therefore, emission at different wavelengths is possible within the same chip.

In biophotonics, providing an arrayed intensity distribution can be applied not only to fluorescence excitation of spot arrays on a biochip but also to parallel optical trapping and manipulation of biological specimens such as cells. As an example for this application, a liquid crystal display as a phase-only SLM has been used and the phase perturbation of the wavefront has been converted to an amplitude perturbation by means of a phase contrast filter (Mogensen and Glückstad 2000). A computing method for optimization of holographic optical traps and a for quantification of the performance of the holograms allowing them to be optimized in real time have been developed (Polin *et al.* 2005).

Acknowledgments

The authors are grateful to Khaled Aljasem, Jens Fiala, Christoph Friese, Carsten Glasenapp, David Kallweit, Nadja Katthagen, and Armin Werber of the Laboratory for Micro-optics in the Department of Microsystems Engineering for preparing some of the figures and tracking down some of the references; Jürgen Steigert, Jens Ducrée, Roland Zengerle of the Laboratory for MEMS applications in the same department for the pictures of the biodisk; Eiji Higurashi of the University of Tokyo for the photos of his blood flow sensor; and David Dickensheets of Montana State University for the diagram of his integrated confocal sensor head.

References

Adams M L, Enzelberger M, Quake S R, Scherer A 2003 Microfluidic integration on detector arrays for absorption and fluorescence micro-spectrometers. *Sens. Actuators A* **104**, 25–31

Arata H F, Noji H, Fujita H 2005 Direct temperature measurement in femtoliter chamber using fluorescent dye.

In: Proc. IEEE/LEOS Int. Conf. Optical MEMS, Oulu, Finland, pp. 77–8
Ashkin A 1992 Forces on a single-beam gradient laser trap on a dielectric sphere in the ray optics regime. *Biophys. J.* **61**, 569–82
Atkins P W, de Paula J 2006 *Physical Chemistry*, 8th edn. Oxford University Press, Oxford
Badugu R, Lakowicz J R, Geddes C D 2004 Noninvasive continuous monitoring of physiological glucose using a monosaccharide-sensing contact lens. *Anal. Chem.* **76(3)**, 610–18
Baechi D, Buser R, Dual J 2002 A high density microchannel network with integrated valves and photodiodes. *Sens. Actuators A* **95**, 77–83
Betzig E, Trautman J K 1992 Near-field optics: Microscopy, spectroscopy, and surface modification beyond the diffraction limit. *Science* **257(5067)**, 189–95
Blom H, Johansson M, Hedman A-S, Lundberg L, Hanning A, Hard S, Rigler R 2002 Parallel fluorescence detection of single biomolecules in microarrays by a diffractive-optical-designed 2×2 fan-out element. *Appl. Opt.* **41**, 3336–42
Bohren C F, Huffman D R 1998 *Absorption and Scattering of Light by Small Particles*. Wiley Interscience, New York
Born M, Wolf E 1999 *Principles of Optics*, 7th edn. Cambridge University Press, Cambridge
Bouma B E, Tearne G J 2002 *Handbook of Optical Coherence Tomography*. Marcel Dekker, New York
Brandenburg A, Krauter R, Künzel K, Stefan M, Schulte H 2000 Interferometric sensor for detection of surface-bound bioreactions. *Appl. Opt.* **39(34)**, 6396–405
Cady N C, Stelick S, Kunnavakkam M V, Batt C A 2005 Real-time PCR detection of *Listeria monocytogenes* using an integrated microfluidics platform. *Sens. Actuators B* **107**, 332–41
Chang-Yen D A, Gale B K 2005 Design and fabrication of a multianalyte-capable optical biosensor using a multiphysics approach. In: *Proc. 3rd IEEE Spec. Top. Conf. Microtechnology in Medicine and Biology,* Honolulu, HI, USA
Chang-Yen D A, Gale B K 2006 Design, fabrication, and packaging of a practical multianalyte-capable optical biosensor. *J. Microlithogr. Microfabricat. Microsyst.* **5(2)**, 021105
Chiou P Y, Moon H, Toshiyoshi H, Kim C-J, Wu M C 2003 Light actuation of liquid by optoelectrowetting. *Sens. Actuators A* **104**, 222–8
Chiou P Y, Ohta A T, Wu M C 2005 Massively parallel manipulation of single cells and microparticles using optical images. *Nature* **436**, 370–2
Cho S K, Moon H, Kim C-J 2003 Creating, transporting, cutting, and merging liquid droplets by electrowetting-based actuation for digital microfluidic circuits. *J. Microelectromech. Syst.* **12(1)**, 70–80
Dario P, Carrozza M C, Pietrabissa A 1999 Development and *in vivo* testing of a miniature robotic system for computer-assisted colonoscopy. *Comput. Aided Surg.* **4(1)**, 1–14
Dario P, Carrozza M C, Benvenuto A, Menciassi A 2000a Micro-systems in biomedical applications. *J. Micromech. Microeng.* **10**, 235–44
Dario P, Carrozza M C, Marcacci M, D'Attanasio S, Magnami B, Tonet O, Megali G 2000b A novel mechatronic tool for computer-assisted arthroscopy. *IEEE Trans. Inf. Tech. Biomed.* **4(1)**, 15–29
Dickensheets D L, Kino G S 1998 Silicon-micromachined scanning confocal optical microscope. *IEEE/ASME J. Microelectromech. Syst.* **7(1)**, 38–47
Divetia A, Hsieh T-H, Zhang J, Chen Z, Bachmann M, Li G-P 2005 Dynamically focused optical coherence tomography for endoscopic applications. *Appl. Phys. Lett.* **86**, 103902
Doble N, Yoon G, Chen L, Bierden P, Singer B, Olivier S, Williams D R 2002 Use of a microelectromechanical mirror for adaptive optics in the human eye. *Opt. Lett.* **27(17)**, 1537–9
Dong L, Agarwal A K, Beebe D J, Jiang H 2006 Adaptive liquid microlenses activated by stimuli-responsive hydrogels. *Nature* **442**, 551–4
El-Ali J, Sorger P K, Jensen K F 2006 Cells on chips. *Nature* **442**, 403–11
Friese C, Mader D, Zappe H 2004 Micro-mirror arrays for adaptive optics fabricated in polymer technology. *In: Proc. 2004 IEEE/LEOS Int. Conf. Optical MEMS,* Takematsu, Japan, pp. 214
Fu L, Jain A, Xie H, Cranfield C, Gu M 2006 Nonlinear optical endoscopy based on a double-clad photonic crystal fiber and a MEMS mirror. *Opt. Express* **14(3)**, 1027–32
Furihata H 2001 New technologies in endoscopic surgery. *Minim. Invasive Ther. Allied Technol.* **10(4–5)**, 199–203
Gambin Y, Legrand O, Quake S R 2006 Microfabricated rubber microscope using soft solid immersion lenses. *Appl. Phys. Lett.* **88**, 174102
Geddes C D, Lakowicz J R 2004 Contact lenses may provide window to blood glucose. *Biophotonics Int.* **February(2)**, 50–3
Glasenapp C, Mönch W, Krause H, Zappe H 2006 Biochip reader with dynamic holographic excitation and hyperspectral fluorescence detection. *J. Biomed. Opt.* **12(1)**, 014038
Glukhovsky A 2004 The development and application of wireless capsule endoscopy. *Int. J. Med. Robot. Comput. Assist. Surg.* **1(1)**, 114–23
Grumann M, Steigert J, Riegger L, Moser I, Enderle B, Riebeseel K, Urban G, Zengerle R, Ducrée J 2006 Sensitivity enhancement for colorimetric glucose assays on whole blood by on-chip beam guidance. *Biomed. Microdevices* **8(3)**, 209–14
Gunji M, Washizu M 2005 Self-propulsion of a water droplet in an electric field. *J. Phys. D Appl. Phys.* **38**, 2417–23
Hafeman D G, Harkins I V J B, Witkowski II C E, Lewis N S, Warmack R J 2006 Optically directed molecular transport and 3d isoelectric positioning of amphoteric biomolecules. *Proc. Natl Acad. Sci.* **103(17)**, 6436–41
Haga Y, Esashi M 2004 Biomedical microsystems for minimally invasive diagnosis and treatment. *Proc. IEEE* **92(1)**, 98–114
Hale G M, Querry M R 1973 Optical constants of water in the 200 nm to 200 μm wavelength region. *Appl. Opt.* **12**, 555–63
Hammer D X, Ferguson R D, Begelow C E, Iftimia N V, Ustun T E 2006 Adaptive optics scanning laser ophthalmoscope for stabilized retinal imaging. *Opt. Express* **14(8)**, 3354–67
Han J 1999 Fabry–Perot cavity chemical sensors by silicon micromachining techniques. *Appl. Phys. Lett.* **74(3)**, 445–7
Hecht E 2003 *Optics*, 4th edn. Addison-Wesley, Boston, MA
Hell S, Stelzer H K 1992 Properties of a 4Pi confocal fluorescence microscope. *J. Opt. Soc. Am. A* **9(12)**, 2159–66
Herzig H P (*ed*.) (1997) *Micro-optics: Elements, Systems and Applications.* Taylor & Francis, London
Higurashi E, Sawada R, Ito T 2002 Hybrid integrated optical sensor for noninvasive blood flow monitoring. *In: Proc. 2002 IEEE/LEOS Int. Conf. Optical MEMS,* Lugano, Switzerland, pp. 33–4
Homola J (*ed*.) (2006) *Surface Plasmon Resonance Based Sensors.* Springer-Verlag, Berlin
van de Hulst H C 1982 *Light Scattering by Small Particles*. Dover Publications, New York
Iddan G, Meron G, Glukhovsky A, Swain P 2000 Wireless capsule endoscopy. *Nature* **405**, 417
Iizuka K 2002 *Elements of Photonics.* Wiley-Interscience, New York, Vols. I & II
James M W, Gu C, Gu M 2004 Scanning total internal reflection fluorescence microscopy under one-photon and two-photon excitation: Image formation. *Appl. Opt.* **43**, 1063–71
Janasek D, Franzke J, Manz A 2006 Scaling and the design of miniaturized chemical-analysis systems. *Nature* **442**, 374–80

Jeong K-H, Kim J, Lee L 2006 Biologically inspired artificial compound eyes. *Science* **312**, 557–61

Johannessen E A, Wang L, Cui L, Tang T B, Ahmadian M, Astaras A, Reid S W J, Yam P S, Murray A F, Flynn B W, Beaumont S P, Cumming D R S, Cooper J M 2004 Implementation of multi-channel sensors for remote biomedical measurements in a microsystems format. *IEEE Trans. Biomed. Eng.* **51(3)**, 525–35

Jones T B, Gunji M, Washizu M, Feldman M J 2001 Dielectrophoretic liquid actuation and nanodroplet formation. *J. Appl. Phys.* **89(3)**, 1–8

Katsumata K, Haga Y, Minami K, Esashi M 2000 Micromachined 125 μm diameter ultra miniature fiber-optic pressure sensor for catheter. *Trans. Inst. Electr. Engrs Jpn* **120E(2)**, 58–63

Ke C, Berney H, Mathewson A, Sheehan M M 2004 Rapid amplification for the detection of *Mycobacterium tuberculosis* using a non-contact heating method in a silicon microreactor based thermal cycler. *Sens. Actuators B* **102**, 308–14

Kim J, Serpe, Lyon L A 2004 Hydrogel microparticles as dynamically tunable microlenses. *J. Am. Chem. Soc.* **126**, 9512–13

Kim J, Nayak S, Lyon L A 2005 Bioresponsive hydrogel microlenses. *J. Am. Chem. Soc.* **127**, 9588–92

Kobayashi T, Kanamori Y, Hane K 2005 A surface emitting solid polymer dye laser using a guided mode resonant grating for bio-optical detections. *In: Proc. IEEE/LEOS Int. Conf. Optical MEMS,* Oulu, Finland, pp. 101–2

Krogmann F, Mönch W, Zappe H 2006 A MEMS-based variable micro-lens system. *J. Opt. A Pure Appl. Opt.* **8**, S330–6

Kuo W C, Chou C, Wu H-T 2003 Optical heterodyne surface-plasmon resonance biosensor. *Opt. Lett.* **28(15)**, 1329–31

Lechuga L M, Tamayo J, Álvarez M, Carrascosa L G, Yufera A, Doldán R, Peralías E, Rueda A, Plaza J A, Zinoviev K, Domínguez C, Zaballos A, Moreno M, Martínez-A C, Wenn D, Harris N, Bringer C, Bardinal V, Camps T, Vergnenègre C, Fontaine C, Díaz V, Bernad A 2006 A highly sensitive microsystem based on nanomechanical biosensors for genomics applications. *Sens. Actuators B* **118(1–2)**, 2–10

Lehr H-P, Brandenburg A, Sulz G 2003 Modeling and experimental verification of the performance of TIRF-sensing systems for oligonucleotide microarrays based on bulk and integrated optical planar waveguides. *Sens. Actuators B Chem.* **92**, 303–14

Liang J, Williams D R, Miller D T 1997 Supernormal vision and high-resolution retinal imaging through adaptive optics. *J. Opt. Soc. Am. A* **14(11)**, 2884–92

Ligler F S, Rowe-Taitt C A (eds.) (2002) *Optical Biosensors: Present and Future.* Elsevier, Amsterdam

Liu D, Garimella S V, Werely S T 2005 Infrared micro-particle image velocimetry in silicon-based microdevices. *Exp. Fluids* **38**, 385–92

Luff B J, Harris R D, Wilkenson J S, Wilson R, Schiffrin D J 1996 Integrated optical directional coupler biosensor. *Opt. Lett.* **21(8)**, 618–20

Luff B J, Wilkenson J S, Piehler J P, Hollenbach U, Ingenhoff J, Fabricius N 1998 Integrated optical Mach–Zehnder biosensor. *J. Lightwave Technol.* **16**, 583–92

Maisenhölder B, Zappe H, Kunz R E, Riel P, Moser M, Duveneck G L 1996 A GaAs/AlGaAs-based Mach–Zehnder interferometer as integrated optical immunosensor. *Biomedical Systems and Technologies. SPIE Proceedings*, Vienna, Austria, Vol. 2928, p. 144

Maisenhölder B, Zappe H, Kunz R E, Riel P, Moser M, Edlinger J 1997a A GaAs/AlGaAs refractometer platform for integrated optical sensing applications. *Sens. Actuators B* **39(1–3)**, 324–9

Maisenhölder B, Zappe H, Moser M, Riel P, Kunz R E, Edlinger J 1997b A monolithically integrated optical interferometer for refractometry. *Electron. Lett.* **32(11)**, 986–8

Mehrvar M, Bis C, Scharer J M, Moo-Young M, Luong J H 2000 Fiberoptic biosensors – Trends and advances. *Anal. Sci.* **16**, 677–92

Mehta D S, Lee C Y, Choiu A 2001 Multipoint parallel excitation and CCD-based imaging system for high-throughput fluorescence detection of biochip micro-arrays. *Opt. Commun.* **190**, 59–68

Menciassi A, Moglia A, Gorini S, Pernorio G, Stefanini C, Dario P 2005 Shape memory allow clamping devices of a capsule for monitoring tasks in the gastrointestinal tract. *J. Micromech. Microeng.* **15**, 2045–55

Meron G 2000 The development of the swallowable video capsule. *Gastrointest. Endosc.* **52(6)**, 817–19

Minas G, de Graaf G, Wolffenbuttel R F, Correia J H 2006 An MCM-based microsystem for colorimetric detection of biomolecules in biological fluids. *IEEE Sens. J.* **6(4)**, 1003–9

Misiakos K, Kakabakos S E, Petrou P S, Ruf H H 2004 A monolithic silicon optoelectronic transducer as a real-time affinity biosensor. *Anal. Chem.* **76(5)**, 1366–73

Mitra B, Wilson C G, Que L, Selvaganapathy P, Gianchandani Y B 2006 Microfluidic discharge-based optical sources for detection of biochemicals. *Lab Chip* **6**, 60–5

Mogensen P C, Glückstad J 2000 Dynamic array generation and pattern formation for optical tweezers. *Opt. Commun.* **175**, 75–81

Mönch W, Dehnert J, Jaufmann E, Zappe H 2006a Flory–Guggins swelling of polymer Bragg mirrors. *Appl. Phys. Lett* **89(16)**, 164104

Mönch W, Dehnert J, Prucker O, Rühe J, Zappe H 2006b Tunable Bragg filters based on polymer swelling. *Appl. Opt.* **45(18)**, 4284–90

Motamedi M E (*ed*.) (2005) *MOEMS.* SPIE Press, Bellingham, WA

Mugele F, Baret J-C 2005 Electrowetting: From basics to applications. *J. Phys. Condens. Matter* **17**, R705–74

Namasivayam V, Lin R, Johnson B, Brahmasandra S, Razzacki Z, Burke D T, Burns M A 2004 Advances in on-chip photodetection for applications in miniaturized genetic analysis systems. *J. Micromech. Microeng.* **14**, 81–90

Pauchard A, Rochas A 2005 Single-photon counters get a second wind. *Photonics Spectra, 102–6*

Piyawattanametha W, Baretto R P J, Ko T H, Flusberg B A, Cocker E D, Ra H, Lee D, Solgaard O, Schnitzer M J 2006 Fast-scanning two-photon fluorescence imaging based on a microelectromechanical systems two-dimensional scanning mirror. *Opt. Lett.* **31(13)**, 2018–20

Polin M, Ladavac K, Lee S-H, Roichman Y, Grier D G 2005 Optimized holographic optical traps. *Opt. Express* **13(15)**, 5831–45

Potsaid B, Bellouard Y, Wen J T 2005 Adaptive scanning optical microscope (ASOM): A multidisciplinary optical microscope design for large field of view and high resolution imaging. *Opt. Express* **13(17)**, 6504–18

Prahl S 2006 Optical absorption of hemoglobin. URL. http://omlc.ogi.edu/spectra/hemoglobin/index.html

Prasad P N 2003 *Introduction to Biophotonics*. Wiley-Interscience, Hoboken, NJ

Psaltis D, Quake S R, Yang C 2006 Developing optofluidic technology through the fusion of microfluidics and optics. *Nature* **442**, 381–6

Reesink K D, van der Nagel T, Bovelander J, Jansen J R C, van der Veen F H, Schreuder J J 2002 Feasibility study of a fiber-optic system for invasive blood pressure measurements. *Cathet. Cardiovasc. Interv.* **57(2)**, 272–6

Ribeiro J C, Minas G, Turmezei P, Wolffenbuttel R F, Correia J H 2005 A SU-8 fluidic microsystem for biological fluids analysis. *Sens. Actuators A* **123–124**, 77–81

Rohrbach A, Huisken J, Stelzer E H K 2003 Optical trapping of small particles. In: Kao F-J and Török P (eds.) *Optical Imaging and Microscopy* Springer-Verlag, Berlin, pp. 357–88

Roorda A, Williams D R 1999 The arrangement of the three cone classes in the living human eye. *Nature* **397**, 520–2

Roulet J-C, Völkel R, Herzig H P, Verpoorte E, de Rooij N F, Dändliker R 2001 Microlens systems for fluorescence detection in chemical microsystems. *Opt. Eng.* **40(5)**, 814–21

Schabmueller C G J, Pollard J R, Evans A G R, Wilkinson J S, Ensell G, Brunnschweiler A 2001 Integrated diode detector and optical fibres *in situ* detection within micromachined polymerase chain reaction chips. *J. Micromech. Microeng.* **11**, 329–33

Schena M, Shalon D, Davis R W, Brown P O 1995 Quantitative monitoring of gene expression patterns. *Science* **270**, 467–70

Semwogerere D, Weeks E R 2004 *Confocal Microscopy. Encyclopedia of Biomaterials and Biomedical Engineering*. Taylor & Francis, London

Seo J, Lee L P 2004 Disposable integrated microfluidics with self-aligned planar microlenses. *Sens. Actuators B* **99**, 615–22

Sepúlveda B, Sánchez del Río J, Moreno M, Blanco F J, Mayora K, Domínguez C, Lechuga L M 2006 Optical biosensor microsystems based on the integration of highly sensitive Mach–Zehnder interferometer devices. *J. Opt. A* **8**, S561–6

Sinzinger S, Jahns J 2003 *Microoptics*, 2 edn. Wiley-VCH, Weinheim

Slavík R, Homola J, Vaisocherová H 2006 Advanced biosensing using simultaneous excitation of short and long range surface plasmons. *Meas. Sci. Technol.* **17**, 932–8

Slyadnev M N, Tanaka Y, Tokeshi M, Kitamori T 2001 Photothermal temperature control of a chemical reaction on a microchip using an infrared diode laser. *Anal. Chem.* **73**, 4037–44

Steigert J, Grumann M, Brenner T, Mittenbühler K, Nann T, Rühe J, Moser I, Haeberle S, Riegger L, Riegler J, Zengerle R, Ducrée J 2005 Integrated sample preparation, reaction, and detection on all high-frequency centrifugal microfluidic platform. *J. Assoc. Lab. Automat.* **10(5)**, 331–41

Su F, Chakrabarty K, Fair R B 2006 Microfluidics-based biochips: Technology, issues, implementation platforms, and design-automation challenges. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **25(2)**, 211–23

Tanaka Y, Slyadnev M N, Hibara A, Tokeshi M, Kitamori T 2000 Non-contact photothermal control of enzyme reactions on a microchip by using a compact diode laser. *J. Chromatogr. A* **894**, 45–1

Thrush E, Levi O, Cook L J, Deich J, Kurtz A, Smith S J, Moerner W E, Harris J S 2005 Monolithically integrated semiconductor fluorescence sensor for microfluidic applications. *Sens. Actuators B* **105**, 393–9

Tohyama O, Kohashi M, Sugihara M, Itoh H 1998 A fiber-optic pressure microsensor for biomedical applications. *Sens. Actuators A* **66**, 150–4

Tran P H, Mukai D S, Brenner M, Chen Z 2004 *In vivo* endoscopic optical coherence tomography by use of a rotational microelectromechanical system probe. *Opt. Lett.* **29(11)**, 1236–8

Tyson R 1997 *Principles of Adaptive Optics*. Academic Press, New York

Valentini G, D'Andrea C 2000 Time-resolved DNA-microarray reading by intensified CCD for ultimate sensitivity. *Opt. Lett.* **25**, 1648–50

Vdovin G, Sarro P, Middelhoek S 1999 Technology and applications of micromachined adaptive mirrors. *J. Micromech. Microeng.* **9(2)**, R8–20

Visscher K, Brakenhoff G J 1991 Theoretical study of optically induced forces on spherical particles in a single beam trap i: Rayleigh scatterers. *Optik* **89**, 174–80

Visscher K, Brakenhoff G J 1992 Theoretical study of optically induced forces on spherical particles in a single beam trap ii: Mie scatterers. *Optik* **90**, 57–60

Vo-Dinh T, Cullum B M, Stokes D L 2001 Nanosensors and biochips: Frontiers in biomolecular diagnostics. *Sens. Actuators B* **74**, 2–11

Vollmer F, Braun D, Libchaber A, Khoshsima M, Teraoka I, Arnold S 2002 Protein detection by optical shift of a resonant microcavity. *Appl. Phys. Lett.* **80(21)**, 4057–9

Werber A, Zappe H 2005 Tunable microfluidic microlenses. *Appl. Opt.* **44(16)**, 3238–45

Whitesides G W 2006 The origins and the future of microfluidics. *Nature* **442**, 368–73

Wolfenbuttel R F 2005 MEMS-based optical mini- and microspectrometers for the visible and infrared spectral range. *J. Micromech. Microeng.* **15**, S145–52

Wolthuis R A, Mitchell G L, Saaski E, Hartl J C, Afromowitz M A 1991 Development of medical pressure and temperature sensors employing optical spectrum modulation. *IEEE Trans. Biomed. Eng.* **38(10)**, 974–80

Wu S Y, Ho H P, Law W C, Lin C, Kong S K 2004 Highly sensitive differential phase-sensitive surface plasmon resonance biosensor based on the Mach–Zehnder configuration. *Opt. Lett.* **29(20)**, 2378–80

Xie T, Xie H, Fedder G K, Pan Y 2003 Endoscopic optical coherence tomography with a microelectromechanical systems mirror for detection of bladder cancers. *Appl. Opt.* **42(31)**, 6422–6

Zappe H 1995 *Introduction to Semiconductor Integrated Optics*. Artech House, Boston, MA

Zappe H 1999 Semiconductor optical sensors. In: Baltes H, Göpel W, and Hesse J (eds.) *Sensors Update.* Wiley-VCH, Vol. 5, Chap. 1, pp. 3–44

Zappe H 2004 *Laser Diode Microsystems*. Springer-Verlag, Berlin

Zappe H, Hess M, Moser M, Hövel R, Gulden K, Gauggel H P, Monti di Sopra F 2000 Narrow-linewidth vertical-cavity surface-emitting lasers for oxygen detection. *Appl. Opt.* **39(15)**, 2475

Zhang C, Xu J, Ma W, Zheng W 2006a PCR microfluidic devices for DNA amplification. *Biotechnol. Adv.* **24**, 243–84

Zhang Y, Poonja S, Roords A 2006b MEMS-based adaptive optics scanning laser ophthalmoscopy. *Opt. Lett.* **31(9)**, 1268–70

Zworkin V K 1957 Radio pill. *Nature* **179**, 898

Biographies

Wolfgang Mönch was born in 1970 in Oettingen, Germany. He studied Physics at the University of Konstanz and graduated there in 1996. From 1996 to 1999, he worked with Prof. Herminghaus at the Max Planck Institute for Colloids and Interfaces in Berlin, Germany, on the dynamics of liquids on structured substrates and earned his doctorate in 1999. In 2000/2001, he investigated elastic instabilities of rubber films at the University of Ulm. Since July 2001, he has been a senior scientist in Professor Zappe's group at the Department of Microsystems Engineering, University of Freiburg, Germany. His current research interests are biophotonics and applications of soft matter in tunable optical microsystems and novel tuning mechanisms in micro-optics.

Hans Zappe was born in Paris and brought up in New York. He studied Electrical Engineering at the Massachusetts Institute of Technology (BSc and MSc, 1983) and at the University of California, Berkeley (PhD, 1989). He has worked on silicon VLSI at the IBM General Technology Division (Burlington, VT, USA), on GaAs electronics and high-speed lasers at the Fraunhofer Institute for Applied Solid State Physics (Freiburg, Germany), and on integrated optical microsystems and surface-emitting lasers at the Centre Suisse d'Electronique et de Microtechnique (CSEM, Zurich, Switzerland). Since 2000, Prof. Zappe has been Professor of Micro-optics in the Department of Microsystems Technology at the University of Freiburg, Germany. His current research specialties are in the areas of optical microsystems for medicine, tunable micro-optics, and the use of novel optical materials.

3.10 RF MEMS

Nickolas Kingsley and John Papapolymerou, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Glossary

IC Integrated Circuit
LCP Liquid Crystal Polymer
RF MEMS Radio Frequency Microelectromechanical Systems
SOC System-on-Chip
SOP System-on-Package

3.10.1 RF MEMS Switches

Switches are an integral part of electronics. It is no surprise that there are countless types of switches, which can range from low-cost PIN diodes to high-power electromagnetic relays. Microelectromechanical systems (MEMS) switches are quickly becoming very popular with microwave engineers, and they have already been used in antennas, filters, phase shifters, oscillators, multiplexers, and other traditional (radio frequency) RF devices. Their popularity has grown because of their wider bandwidth, better isolation, lower loss at high frequencies, improved linearity, and lower power consumption compared with traditional switch devices (Goldsmith *et al.* 2001).

3.10.1.1 Introduction

MEMS switches as we know them today were conceptualized in the late 1980s and early 1990s. The first published articles were entirely conceptual and pointed out some of the design challenges and potential uses of MEMS devices (Koester *et al.* 1996). MEMS switches in particular were of great interest to RF engineers for their potential to reduce the total area, power consumption, and cost of their devices. A picture of an RF MEMS switch is shown in **Figure 1**.

MEMS were initially fabricated exclusively on silicon because integrated circuit (IC) fabrication at the time was on silicon. The material properties and fabrication processes for silicon were already well known.

MEMS devices quickly surpassed the RF performance of their solid-state equivalents. Even early MEMS switches had an insertion loss of 0.15 dB at 20 GHz, compared with an on-state insertion loss of approximately 1 dB for a typical GaAs-FET or PIN diode switch at the same frequency (Brown 1997).

For applications where the frequency of interest is less than 1 GHz, solid-state switches are still the switching element primarily used. They are cheap, have low loss, are easy to integrate, and are widely available. Above a few gigahertz, the losses begin to increase, and it becomes more difficult to integrate the switches. This is where the advantages of MEMS switches

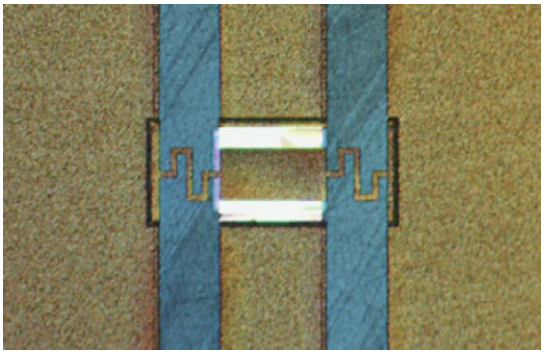


Figure 1 A single radio frequency microelectromechanical systems (RF MEMS) switch is shown in a coplanar waveguide structure. The metal pads on the *left* and the *right* side of the image serve as the DC and the RF ground. The center metal serves as the DC and RF carrier. A thin layer of silicon nitride prevents a direct connection between the layers during switch actuation. (Source: Thompson D, Kingsley N, Wang G, Papapolymerou J, Tentzeris M 2005 RF characteristics of thin-film liquid crystal polymer (LCP) packages for RF MEMS and MMIC integration. *IEEE MTT-S Int. Microwave Symp. Dig.,* © 2005; reproduced with permission from IEEE.)

become apparent. They are not as fast nor are they as robust as their solid-state cousins, but they can more than make up for that in electrical performance. A MEMS switch can easily have losses in tenths of a decibel up to 40 GHz. Switching times are typically measured in tens of microseconds for hundreds of billions of cycles. In recent years, MEMS switches that can handle multiwatt power levels have been reported (Que *et al.* 2004).

The holy grail for many MEMS designers is the cell phone market. The global wireless market generated $555 billion in 2005 and this number is expected to grow to $800 billion by 2010 (Reuters 2006). This presents a large funding source for any company that can integrate RF MEMS into a cell phone. Utilizing a bank of RF MEMS switches, capacitors, and inductors, a cell phone could offer fantastic reconfigurability. The cell phone could conceivably work at any frequency, on any channel, for any standard, and in any location. Dropped calls would be a thing of the past. Many performance limitations would arise when trying to do this with solid-state switches.

3.10.1.2 Operation

RF MEMS switches are one type of MEMS devices that utilize either a single-supported (cantilever) or a double-supported (air-bridge) beam suspended over a metal pad. Because a MEMS switch uses only a single moving part, it is one of the simplest MEMS in use today. By comparison, a sensor can have dozens of moving parts. Switches can come in various shapes, sizes, and materials. There are two main types of actuation for RF MEMS switches: thermal and electrostatic.

Most materials expand when heated and contract when cooled. This is the basic principle behind thermal switches. A resistive material is integrated on the switch membrane. Electrical current is passed through the switch, which heats the resistive material and causes it to expand. This expansion deflects the beam. When the current is reduced (or eliminated), the switch returns to the steady state. This switch topology is not widely used because it is much slower, has a lower bandwidth, consumes more power, and is more difficult to control than with electrostatic actuation; also, it results in higher losses. It can have a low actuation voltage, which could make it attractive for system-on-chip applications.

Electrostatic actuation relies on the principle that opposite charges attract. With one metal beam suspended above a metal pad, a voltage is applied to the beam while the pad is grounded or *vice versa* (as shown in **Figure 2**). A static charge will form from the voltage potential, and this creates an electrostatic

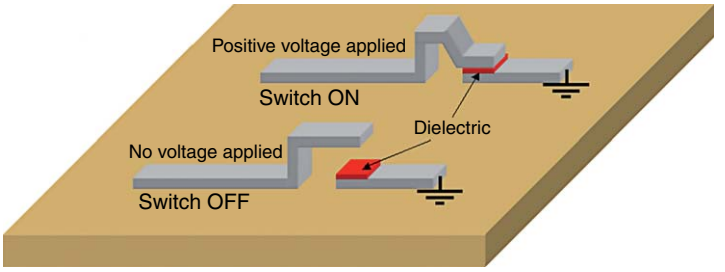


Figure 2 Basic operation of a single-supported, capacitive microelectromechanical systems (MEMS) switch. When no voltage is applied to the membrane, no actuation occurs. When a voltage is applied, the voltage potential creates an electrostatic force that pulls the membrane toward the grounded sign line beneath it. A thin layer of dielectric material prevents direct contact between the layers.

force between the layers. As the voltage potential is increased, the electrostatic force strengthens. When this force exceeds the beam's ability to resist deflection, the metal layers are pulled together. If the metal layers are allowed to make direct contact, this switch is ohmic and DC current is able to flow through the switch. If direct contact is prevented by a thin layer of dielectric (usually silicon nitride), then the switch is capacitive and no DC current is allowed to flow. Because a capacitor is the basis for this design, the frequency must be sufficiently high so that the RF energy can pass through. Switches of this type are typically used from 5 to 100 GHz. Varying the thickness of the dielectric layer is one way of tuning the resonant frequency of these switches. Filter designers are particularly fond of capacitive switches because there is no resistance and this gives a higher Q-factor.

MEMS switch designers have a number of variables at their disposal that can be optimized for a given application (Senturia 2001). There is a fundamental trade-off between isolation, switching speed, and actuation voltage. The best way of improving the isolation of a switch is to increase the vertical distance between the beam and the metal pad. This distance is typically 1–3 μm. As this distance is increased, the switching time and the actuation voltage also increase. Decreasing this distance will likewise decrease the switching time and the actuation voltage. The speed and voltage can also be improved by changing the beam material. Very pliable metals, such as aluminum, will switch more easily than stiffer metals, such as gold. The stiffness of a material is denoted by its Young's modulus (the higher the value, the stiffer the material).

Equations for predicting the bending of cantilever and double-supported beams have been around for decades. Unfortunately, trying to apply simplistic equations to complex MEMS devices can be cumbersome. One of the most important mechanical parameters of a MEMS switch is the pull-down voltage. This quantity can be estimated by treating the MEMS switch as a mechanical spring. In order to calculate the pull-down voltage, one must equate the electrostatic force pulling down on the beam

$$f_{\text{down}} = \frac{\varepsilon A V^2}{2g^2} \quad [1]$$

and the force pushing up from the spring (Hooke's law)

$$f_{\text{up}} = -k(g_0 - g) \quad [2]$$

For these equations, ϵ is the permittivity, A the area, V the voltage, k the spring constant, g_0 the initial gap, and g the evaluated gap. We can use these simple, spatially independent equations because we know the charge density (and therefore the force) is uniform across the capacitive region. It is well known that for parallel plate electrostatic actuation, when the gap reduces to two-thirds of the original gap, the beam becomes unstable and experiences a pull-in effect. That is, when the gap reaches a certain threshold, namely, two-thirds the original gap, the switch will snap down. Magnets experience the same effect. As two magnets of opposite polarity are brought closer together, the attractive force is barely noticeable until they reach a certain distance apart. At this point they snap together and the force between them is strong.

Equating eqns [1] and [2] where the gap (g) is two-thirds the original gap (g_0) and solving for the pull-down voltage gives

$$V_{\text{PD}} = \sqrt{\frac{8kg_0^3}{27\varepsilon A}} \quad [3]$$

In order to reduce the pull-down voltage, design engineers can reduce the spring constant, reduce the switch gap, or increase the area of the switch.

3.10.1.3 Comparison of Technologies: Electromechanical vs. Solid State

There are various switching elements available on the market today. In order to choose the right type of switch, one must consider the required performance specifications, such as frequency, bandwidth, linearity, power handling, power consumption, switching speed, signal level, and allowable losses. A comparison of electrical performance parameters for a typical RF MEMS, PIN diode, and FET switching element is summarized in **Table 1** (NASA 2003).

PIN diodes are useful switching elements because they are widely available commercially, have fast switching speeds, are of low cost, and are rugged. One of the main limitations of PIN diodes is the insertion loss. Above a few gigahertz, PIN diodes can start to have quite a bit of insertion loss. This becomes significant above X-band (8–10 GHz) when the skin effect causes an increase in the resistance of the switch. At any high frequency, diodes tend to have issues with linearity, bandwidth, and power consumption.

Like PIN diodes, FETs are widely used because of their availability, fast speed, low cost, and durability.

Table 1 Comparison of electrical performance for a typical RF MEMS, PIN diode, and FET switch

Parameter	*RF MEMS*	*PIN diode*	*Field-effect transistor*
Voltage (V)	20–80	±3–5	3–5
Current (mA)	0	0–20	0
Power consumption (mW)	<0.5	5–100	−0.5 to 0.1
Switching time	1–300 uS	1–100 ns	1–100 ns
Cup (series) (fF)	1–6	40–80	70–140
Rs (series) (Ω)	0.5–2	2–4	4–6
Capacitance ratio	40–500	10	N/A
Cutoff frequency (THz)	20–80	1–4	0.5–2
Isolation (1–10 GHz)	Very high	High	Medium
Isolation (10–40 GHz)	Very high	Medium	Low
Isolation (60–100 GHz)	High	Medium	None
Insertion loss (1–100 GHz) (dB)	0.05–0.2	0.3–1.2	0.4–2.5
Power handing (W)	<1	<10	<10
Third-order intercept (dBm)	+66–80	+27–45	+27–45

They consume much less power than PIN diodes, but they are not available for the same frequency range as diodes. That is, they have limited use in Ka-band (26–40 GHz) and are practically unusable above U-band (40–60 GHz) (NASA 2003).

MEMS switches are quickly becoming the preferred switching element for RF devices. They offer the lowest insertion loss, the highest isolation, an extremely high linearity, a negligible power consumption, and a small size. The switching time, power handling capability, and packaging requirements are the three main limitations in their use. The switching time for a MEMS device usually varies with isolation because of physical constraints (the better the isolation, the slower the switching time). However, for a microwave system that can tolerate a switching time in the micro- to millisecond range, MEMS are suitable. Switching times in the hundreds of nanoseconds have also been demonstrated (Mercier *et al.* 2005). Furthermore, if signal amplification in a wireless system can be done right before propagation (thereby eliminating the exposure of the switching elements to high power), then MEMS are also applicable. Packaging MEMS is not as straightforward as packaging solid-state devices. This topic is discussed in Section 3.10.4.

3.10.1.4 Limitations

One of the most difficult challenges for MEMS designers is overcoming dielectric charging. All electrostatic MEMS switches use some sort of dielectric to maintain the voltage potential. Over time this dielectric will store charge and the switch will remain in the actuated state. This charge naturally dissipates into the substrate, but it can take anywhere from milliseconds to hours for this to occur, depending on the actuation voltage, the substrate material, and the extent of the charging. Lower actuation voltages and materials with good charge mobility tend to dissipate charge faster. Cycle testing is a good metric for determining the lifetime of a switch, but it is not a good indicator of the reliability. Switch-cycle lifetimes in the hundreds of billions of cycles have been demonstrated (Radant MEMS 2006). A switch that can operate for a hundred billion cycles is sure to have a long lifetime, but if the switch becomes stuck after a minute in the actuated state, then it is not a very reliable switch. For a system that is not continuously being reconfigured, this is a big problem. Researchers are currently investigating better ways of removing this charge accumulation (Papaioannou *et al.* 2005). Many techniques that are being proposed involve the use of novel dielectric materials.

Another solution to the problem of long-duration actuation is to use a hybrid actuation technique, which uses electrostatic and thermal mechanisms. Electrostatic actuation is used first to quickly pull down the switch. The switch is then held down using thermal actuation and the electrostatics are turned off. This switch has the fast switching time of an electrostatic switch; however, the drawback is that there is a slight increase in the power consumed (by the resistive material). Because the static charge is only present for a very short time, this switch can be much more reliable.

Robustness is another big challenge for engineers. In order for a product to be successful, it needs to be able to survive a certain amount of abuse. Modern cell phones can survive years of abuse from keys,

pocket change, multiple falls, and sometimes even brief exposures to water. Could a MEMS switch survive the same? Many experts say that as their size is so small, they are almost unaffected by everyday vibrations. Other experts are more skeptical.

One last limitation worth mentioning is the power handling capability of MEMS. Typically, the geometry of a switch requires the RF signal to travel directly under the metal membrane. As the power level for the RF signal is increased, it can start to have an effect on the membrane. At some power level, the RF signal will become strong enough to self-actuate the switch. This is a fundamental limitation and is something that is also actively under investigation. At the moment, switches are typically rated in the tens or hundreds of milliwatts (Radant MEMS 2006) although power levels over a watt have been demonstrated (Que *et al.* 2004).

3.10.2 System-Level Approach to RF MEMS

Many commercial companies are investing top dollar to integrate RF MEMS into their products. Like many MEMS researchers, they have realized that in a device where low power consumption, low loss, low cost, and high performance are essential, there is no substitute for a MEMS switch. NASA, MIT Lincoln Labs, Intel, IBM, and Raytheon are just a select few who are trying to incorporate MEMS into their state-of-the-art systems. Radant MEMS and Memtronics (MEMtronics Corporation 2005) are two examples of commercial companies that have system-level switches available for purchase. This section will discuss several systems where RF MEMS switches have been successfully integrated.

3.10.2.1 MEMS on Organic Substrates

As consumers push for improved wireless systems, there is a related push from microwave design engineers for lightweight, low-loss, and low-cost substrate materials. Unfortunately, this creates an interesting problem. In a typical wireless system, one would find various active and passive devices. Most of the active devices require expensive substrate materials such as GaAs or InP. Making an entire wireless system with these materials would be impractical. Alternatively, active devices can be integrated onto a less expensive substrate with the passive devices. It is very difficult to find a substrate material that is of low cost, has low loss at high frequencies, and can easily package active devices.

Fortunately, liquid crystal polymer (LCP) is one such material. LCP is a thin-film, flexible, organic polymer composed of oxygen, hydrogen, and carbon atoms. The exact chemical formula is proprietary and varies with the manufacturer. LCP is commercially available from Rogers Corporation (R/flex® 3000), Solvay Advanced Polymers (XYDAR® LCP), RTP Company (RTP Compounds LCP), DuPont Engineering Polymers (Zenite® LCP), Ticona (Vectra® LCP), and several others.

LCP has been characterized up to 110 GHz as a low-loss, low-cost material (Thompson *et al.* 2004). The line loss for a microstrip structure is shown in **Figure 3**. Because LCP is strong and lightweight, with a permittivity of around 3, it is ideal for antenna arrays (surface wave propagation is usually less for substrates with low permittivity). Because LCP is flexible, systems on LCP can be kept flat or conformed to a curved surface (Kingsley and Papapolymerou 2006b).

As applications go higher in frequency, there is a distinct advantage to using MEMS as the switching element. One challenge with using LCP, or any polymer, as a substrate material is the difficulty in fabricating MEMS directly on the substrate. There is a good reason why MEMS engineers choose optically flat substrate materials, such as silicon or quartz, for their devices. Fabricating on a flexible, organic substrate is not as straightforward as using a smooth, flat substrate such as silicon. Being a flexible material, it is prone to curling. This effect becomes more pronounced throughout processing due to temperature fluctuations arising from the various baking, deposition, and etching steps. Because hard contact optical lithography with a 3- to 5-μm resolution cannot be performed on a curled substrate, it is necessary to mount the sample to a flat, clean room grade material before processing. Temporary mounting can be done using a spin-on or a roll-on adhesive. Permanent mounting can be done using a thermal bonding machine. Because the substrate is also a polymer, surface roughness is an issue. The surface roughness is usually on the order of 2–10 μm. Given that the MEMS switch is generally suspended 1–3 μm above the substrate, the surface roughness can be large enough to prevent the switch from deflecting. To solve this problem, each sample is mechanically polished using alumina or oxide slurry. After polishing, the sample will have a surface roughness between 10 and 50 nm, which is smooth enough for MEMS switch operation (Kingsley *et al.* 2005).

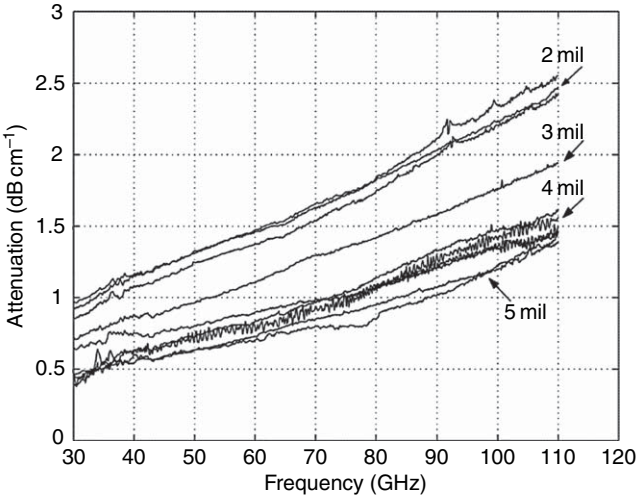


Figure 3 Measurement results of the attenuation of microstrip lines on various thicknesses of liquid crystal polymer (LCP). The loss increases as the substrate becomes thinner. (Source: Thompson D, Tantot O, Jallageas H, Ponchak G, Tentzeris M, Papapolymerou J 2004 Characterization of liquid crystal polymer (LCP) material and transmission lines on LCP substrates from 30 to 110 GHz. *IEEE Trans. Microwave Theory Tech.* **52**, 1343–52, © 2004; reproduced with permission from IEEE.)

3.10.2.2 Phased Antenna Arrays

Integrating switching elements into an antenna array is nothing new. However, using solid-state switches at the front-end of an antenna adds an element of noise. Furthermore, solid-state varactors have been used to tune antennas at the front-end, but this too adds a level of distortion that is often unacceptable in a communication system.

As the need for reduced bandwidths and beam steering increases, so does the use of planar antenna arrays. This type of antenna system allows for excellent control of the beam properties, especially when integrated with phase shifters and attenuators. All of these components require switching elements. As an array becomes larger, the number of components increases, and the required number of switches also increases. Using solid-state switches can add an enormous amount of loss to the array. This in turn will require active devices to compensate for this loss. These factors will increase the cost and create other problems such as heat dissipation and distortion. RF MEMS do not have these limitations. They do not consume power. They do not generate heat. They do not distort RF signals. They are an ideal component for planar antenna systems.

3.10.2.3 Multibit Phase Shifter

To date, there have been publications on multibit phase shifters on various materials, including silicon and GaAs (Rebeiz *et al.* 2002). In addition, various switching elements have been used, including FETs, PIN diodes, and in recent years MEMS. Given that phase shifters typically employ several switching elements, MEMS switches are a natural fit because they offer excellent isolation when in the off state and very low loss in the on state.

In a never-ending quest to lower the cost of devices, engineers often try to make devices on cheap substrates such as LCP. The first MEMS phase shifter on a flexible, organic substrate was presented at the 35th European Microwave Conference (Kingsley *et al.* 2005). That work was improved by greatly shrinking the size, lowering the loss, expanding to four bits, and integrating into an all-organic package. The improved geometry was 2.8 times smaller and 4 times shorter than a traditional switched-line phase shifter. In addition to the size reduction, the line length and the number of switches traversed compared with a traditional implementation were each reduced by a factor of 2. This resulted in half the line loss and half the switch loss by using this implementation (Kingsley and Papapolymerou 2006a). The size improvement over a traditional implementation is shown in **Figure 4**.

Works that claim small, reduced, or miniature size typically use high dielectric materials, such as silicon or GaAs, that have permittivities between 11 and 13. This is because the wavelength of a microstrip line is inversely related to the square root of the permittivity. Microstrip phase shifters on high-dielectric

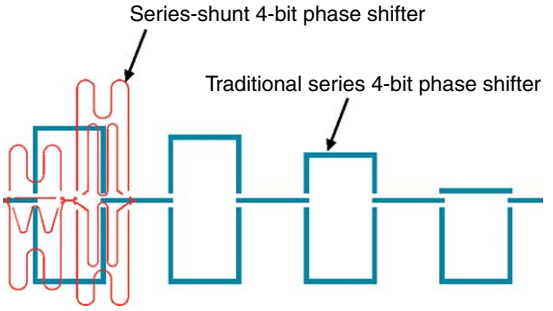


Figure 4 A series-shunt switched-line phase shifter (red) is compared to the traditional layout (blue). Not only are the overall dimensions considerably smaller but also the insertion loss is much less. Microelectromechanical systems (MEMS) serve as the switching elements, which further enhance the performance. The substrate material is liquid crystal polymer. (Source: Kingsley N, Papapolymerou J 2006a Organic ''wafer-scale'' packaged miniature four-bit RF MEMS phase shifter. *IEEE Trans. Microwave Theory Tech.* **54(3)**, © 2006; reproduced with permission from IEEE.)

materials will always be much smaller than those on low-dielectric materials. This approach attempts to level the playing field by introducing a design that is roughly the same size as a phase shifter on a high-permittivity material but is made on a substrate that is a fraction of the cost. The phase shifter fabricated on LCP is shown in **Figure 5**.

Because the phase shifter uses low-loss MEMS switches and is fabricated on a low-loss substrate, this device is capable of being very low-loss up to high frequencies. The phase shifter shown in **Figure 5** was designed for a 14-GHz center frequency. The measurement results are shown in **Figure 6**.

3.10.3 MEMS Packaging

There is an increased interest in moving toward system-on-package (SOP) RF front-end technologies. SOP offers design simplicity, lower cost, higher system function integration, better electrical performance, and various 3D packaging capabilities (Tummala 2004). Packaging is necessary for all devices, but it is particularly important for MEMS. On a hot humid day, many MEMS devices will fail in minutes if not protected. This is because at the micro level, changes in moisture and temperature can have a pronounced effect. MEMS are also sensitive to particulate matter and airflow. Dust particles can be less than a micron in diameter and invisible to the naked eye, but they can be like flying boulders to a sample of MEMS. For these reasons, MEMS must be packaged.

Electronic packaging can account for a good portion of the overall system area, mass, and cost. This is especially true for MEMS because they are ultrasensitive to environmental effects. The challenge for the package is to provide all crucial functions required by the component without limiting the performance of

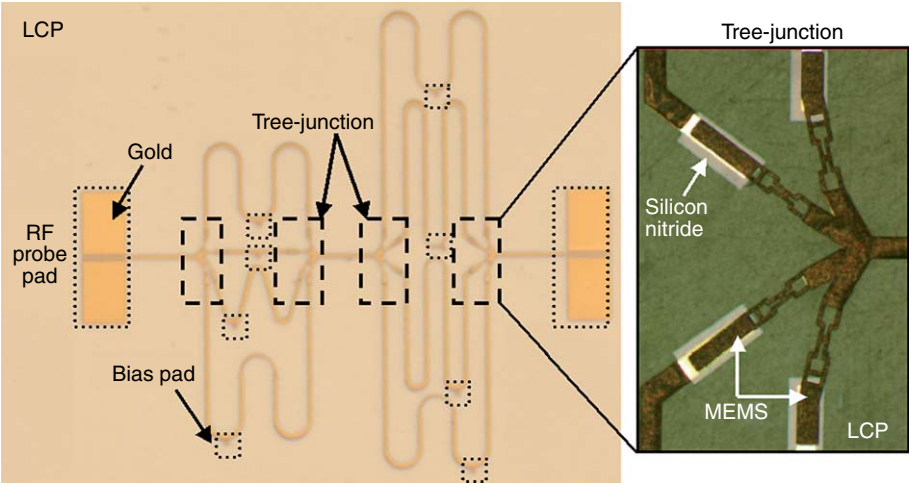


Figure 5 Four-bit microelectromechanical systems (MEMS) phase shifter fabricated on liquid crystal polymer (LCP). The switches used are a single-pole, four-throw implementation arranged into a tree-junction. Radio frequency (RF) probe pads allow for co-planar waveguide (CPW) probe measurements. DC bias pads are used to apply the DC actuation voltage. (Source: Kingsley N, Papapolymerou J 2006a Organic ''wafer-scale'' packaged miniature four-bit RF MEMS phase shifter. *IEEE Trans. Microwave Theory Tech.* **54(3)**, © 2006; reproduced with permission from IEEE.)

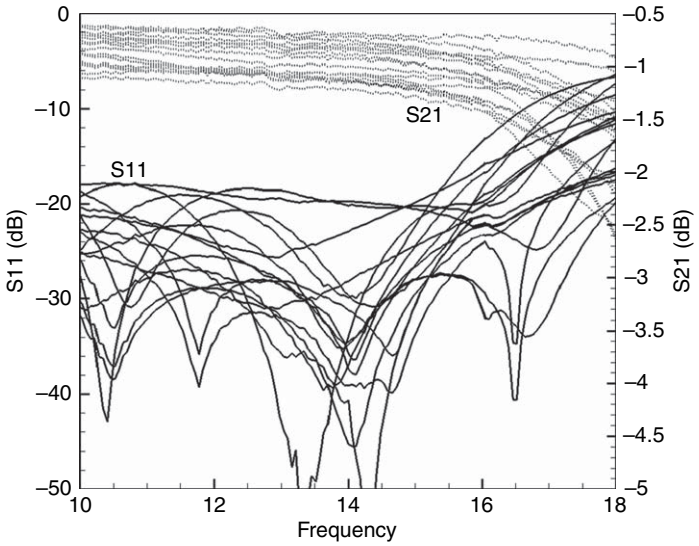


Figure 6 Measurement results for the four-bit microelectromechanical systems (MEMS) phase shifter shown in **Figure 5** (all 16 possible combinations are shown). The worst case S11 is −19.7 dB at the design frequency. The average insertion loss is 0.96 dB, which gives a per-bit loss of only 0.24 dB. (Source: Kingsley N, Papapolymerou J 2006a Organic ''wafer-scale'' packaged miniature four-bit RF MEMS phase shifter. *IEEE Trans. Microwave Theory Tech.* **54(3)**, © 2006; reproduced with permission from IEEE.)

the part. A slight crack in a seal, too much outgas from an epoxy, or too much exposure to heat can all damage a MEMS during bonding. Extreme care must be taken to isolate the device from the packaging technique, which is a difficult task because the smallest possible package is often preferred.

The semiconductor industry typically uses chip package standards. One package of a given size, shape, and number of leads can house any number of CMOS chips. The situation with MEMS is different. Some MEMS, such as pressure and chemical sensors, need to be able to sense the environment, but they also need to be protected from the elements. Other devices, such as switches, need to be completely sealed off from the environment. Some devices need to be packaged in air. Some need to be packaged under vacuum. For every application, there is a different set of package requirements. Packaging MEMS can be an expensive endeavor. Fortunately, there are several popular methods that are discussed here.

3.10.3.1 Silicon Packages

If MEMS are to be integrated directly on the wafer with CMOS or other silicon devices, then the ability to package with silicon can be useful. There are two main issues with using this method: impedance matching to the packaging layer and bonding the wafers together.

RF MEMS must be matched to the impedance of the system. This is usually not a problem because the dimensions can be tailored to match the system impedance. Therefore, cavities can be etched into a second silicon wafer and placed over the MEMS. However, when an air cavity is created around the MEMS switch, this creates a dielectric discontinuity. That is, before and after the switch, the RF signal propagates between two silicon layers. The impedance is known and constant throughout. While the signal propagates through the switch, it is sandwiched between a layer of air (permittivity of 1) and silicon (permittivity around 12). The impedance here is very different, and this will cause reflections in the system. These losses must be tolerated or a sophisticated matching network must be designed to minimize the effect.

Provided that the RF issues can be resolved, the wafers must be hermetically bonded. Bonding silicon to silicon is called fusion bonding, and it typically takes place at temperatures around 1000°C. Unfortunately, this temperature is much too high for RF MEMS switches, which can have problems with temperatures over 200°C. Instead, a different

bonding technique called eutectic bonding can be used. In this method, gold is deposited on both sides of the silicon wafers. Under pressure, gold and silicon will bond between 200°C and 300°C (eutectic bonding). This temperature is low enough for many MEMS devices, but with gold added to the processing, this will increase the cost and the complexity of production. Also, adding a great deal of metal to a wafer with RF devices can affect the devices themselves. This must be taken into account, which will also raise the design costs.

3.10.3.2 Ceramic and Metal Packages

Ceramic and metal packages are perfectly hermetic and robust, but they are costly. They are the preferred method for many high-volume applications because the ceramic cavity can be premolded and soldered or glued in place. Even though this process can be done quickly, it is not as efficient as wafer-scale techniques. To lower the production cost, chip-level packaging is becoming very popular where an entire device rather than just the MEMS is capped with ceramic (Tummala 2004).

Ceramic packaging is particularly useful for silicon devices because both materials expand and contract at nearly the same rate with temperature. Because they are mechanically durable and chemically resistant, they are useful for many applications that require a degree of robustness.

3.10.3.3 Organic Packages

Organic packaging eliminates many of the problems associated with silicon or ceramic packaging. In particular, cost can be reduced on both the design and the production side. Because polymers tend to have lower dielectric constants than semiconductors, there is less concern about impedance mismatches. Polymers are cheap and easy to work with, which will lower cost. One limitation of using organic packages is the moisture absorption. Unlike silicon, ceramic, and metal, polymers are only near-hermetic. This means that some small fraction of moisture may enter the package and cause problems for the devices. If the MEMS device is enclosed in a larger, hermetically sealed system-level package, then organic packaging is still suitable.

One popular material for organic packaging is LCP. Like in silicon, devices can be fabricated on and packaged in LCP for an all-organic system. With a permittivity near 3, dielectric discontinuities are minimal. Layers of LCP can be bonded using epoxy or heat via thermocompression bonding. LCP is ideal for thermocompression bonding because it can be manufactured to melt at either 315°C (high-melt LCP) or 290°C (low-melt LCP). Typically, the low-melt LCP is used as the bond ply and the high-melt LCP is used as the substrate and the superstrate layers. When the substrate (with MEMS), bond ply, and superstrate layers are sandwiched together and placed inside a thermocompression bonding machine at a temperature between 290°C and 315°C, the low-melt LCP bond ply will melt and adhere uniformly to the outer core layers. This creates the all-organic, near-hermetic package (Thompson *et al.* 2005). Other materials require high voltages, laser welders, or internal metal rings to adhere to the packaging layer.

LCP is an ideal material for SOP technologies. The water absorption rate for LCP is only 0.02–0.04% (Rogers 2005). Low water absorption is often necessary because absorbed water can alter the permittivity of the substrate material, which can drastically change the device performance. Like silicon packaging, it is very attractive because it allows for wafer-scale packaging. Organic materials show promise for biopackaging, which could be used inside of a human body. Because LCP is flexible, it is useful for bendable RF systems as demonstrated in **Figure 7**.

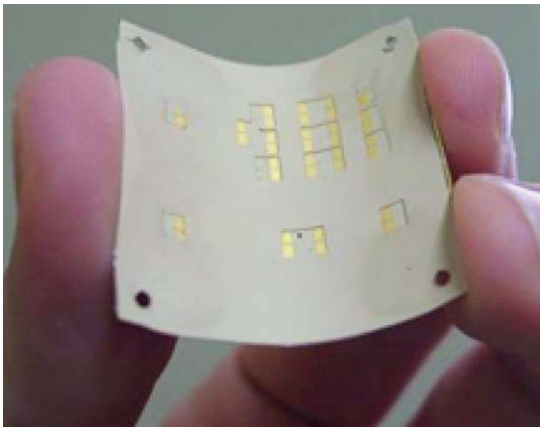


Figure 7 The phase shifter shown in **Figure 5** was packaged using liquid crystal polymer (LCP). The system can be flexed without damaging the phase shifter or switches. The presence of the packaging layer adds less than 0.045 dB of insertion loss and only 1.1 dB of return loss. This amount of loss is practically negligible.

References

Brown E, 1997 RF MEMS for digitally-controlled front-end components. *Second Annual IEEE International Conference on Innovative Systems in Silicon,* Austin, TX, USA, p. 338

Goldsmith C, Kleber J, Pillans B, Forehand D, Malczewski A, Frueh P 2001 RF MEMS: Benefits and challenges of an evolving RF switch technology. *23rd Annual Technical Gallium Arsenide Integrated Circuit (GaAs IC) Symposium Digest*, Baltimore, MD, USA, October 21–24, 2001, pp. 147–8

Kingsley N, Papapolymerou J 2006a Organic ''wafer-scale'' packaged miniature four-bit RF MEMS phase shifter. *IEEE Trans. Microwave Theory Tech.* **54(3)**, 1229–36

Kingsley N, Papapolymerou J 2006b Multibit MEMS phase shifter on flexible, organic substrate for microwave antenna systems. *15th IST Mobile and Wireless Communications Summit*, Myconos, Greece, June 4–8, 2006

Kingsley N, Wang G, Papapolymerou J 2005 14 GHz microstrip MEMS phase shifters on flexible, organic substrate. *35th European Microwave Conference*, October 2005

Koester D, Markus K, Walters M 1996 MEMS: Small machines for the microelectronics age. *Computer* **29**, 93–4

MEMtronics Corporation 2005 RF MEMS switches http://www.memtronics.com

Mercier D, Van Caekenberghe K, Rebeiz GM 2005 Miniature RF MEMS switched capacitors. *IEEE MTT-S Int. Microwave Symp. Dig.*, Long Beach, CA, USA, June 12–17, 2005

NASA Electronic Parts and Packaging Program 2003 RF MEMS switches – Description of technology http://nepp.nasa.gov/index_nasa.cfm/813

Papaioannou G, Exarchos M-N, Theonas V, Wang G, Papapolymerou J 2005 Temperature study of the dielectric polarization effects of capacitive RF MEMS switches. *IEEE Trans. Microwave Theory Tech.* **53(11)**, 3467–73

Que L, Udeshi K, Park J, Gianchandani Y B 2004 A bi-stable electro-thermal RF switch for high power applications. *17th IEEE Int. Conf. Micro Electro Mechanical Systems*, Maastricht, The Netherlands, pp. 797–800

Radant MEMS 2006 Radant MEMS switch. http://www.radantmems.com/radantmems/switchperformance.html

Rebeiz G M, Tan G-L, Hayden J S 2002 RF MEMS phase shifters: Design and applications. *IEEE Microwave Mag. 3 (2), 72–81*

Reuters 2006 Half the world will use a cell phone by 2009. http://www.mobiledia.com/news/43104.html

Rogers Corporations 2005 RO3000 series high frequency circuit materials. http://www.rogerscorporation.com/mwu/pdf/3000data.pdf

Senturia S, 2001 *Microsystem Design*. Springer Science and Business Media, New York

Thompson D, Tantot O, Jallageas H, Ponchak G, Tentzeris M, Papapolymerou J 2004 Characterization of liquid crystal polymer (LCP) material and transmission lines on LCP substrates from 30 to 110 GHz. *IEEE Trans. Microwave Theory Tech.* **52**, 1343–52

Thompson D, Kingsley N, Wang G, Papapolymerou J, Tentzeris M 2005 RF characteristics of thin-film liquid crystal polymer (LCP) packages for RF MEMS and MMIC integration. *IEEE MTT-S Int. Microwave Symp. Dig.*, Long Beach, CA, USA

Tummala R R 2004 SOP: What is it and why? A new microsystem-integration technology Paradigm – Moores law for system integration of miniaturized convergent systems of the next decade. *IEEE Trans. Adv. Packag.* **27**, 241–9

Biographies

Nickolas Kingsley received his BS (2002), MS (2004) and PhD (2007) degrees in electrical engineering from the Georgia Institute of Technology. He was a member of the Georgia Tech MiRCTECH Research Group and the Georgia Electronic Design Center (GEDC). He holds four invention disclosures for multilayer LCP technologies. He has authored or coauthored over a dozen publications in peer-reviewed journals and conferences. He is the recipient of the 2005 trainer-of-the-year award (honorary mention) from the Georgia Tech Microelectronic Research Center. He also received the 2002 President's Undergraduate Research Award from Georgia Tech and the 2001 Armada Award from Compaq Computer Corporation for engineering excellence. His research has won numerous accolades from Georgia Tech at the University, College of Engineering, and School of Electrical and Computer Engineering level. He is a student member of IEEE, IEEE APS, IEEE MTT-S, and Order of the Engineer.

Since June 2007, Dr. Kingsley has served as a principal engineer with Auriga Measurement Systems in Lowell, MA, USA.

John Papapolymerou received his BSEE degree from the National Technical University of Athens, Athens, Greece, in 1993, the MSEE and Ph.D. degrees from the University of Michigan, Ann Arbor, in 1994 and 1999, respectively. From 1999 to 2001 he was an assistant professor at the Department of Electrical and Computer Engineering at the University of Arizona, Tucson, and during the summers of 2000 and 2003 he was a visiting professor at the University of Limoges, France. From 2001 to 2005 he was an assistant professor at the School of Electrical and Computer Engineering of the Georgia Institute of Technology, where he is currently an associate professor. His research interests include the implementation of micromachining techniques and MEMS devices in microwave, millimeter-wave, and terahertz circuits and the development of both passive and active planar circuits on semiconductor (Si/SiGe, GaAs) and organic substrates (LCP, LTCC) for SOC/SOP RF front-ends.

Dr. Papapolymerou received the 2004 Army Research Office (ARO) Young Investigator Award, the 2002 National Science Foundation (NSF) CAREER award, the best paper award at the third IEEE International Conference on Microwave and Millimeter-Wave Technology (ICMMT2002), Beijing, China, and the 1997 Outstanding Graduate Student Instructional Assistant Award presented by the American Society for Engineering Education (ASEE), The University of Michigan Chapter. His student also received the best student paper award at the 2004 IEEE Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems, Atlanta, GA. He has authored or coauthored over 120 publications in peer-reviewed journals and conferences. He currently serves as the Vice-Chair for Commission D of the US National Committee of URSI and as an associate editor for *IEEE Microwave and Wireless Component Letters* and *IEEE Transactions on Antennas and Propagation*. During 2004 he was the Chair of the IEEE MTT/AP Atlanta Chapter.

3.11 Ink Jets

Joachim Heinzl, Technische Universität Munich, Garching, Germany

Glossary

AM Amplitude Modulation
CMYK Cyan Magenta Yellow Black
DGI Digital Graphic Incorporation
DOD Drop on Demand
dpi Dots per Inch
FEFCO European Federation of Corrugated Board Manufacturers, Brussels, Belgium
FGB TUM Chair of *Feingerätebau* (Precision-Engineering) TUM, since 2005 MIMED
FLOW 3D Software tool
FM Frequency Modulation
FPDs Flat Panel Displays
HP Hewlett-Packard
IJT Ink-Jet Technology
kbps Kilobit per Second
Lc, Lm Light Cyan, Light Magenta
MAGIC Multiple Array Graphic Ink-Jet Color
MEMS Microelectromechanical Systems
MIMED Chair of *Mikrotechnik und Medizingerätetechnik*, TUM
NIP Nonimpact Printing
npi Nozzles per Inch
OLED Organic Light-Emitting Diode
pl Picoliter, $1\,\text{pl} = 10^{-15}\,\text{m}^2$
ppm Pages per Minute
PZT Lead Zirconate Titanate
RFIDs Radio Frequency Identification Devices
SEM Scanning Electron Microscope
SID Society for Information Display, Los Angeles CA, USA
SIMULINK Software Tool
SOHO Small Office and Home Office
TDPS Total Drop Per Second
TUM *Technische Universität* Munich, Germany
UV Ultraviolet
VisualBasic Programming Language
wc water column

3.11.1 Introduction and Definition

Ink-jet printing has perhaps been the most successful field of microsystems during the last 10 years. This is due to the breakthrough that drop-on-demand (DOD) ink-jet has achieved in the field of desktop printing. The increase in the rates of sales and revenue is considerable, and consumables like ink cartridges and special paper have an especially large share in it. Therefore, one must distinguish between microsystems and printer hardware plus consumables with respect to their share in this huge market.

3.11.1.1 The Market

According to the Nexus Market Analysis for microelectromechanical systems (MEMS), ink-jet printheads will rise from \$1.8 billion in 2004 to \$2.5 billion by 2009 (Nexus 2005). According to a different market study (McWilliams 2006), in 2004 microfluidic MEMS had a share of over 44% of the global MEMS market and MEMS ink-jet heads totaled \$1.8 billion or almost 40% of the global MEMS market in 2004. These figures obviously take into account only the MEMS themselves. If the complete printer hardware and consumables are taken into consideration, different figures are obtained.

In October 2002, Marc Boer, a consulting partner at IT Strategies, predicted an increase of worldwide revenues for ink-jet printers and inks from \$28.5 billion in 2001 to almost \$41 billion by 2006 (Boer 2002). In 2006 he predicted a growth from \$42.1 billion in 2004 to \$58 billion by 2009. The estimate for 2006 had been reached in 2004. This shows the development for wide- and narrow-format ink-jet hardware, media, and ink. For the ink, he predicted an increase from \$13.4 billion in 2002 to more than \$21.5 billion by 2006. In 2006 he predicted an increase from \$20 billion in 2004 to almost \$33 billion by 2009, and for media from \$7 billion to \$11 billion in the same period (Boer 2006). Not included in these estimates is the emerging industrial application of ink-jet technology, "perhaps the biggest opportunity" as Boer says.

3.11.1.2 Drop-on-Demand Systems

There are various systems of ink-jet printing. Only DOD systems, i.e., systems that eject single drops under electrical control, will be considered here. DOD systems use the vacuum method. This means that the ink in the printhead is under negative pressure. The printhead consists of a number of nozzles. Each of them communicates with a cavity filled with ink. Each of them is triggered by an actuator generating shock waves. The cavities are refilled from an ink cartridge by capillarity. Viscosity and surface tension of the ink are tuned in a way that the printheads can do without tight valves. Surface tension and vacuum at

the nozzle take care that the ink does not escape from the nozzle. Drops are only ejected when a shock wave coming from the cavity reaches the nozzle. Due to negative pressure, the surface tension forms a negative meniscus in the nozzle orifice.

This negative pressure can be created in several ways, either by the ink cartridge that is situated some centimeters below the nozzles (Heinzl *et al.* 1979), or by capillarity that keeps the ink (Hara 1985), or by a sponge that absorbs the ink (Baker *et al.* 1988) or by a valve (Nielson and Bolmgren, 1986) that holds the ink back in the cartridge until the vacuum reaches a threshold. The advantage of negative pressure is that many nozzles can be arranged side by side on one and the same nozzle plate without the ink on the nozzle plate disturbing the printing. As soon as the ink on the nozzle plate comes into fluid contact with the ink in a nozzle, it is sucked back by the negative pressure. Only this little trick makes it possible to arrange many nozzles close together.

These systems have made typewriters and needle printers obsolete and are just replacing classical paper prints in photography. Increasing reliability will allow more and more industrial applications, leading to quantum leaps in production strategies.

Apart from the above DOD systems, there are systems with tight-closing valves. They are used for large-scale applications like printing on tarpaulins, covers, or carpets. Also, there are continuous jets using high or low pressure. They produce a print using a continuous stream of droplets, which are selected and deflected (Le 1998). These systems will not be considered here.

Figure 1 shows the difference between nozzles used in high pressure, low pressure, and vacuum methods. High-pressure methods make the jet break into single droplets when leaving the nozzle. This is done either by the surface tension itself, with the droplets offering less surface than the jet (Rayleigh 1878), or by pressure waves of high frequency encouraging and synchronizing the breaking up of the jet. Pressures up to 3×10^6 Pa (300 m wc) create droplets of a volume of 2–10 pl with frequencies up to 1 MHz and reaching a speed of up to 40 ms^{-1} (Heinzl and Hertz 1985, p. 101ff).

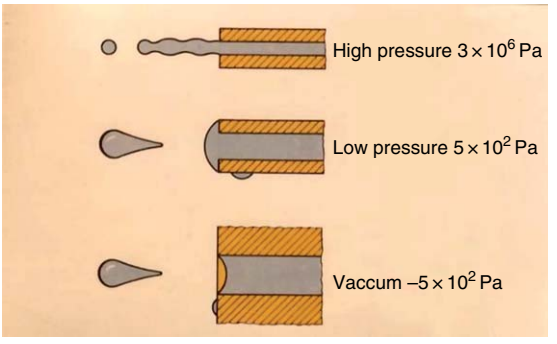


Figure 1 Nozzles at high pressure, low pressure, and vacuum method. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

Low-pressure methods use the positive meniscus of the ink in the nozzle. An electrical field draws out and deflects the ink. Here either the entire nozzle or at least its outer surface must be hydrophobic and remain so. Typical data are 5×10^2 Pa (5 cm wc) in the nozzle, droplet volumes of 20 pl, and frequencies of up to 5 kHz (Ascoli *et al.* 1975).

Vacuum methods use a nozzle that has at least a hydrophilic inner surface. So the negative meniscus remains in the nozzle in spite of the vacuum. Liquids on the surface of the nozzle plate are sucked back into the nozzles by the vacuum. Typical data are a vacuum of -3 to -5×10^2 Pa (-3 to -5 cm wc), a droplet size ranging from 1 to 200 pl, flying speeds from 4 to 12 m s^{-1}, and frequencies from 4 to 20 kHz. Usable frequencies of droplets up to 100 kHz are predicted for DOD systems.

3.11.1.3 Drops and Media

DOD systems are distinguished by the great variety of inks and recording media they allow. There are water-based inks, inks with solvents, ultraviolet (UV)-curable inks, wax inks, and solid inks. UV-curable inks are fixed after application by UV light. Wax and solid inks are melted before jetting. Soluble substances or pigments down to nanoparticles serve as colorants.

Meanwhile, medical drugs, glues, oils, organic substances, plastics, and even metals such as soft solder are jetted. This corresponds to a great variety of recording media from plain paper to coated paper, transparencies, films, corrugated cardboard, and thick card stock. There are applications on glass, ceramics, wood, metal, plastics, and different powders. Depending on the combination of ink and medium, there are differences in the ways by which the droplets hit the medium. **Figure 2** offers a schematic overview.

When water-based inks and plain paper are used, the droplets are deformed when they hit the paper. A meniscus with a flat wetting angle is formed. It defines the area in which the ink is absorbed by the paper, and it appears as the dot. A long tail and satellites of the droplet hit the meniscus and unite. If the printing speed approaches that of the flying droplets, tails and satellites can be distinguished on

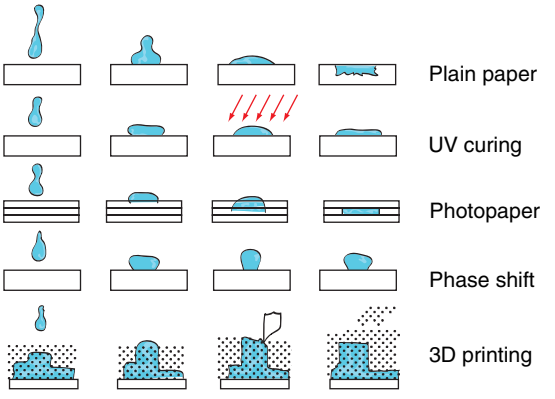


Figure 2 Drops and media. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

the paper. If a neighboring droplet hits the meniscus before the latter is absorbed, the two droplets coagulate, the second drop enlarging the first and shifting it slightly aside. The drops are absorbed in fractions of milliseconds. Drying takes considerably longer.

UV inks can be directly applied to nonabsorbing surfaces. Their viscosity is higher. Therefore it is heated for jetting. When they hit the medium, they cool down quickly and are fixed by UV radiation.

For printing on films or transparencies, at least a gelatin coat is necessary, as long as water-based or solvent inks are used. Without such a coat the ink would take hours to dry, and a major part of the droplets would coagulate. Special photo papers consist of several layers, one of them holds the color.

It takes a temperature of more than 300°C to shoot soft solder on copper or gold conductors. At this temperature, viscosity and surface tension of the soft solder closely resemble those of water ink at room temperature. The drops freeze on the conductor without wetting it, and thus become uniform bumps.

Three-dimensional printing requires the spreading of single layers of granular material. Either the granular material is coated with a hardener, or the hardener is applied to each single layer. A binder is jetted, thus creating layer by layer the structure of the 3D body. The loose granular material serves as support. Normally, the newly created surface is smoothed before the next layer is spread.

3.11.2 Products on the Soho Market

Available ink-jet products range from small printing calculators and a variety of desktop printers, photo printers to all-in-one devices, and large-format printers. The companies Hewlett-Packard (HP), Lexmark, Canon, and Epson dominate this market. The printheads of these devices contain either small heating elements creating vapor bubbles or piezoelements.

3.11.2.1 Desktop Printers

When you buy a desktop printer you do not usually know which technology is hidden in it. The packaging only says ink-jet. It does not give any indication whether a thermal printer, a bubble jet, or a piezojet is inside. Obviously, the term ink-jet is the decisive advertising argument. Canon uses bubble jet technology. HP and Lexmark call the same technology thermal ink-jet.

Only Epson offers printers in piezotechnology. They look almost alike. **Figure 3** shows typical printers from HP, Lexmark, Canon, and Epson: HP deskjet 990c (**Figure 3(a)**), Lexmark Z735 (**Figure 3(b)**), Canon Pixma iP 1600 (**Figure 3(c)**), and Epson Stylus C64 (**Figure 3(d)**). On almost every packaging, you find data such as 4800 dpi (dots per inch), up to 22 ppm (pages per minute), and the number of colors and cartridges.

You can find such printers for prices below €30. Thus, the complete device is hardly more expensive than a new set of cartridges. In order to avoid the customer buying a new printer instead of new cartridges, some companies provide their printers with only partly filled cartridges. Without the sale of replacement cartridges, the sale of the printers would not be economical. The ink, which consists mainly of water, can be sold in the cartridges for a price higher than that of expensive perfumes. That is why a lot of companies try to have a share in the sale of cartridges, even if they themselves do not offer printers. The companies Epson and Canon are especially affected by these substitutes. They do hold a number of patents to protect the technology of their cartridges. These patents are the object of numerous nullity and infringement suits. Many of these patents have already expired.

The cartridges of HP and Lexmark are less affected by substitutes because apart from the ink they contain the complete printhead (Buck *et al.* 1985). Only for heavy-duty printers ink cartridges without printhead are provided by HP (**Figure 4**).

Figure 3 (a) Hewlett-Packard (HP) deskjet 990c, (b) Lexmark Z735, (c) Canon Pixma iP 1600, and (d) Epson Stylus C64. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

Figure 4 Hewlett-Packard (HP) Business Inkjet 1200. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

Figure 5 The Epson Picture mate 500. (Source: Epson Deutschland GmbH or SEIKO EPSON Corporation.)

3.11.2.2 Photo Printers

Especially for the direct printing of photos, a special printer type has been developed. Small, handy printers for the formats 10 cm × 15 cm and 10 cm × 30 cm, partly also for 13 cm × 18 cm are for sale. Epson's picture make 500 offers a resolution of 5760 × 1440 dpi. Figure 5 shows this printer. A 10 × 15 photo print takes 77 s. HP printers such as the Photo Smart 385 or P 475 have a resolution of 4800 × 1200 dpi. The higher figure describes the resolution along the raster lines, whereas the lower

figure is the resolution between the lines. The special photo papers offered for these printers contain a special layering providing a cover layer, a layer that absorbs the colors, a layer for the solvents, and a support layer. According to what is printed on some packages, the specially pigmented inks will not fade in the next hundred years.

In this field, ink-jet printing is competing with thermal transfer. The latter uses color foils in addition to photo paper and provides waterproof prints. It must be said that traditional photo prints on paper are not waterproof either.

3.11.2.3 All-in-One

By combining a scanner and a printer, numerous functions can all be realized in the same device: printing black-and-white, color printing, photo printing, scanning, copying and color copying, and faxing. In this field HP is the market leader. The devices differ in the number of interfaces, especially for memory cards, the number of paper cartridges, the size of the stack for originals, the printing speed, the number of ink cartridges, the scanning resolution, and the transmission speed for faxes. The Photosmart 3110 of HP, with six ink cartridges, 32 ppm, 4800×4800 dpi scanning resolution and 33.6 kbps. Such devices are available for less than €300 in the marketplace. **Figure 6** shows the Brother MFC-820 CW.

3.11.2.4 Large Format Printers

Meanwhile, a great variety of devices for the printing of posters is available. Rolls of special paper are printed on, their breadth ranging from A1 to 60″. The performance of HP printers is $52.8\,m^2\,h^{-1}$ on plain paper and $13.7\,m^2\,h^{-1}$ on special paper, using four colors with 1200×600 dpi. **Figure 7** shows the HP large format printer Designjet. Canon's best device is the PROGRAF 9000 using 30 720 nozzles, a 12-color pigment inking system, 60″ paper breadth, 4-pl drop size, a resolution of 2400×1200 dpi with an enlarged color gamut and a strongly reduced granularity. Most of the devices are equipped with a cutter and a stationary ink refill station. In this field, ink-jet printing is unequaled. Laser printing is not suitable for large formats.

3.11.3 Actuators and design of Printheads

In order to eject a droplet from a nozzle and to produce a dot, you need a shock wave containing a certain amount of energy. This energy is necessary to accelerate the fluid against the vacuum and to overcome the viscosity of the fluid in the nozzle, in the ink cavity and in the feeder, and to build a new fluid surface. A drop of water of 30-pl size that flies at $8\,m\,s^{-1}$ has 9.6×10^{-10} J kinetic energy and 3.1×10^{-10} J surface energy. The power loss caused by the viscosity is far less. Depending on the type of actuator and the design of cavity and nozzle, you have to apply 10^{-8}–10^{-6} J of energy to the actuator in order to eject a drop. In wire matrix printing,

Figure 6 Brother MFC-820 CW. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

Figure 7 Hewlett-Packard (HP) large-format printer DesignJet. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

the mechanical energy is transmitted from an electromagnetic transducer to the tip of a print needle by the motion of the needle. The electrical energy required to print a dot is about 10^{-2} J. Approximately, one-tenth of this energy is dissipated at the tip of the needle and on striking the ribbon. The needle has to be accelerated and has to squeeze the much thicker ink out of the ribbon and to knead it into the paper. The remainder is transformed into heat together with a tiny but especially annoying part that is transformed into noise. In comparison to that, ink-jet printing is nearly soundless. What can be heard in the office is only the movement of the printing carriage and the paper.

3.11.3.1 Efficiency and Power Density

Various actuators are suitable for miniaturization and thereby for the use in a high resolving printhead. There have been experiments with electrostatic, electrodynamic, electromagnetic, piezoelectric, and thermoelectric actuators. There are two kinds of thermoelectric actuators: one uses a phase-change medium and produces bubbles, the other uses a solid-state body and produces thermal dilatation bending like a bimetal.

Obviously, the thermoelectric actuators with bubbles are best suited for miniaturization. The ink, however, must be resistant to boiling, which distinctly restricts the choice of inks. Piezoelectric actuators need more space, but they are more universally applicable. Electrostatic actuators in air have too low a power density. Electrodynamic actuators can only be produced reasonably when there are few windings. They also need strong currents and very low voltages. For this purpose, there is no applicable circuit technique (Berdelle-Hilge 1990). Electromagnetic actuators have the same problem. They have a low efficiency of not more than a few percent. In comparison, piezoceramic offers efficiencies up to 25% and a high power density. An even higher power density can be found in thermoelectric transducers when they produce bubbles.

Figure 8 shows an experiment for measuring the efficiency of piezoelectric actuators. A piezoblade or a piezostack is fixed on one side. At its free end, there is a sphere that hangs on a thread. While the piezoceramic is charged, current and voltage are measured, and thus the energy input is calculated. The sphere is flung off during charging. The kinetic energy that is imparted to the sphere is determined by means of the deflection, i.e., the ascent height of the swing. The mass of the sphere can be varied until the optimal ratio between mechanical work and electric energy is reached. In this way, by using a large signal, an efficiency of 25% was observed (Heider 1975).

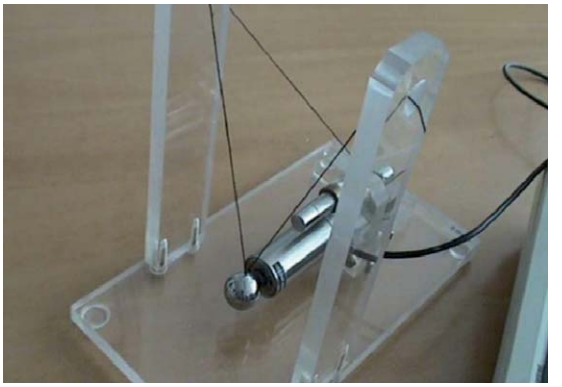


Figure 8 Efficiency experiment with piezoactuators. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

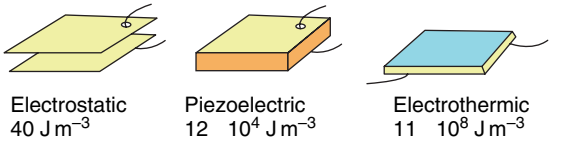


Figure 9 Comparison of power density. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

Figure 9 shows a comparison of power density among electrostatic, piezoelectric, and thermoelectric actuators. The forces applied to the plates of a parallel plate capacitor are calculated up to the insulating strength of air. Where piezoceramic is concerned, power density is calculated from the change of length and Young's modulus. For the vapor bubbles, the increase of volume during vaporization and the heat of evaporation are taken into account. While expanding to atmospheric pressure, the vapor film increases its volume by the factor 1700. It is assumed that the evaporating fluid film and the heating element have roughly the same thickness. The power densities of electrostatic, piezoelectric, and thermoelectric transducers with evaporation differ from one another by a factor of 3^{10} or 4^{10}.

3.11.3.2 Thermoelectric Actuators

In 1979, Endo and Hara of Canon (Endo *et al.* 1977) invented the DOD method with droplets ejected

from a nozzle by growth and collapse of a vapor bubble generated by a small heater element located near the nozzle. They called it bubble jet technology. At the same time HP developed a similar technology (Vaught *et al.* 1984). They called it thermal ink-jet.

Thin resistor layers serve as heater elements. Layer t+hickness typically is about 100 nm. The heater of the first printheads of HP measured about 100 µm × 100 µm. Today's printheads need only 1% of this area. Tantalum aluminum and hafnium boride are typical materials chosen for heaters. Water- or solvent-based color solutions or pigmented versions are used as inks. The heaters reach temperatures greater than 300°C (Runge 1993). While the ink is heated, no residues should remain on the heater, as they might burn or coagulate. In some products, the heater is covered with a protective layer of quartz. The surface is highly stressed, not only thermally, but also mechanically. When the bubbles collapse, cavitation occurs, which cleans the heaters, but may also damage them.

Three types of arrangements can be distinguished: the edgeshooter, the sideshooter, and the backshooter. The edgeshooter ejects droplets tangent to the heater surface. The sideshooter contains a nozzle plate arranged in parallel to the heaters and the droplets are ejected normal to the heater surface. The backshooter contains the heating elements on the nozzle plate. The drops are ejected normal to the heater surface, but the vapor bubbles emerge on the opposite side of the nozzle plate. **Figure 10** shows the three types.

Once again, the advantage of the thermoelectric heaters is that they are small and can be arranged in close proximity to the nozzles. Their drawback is that there is a great demand on the ink and the heater. Sometimes sideshooters are also called roofshooters. Edgeshooters and sideshooters are also used for piezoprintheads if the heaters are replaced by piezomembranes.

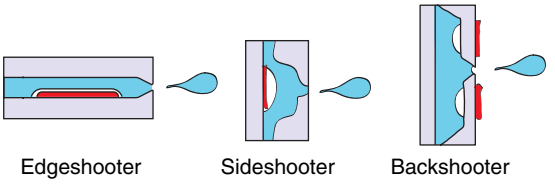


Figure 10 Sideshooter, edgeshooter, and backshooter. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

3.11.3.3 Printheads with Thermoelectric Actuation

The process of evaporation could be observed particularly well with the first printheads. **Figure 11** shows the top view of the HP Thinkjet heater. It is a sideshooter with heating areas of 100 µm × 100 µm. Its nozzle plate has been removed. The heating elements enhance their electronic conductivity a little when the temperature rises. The coefficient is enough to make the heater serve as a sensor at the same time. After starting the pulse, the curve of the current rises slowly with increasing temperature. After 5 µs the nuclei of the bubbles multiply and after 6 µs they form a hermetic film above the heater. The vapor barrier now interrupts the heat transfer to the ink, and accordingly, the temperature of the heater increases rapidly. Therefore, the current curve bends sharply and continues more steeply (Pöppel 1990). After about 20 µs, the bubble collapses and cavitation jets starting from the corners can be detected. After about 200 µs, the initial state is reached again. Frequently, single nuclei remain, which disappear very slowly or not at all.

Figure 12 shows the design of an early Canon bubble jet printhead. Conductive patterns and heaters are applied to the silicon substrate in thin-film deposition technique. The channels and nozzles are made of photoresist. The channels are closed on top by a glass lid fixed with adhesive film. The preparing of the front edge of the nozzles was challenging. Silicon, glass, and photoresist had to be cut, ground, and lapped. Nevertheless, this head was built in great numbers.

Figure 13 shows the design of a backshooter, which was developed to the stage of a prototype. On the substrate, two series parallel converters and the end transistors were integrated on top of the nozzle plate. Therefore, only six contacts were necessary for 128 nozzles. The heaters were arranged on both sides of the nozzles. The cavity was structured by anisotropic etching. The nozzles were made by dry etching (Heinzl *et al.* 1990, Wehl 1995, Zollner 1995).

Figure 14 shows a HP ink cartridge with its contacts and the nozzle plate. It is a consumable. The number of contacts is reduced by a matrix of diodes. **Figure 15** shows a section of the nozzle plate with two rows of nozzles. The long slits in between serve to separate inks of different colors. Behind the slits, there are felt pads that absorb the mobile ink. **Figure 16** shows some nozzles more

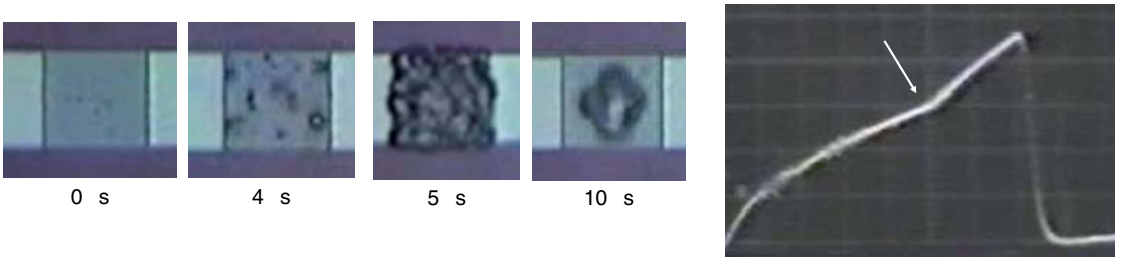


Figure 11 Bubbles and current history. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

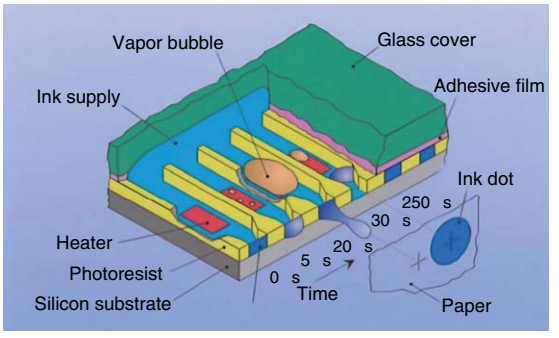


Figure 12 Cannon's early bubble jets. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

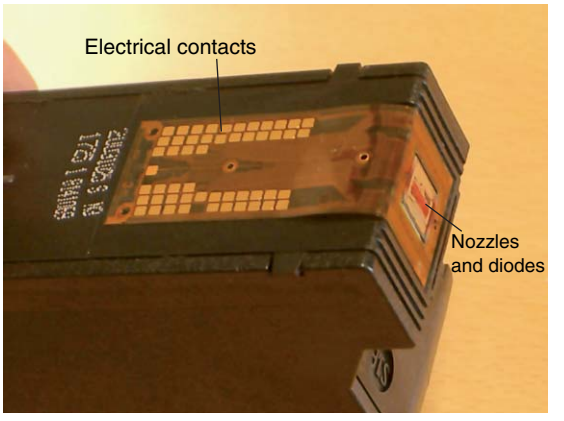


Figure 14 Hewlett-Packard (HP) ink cartridge. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

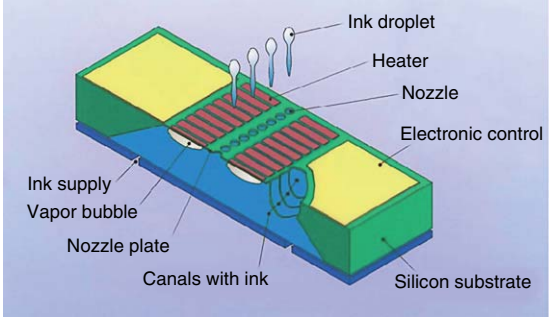


Figure 13 Build-up of a backshooter. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

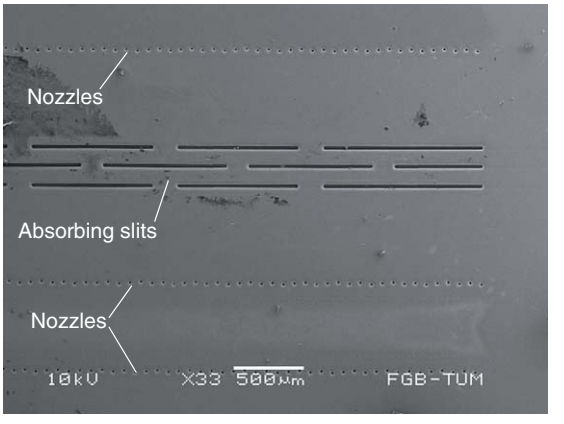


Figure 15 Section of a nozzle plate. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

closely. The single nozzles are slightly staggered so that the pulses can be given at different moments and nevertheless form a vertical line. The nozzle plate is electroplated. One can recognize the starting electrode, upon which the nickel has grown uniformly. The nozzles have a sharp outer edge; toward the cavity they expand donut-shaped.

Figure 17 shows the structure beneath the nozzle plate. Two rows of cavities and heaters can be seen. Behind a structure that resembles a rake, there are two tributaries to the proper ink cavity

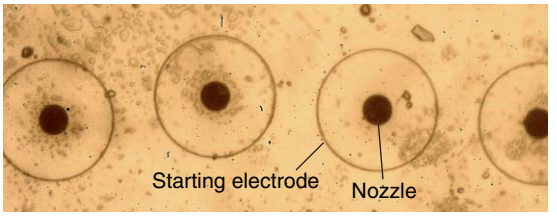


Figure 16 The nozzles. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

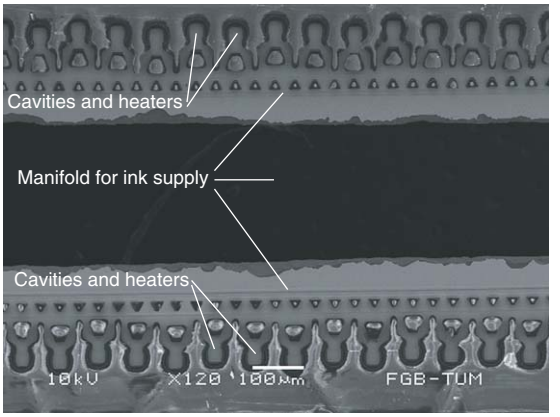


Figure 17 Heating elements and channel structure. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

with the heater. Between the two rows, one can recognize the through hole where the ink comes in.

Figure 18 shows a schematic of the new Canon PIXMA MP800 nozzle element. It is also a sideshooter. The 6000 nozzles are arranged in an area of 20 mm × 16 mm. The bubble cleans the nozzle. **Figure 19** shows the corresponding channel structure designed for the production of 1- and 5-pl drops. Canon has developed a new manufacturing technology for it. First the heater and circuit pattern are formed on a silicon wafer, after that two different photoresists are applied, exposed to light, and developed. While the first photoresist is fully leached out, the complete structure of cavities and nozzles appears on the wafer. The distance between neighboring nozzles is 85 µm.

Sideshooters are predominant in the design of printheads with bubbles because the increasing number of nozzles can be arranged in several rows close together. As we have seen, the production technology can vary. There are electroplated nozzle plates glued

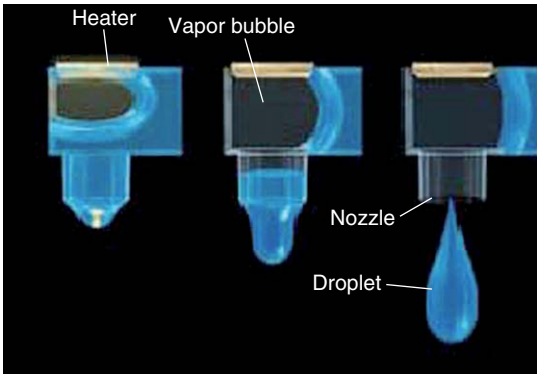


Figure 18 Canon nozzle element. (Source: Canon Deutschland, Europark Fichtenhain A10.)

to a silicon structure (HP). There are nozzle plates built with polyimide foil in which the nozzles and the cavities are formed by laser ablation in various layers. They are mounted on the structures with the heaters (Kahlert 2001). Finally, there are nozzles and cavities of photoresist, which are applied on the wafer in a batch process (Canon).

Literature describes printheads that make use of the bimetal effect. With adequate miniaturization and by using ink as coolant, the electrically actuated bimetal can react with high frequencies. Short pulses can heat it on one side and thus deflect it.

3.11.3.4 Piezoelectric Actuators

A large number of materials show piezoelectric effects. This means that mechanical stress evokes an electric charge on the surface and *vice versa* an electric charge on the surface creates a mechanical stress inside the material (Curie and Curie 1880). Seignette salt, quartz, ice, and turmaline are well known. Apart from that, there are ceramics that show excellent piezoelectric characteristics after sintering and polarizing (Ikeda 1990).

3.11.3.5 Piezoceramics

PZT (lead zirconium titanate) has a high efficiency as discussed in Section 3.11.2.1. In the following, piezoelectricity, polarization, and Curie temperature will be discussed taking PZT as an example. As long as its temperature is higher than the Curie temperature, its elementary cell consists of eight lead atoms that are situated at the corners of a cube. In the center of each of the six faces of the cube, there is an oxygen atom,

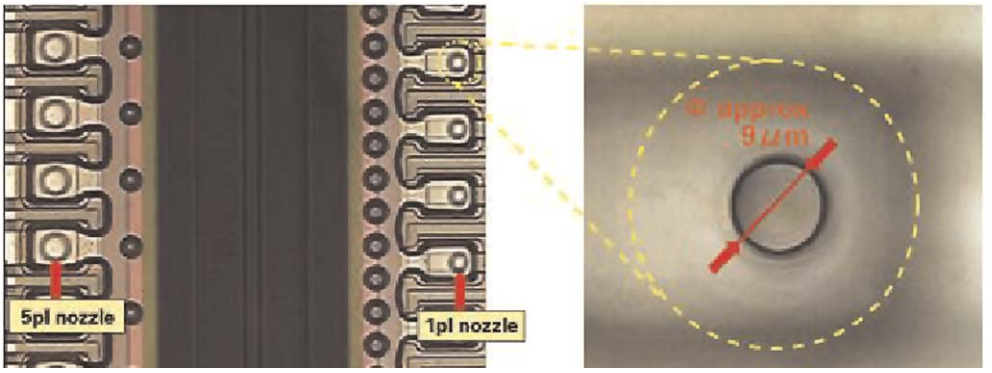


Figure 19 Canon's channel structure and heating elements. (Source: Canon Deutschland, Europark Fichtenhain A10.)

and in the center of the cube there is a titanium or a zirconium atom. In **Figure 20** this elementary cell is shown. Only one lead atom, one oxygen atom, and one titanium or zirconium atom belong to the elementary cell. The other atoms are part of the neighboring cells.

As the ceramic slowly cools down, the atoms move closer together. On reaching the Curie temperature, the central atom instantaneously fits in between the oxygen atoms. The Curie temperature depends on the mixing proportions. For PZT it is between 240°C and 370°C. The length of the cell edge is about 0.4 nm at the Curie temperature. Above the Curie temperature the elementary cell is only electrostrictive. In an electric field, the lead atoms and the central atom move relative to the oxygen atoms. The cube shape is not deformed. Below the Curie temperature, the central atom has to shift one of the oxygen atoms in order to gain space. The other oxygen atoms move closer together, and the cube is deformed into a cubic. An electric field in the direction of the shift intensifies this tendency, making the cubic even longer.

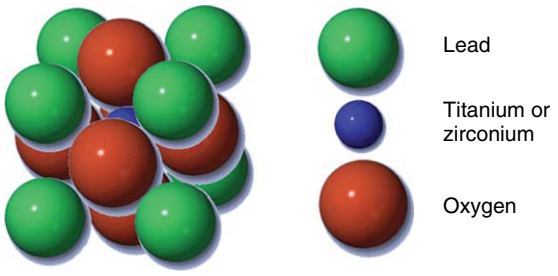


Figure 20 PZT (lead zirconate titanate) elementary cell. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

As ceramics are not monocrystallic, it consists of numerous domains. Within a domain, the elementary cells show the same orientation, but the central atoms can shift in different directions while falling below the Curie temperature. One of the six oxygen atoms must give way, so there are six different directions that can be chosen. When an electric field is applied, the distortions will compensate each other on average, and therefore there is no piezoelectric effect. Only when the distortions are oriented roughly in the same direction, the piezoelectric properties appear. This is achieved by applying a strong electric field at temperatures near the Curie temperature. Then the oxygen atom that lies closest to the direction of the field has to give way while cooling down. This process is called polarization. Electrodes are fixed to the ceramic in order to apply the voltage. The positive direction of the electric field is called direction of polarization. To avoid flashovers, the ceramic can be dipped into a transformer or silicon oil. The polarization moves the central atom to that of the six directions, which is closest to the direction of polarization. Deviations up to 54.7° can occur. If we could succeed in growing a monocrystal and in getting polarization into line with it, the effect would be far stronger. There is as well a difference between small signal and large signal reaction of the ceramic (Reiländer 2003, Wolters 2004).

To describe the directions of PZT, it has become customary to use a rectangular coordinate system with axes 1 and 2 parallel to the electrodes, and axis 3 in the direction of polarization.

3.11.3.6 Longitudinal, Diagonal, and Shear Effects

If voltage is applied in the direction of polarization, the ceramic is lengthened in this direction and thins out crosswise. These effects always occur simultaneously. They are called longitudinal and diagonal effects. If the voltage is applied in the normal direction, the ceramic will shear. A cubic is transformed into a parallelepiped. This effect is called shear effect. To achieve it, though, the electrodes for polarization must be replaced and new electrodes normal to the former electrodes must be put in place. **Figure 21** shows the middle plane of the elementary cell before (**Figure 21(a)**) and after polarization (**Figure 21(b)**) as well as the deformation in the direction of polarization (**Figure 21(c)**) and normal to it (**Figure 21(d)**), both excessively magnified. In reality, the effects are very small.

Flashover in air limits the longitudinal effect to 0.2% in the direction of polarization and the diagonal effect to 0.1% (**Figure 21(f)**) compared to the unloaded dimension (**Figure 21(e)**). In the opposite direction, only a third of these effects is usable. If it exceeds, there is a risk for partial depolarization. Shear effect can be used up to 0.2% in both directions (**Figure 21(g)**).

Piezoceramic can be used in many different ways as an actuator, as shown in **Figure 22**. The red arrows indicate the direction of motion. The longitudinal or the diagonal or the shear effects can be applied. The piezotube (**Figure 22(a)**) with its electrodes on the inner and outer circumferential surface uses both the longitudinal and the diagonal effects (Heinzl and Kattner 1975). An electric field applied in the direction of polarization can make the wall thicker and the circumference as well as the length shorter. The

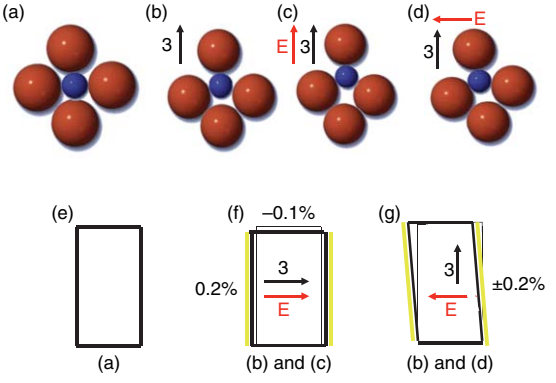


Figure 21 Middle plane of the elementary cell. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

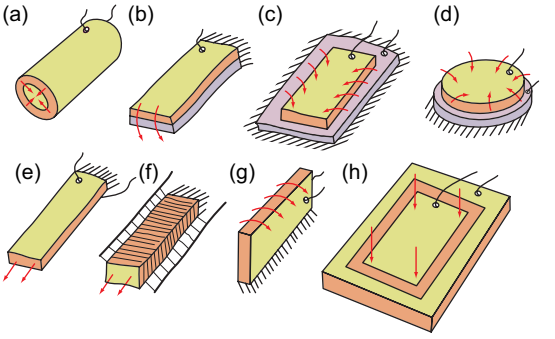


Figure 22 Piezoelements and their deformation. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

dimension can be determined in such a way that the outer diameter remains constant while the inner diameter and the length change (Heinzl 1977).

A structure of two layers of which at least one is a piezoceramic is called a bimorph. If an electric field is applied using the diagonal effect, the ceramic layer and the bimorph bend (Kyser 1981). A small strip fixed on one end acts as a cantilever beam (**Figure 22(b)**). When fixed on all sides, it acts as a membrane. This membrane can form a square, a rectangle (**Figure 22(c)**) or a circle (**Figure 22(d)**), or even a diamond. If there is a distinct distance between the ceramic and the fixation line, the uncovered part of the second layer acts as a solid-state hinge.

If you use it like a ram, either the diagonal or the longitudinal effect can be applied. To have a great dilatation and at same time low voltage, a long, narrow, and very thin cubic is selected. The large lateral surfaces can be used as electrodes and the diagonal effect can be made use of (**Figure 22(e)**). You can also make use of the greater longitudinal effect by stacking many layers upon each other, putting electrodes in between (**Figure 22(f)**). The electrodes can be contacted in such a way that the direction of the field changes from layer to layer. Such multilayer stacks are available with layers as thin as 20 µm. The art in producing these stacks consists in making the electrodes withstand sintering. This can only be achieved by using palladium and silver.

Because the electrodes for operating and polarizing are the same, polarization can be renewed if the longitudinal or diagonal effect is used. In contrast, it is not possible to renew polarization if the shear effect is used. Application of very strong voltages to the operating electrodes is likely to make the shear effect

disappear. To obtain the shear effect, there are two typical arrangements of electrodes. The first is to put electrodes on two opposite sides of a polarized ceramic wall normal to the direction of polarization. In this case, the field lines run across the ceramic, normal to the direction of polarization, thus producing the shear effect (**Figure 22(g)**). The second especially elegant arrangement is to use a thin layer of polarized ceramic, to remove only one of the electrodes completely, and to partition the other electrode. In the gap between two partitions, a great part of the field lines run normal to polarization if different potentials are applied to the different parts of the electrode (Fischbeck 1986). Then the ceramic shears all along the gap (**Figure 22(h)**). These arrangements can be operated as monomorphs.

3.11.3.7 Shock Waves and Drop Generation

As soon as a voltage is applied to the electrodes, the mechanical tension in the ceramic rises without delay. This tension tries to deform the ceramic. The inertial forces in the ceramic must be overcome. This leads to a finite speed of structure-borne sound. Besides, elasticity and inertial forces of the mechanical structures attached to the ceramic must also be overcome. And the ink must be compressed or moved. This can happen in different ways, depending on the mechanical impedances of the materials and structures involved.

If the ceramic is oversized, the ink must be considered as compressible. This was the case with the piezotubes (Wehl 1984). For the greater part, the propagation of pressure is done acoustically. Inside the tube, the pressure is imparted to the ink, and it runs through the cavity at the speed causing a sound. Every change of diameter and/or elasticity of the cavity results in a splitting: One part runs on, the other part is reflected. If the cavity opens to the manifold, the pressure wave is reflected with reversed sign. The running waves are superimposed by the reflections.

As long as the ink in the nozzle is in static condition, it reflects like a closed end. When the ink begins to move in the narrow nozzle, the impedance changes together with the increasing speed. At the beginning of a drop ejection, the cavity acts like a covered organ pipe and then changes into an open one. Drops can be formed by harmonizing the pulses with the waves running to and fro and the pressure history at the nozzle (Heinzl and Hertz 1985, pp. 113–31).

Less energy for drop ejection is needed, if the mechanical impedance of the ceramic is well adapted to the impedance of the filled cavity. The shorter the distance between the actuator and the nozzle and the fewer the changes in cross section, the less is the influence of the acoustics. The actuator then simply forces away the liquid. The actuator can then be regarded as a piston in a leaky cylinder. A change of the electrical field on the actuator results in a new balance of the piston. This balance, however, is not achieved at once. Not before the ink is forced through the nozzle and back through the manifold, can the actuator follow the balance.

Under these conditions, the mass of the piston can be neglected, and the ink can be seen as incompressible. A restrictor is placed between the manifold and the ink cavity with the actuator. It is of the dimension of the nozzle. It has to constrain the flow back to the manifold during the ejection of a drop and it has to allow the refill of the cavity in time. The restrictor can also be a sintered metal forming the cavity between actuator and nozzle (Karlinski 1999). The movement is monitored almost exclusively by forces of inertia and the viscosity of the ink. A small contribution is added by the surface energy needed to build the drop and to remove it.

The drops are formed almost exclusively by the speed that is imparted to the ink at the exit of the nozzle. The ink can only be accelerated or slowed down as long as it is inside the nozzle. As soon as it leaves the nozzle, it can only be shaped further by surface tension, which is counteracted by inertia and viscosity. If the ink that leaves the nozzle is slowly accelerated, the first and slow parts of the drop are overtaken by the faster parts. The slow parts do not move forward and so form a long tail. When the ejection speed finally slows down, the end of the drop becomes longer and thinner and is detached if that results in smaller surfaces.

Better drops can be formed if the ink is first pulled back deep into the nozzle. Inside the nozzle, the ink can be accelerated so that even the first parts of the drop reach the required flying speed. The greater part of the drop should be ejected at that constant speed. In order to separate the drop quickly, the speed of the ink in the nozzle should be reversed as fast as possible. Then the share of the ink that forms a tail is particularly small and can be attracted to the droplet by surface forces. Accordingly, the sequence of pressures applied to the nozzle should start with a short phase of negative pressure followed by a wave of positive pressure, whose duration is proportional

to the size of the droplet, and finally again a short phase of negative pressure to remove the droplet (Heinzl *et al.* 1982). This ideal pressure history can only be approached roughly for the following reasons: mostly the pulses are only monopolar, and the pressure history at the nozzle is not similar to the driving pulse. It depends on the kind of pressure propagation and the part acoustics plays in it.

3.11.3.8 Design of the Printheads with Piezoelectric Actuators

The different kinds of piezoelectric actuators mentioned above have all been used in printheads. In industrial applications, a trend toward shear actuators can be seen. Zoltan (1974) had suggested tubes, Kyser and Sears (1976) used membranes, Nilsson tried a comb with cantilever beams (Bolmgren and Nilsson 1982), Larsson (1983) used push pistons, Bartky (1989) introduced shear walls (Michaelis *et al.* 1989), and Fischbeck (1986) introduced shearing monomorphs.

Figure 23 shows design principles with tubes (**Figure 23(a)**), rectangular bimorphs (**Figure 23(b)**), round bimorphs (**Figure 23(c)**), cantilever beams (**Figure 23(d)**), push pistons (**Figure 23(e)**), shear walls (**Figure 23(f)**), monomorphs (**Figure 23(g)**), and with refill through sintered metal (Karlinski 1999) (**Figure 23(h)**).

The first ink-jet printer that entered the market in considerable numbers was the Siemens Printer Terminal PT80. It was the first to use vacuum and it contained piezotubes (Heinzl and Rosenstock 1977). **Figure 24** shows the printhead in cross section. The tubes surround cylindrical cavities that taper toward the nozzles. The tubes have a 2-mm diameter, a length of 10 mm, and a wall thickness of 0.5 mm. Thus the volume of PZT is 12 mm^3. The power that a contraction of one tube might yield is 1.5×10^{-3} J. To eject a drop, an electric pulse of only 10^{-6} J was necessary while the flying droplet contained an energy of 10^{-8} J. The tubes were extremely oversized. The vacuum was established by positioning the ink cartridge 3 cm below the nozzles. The acoustics of the long cavity were tuned by the position of the tube.

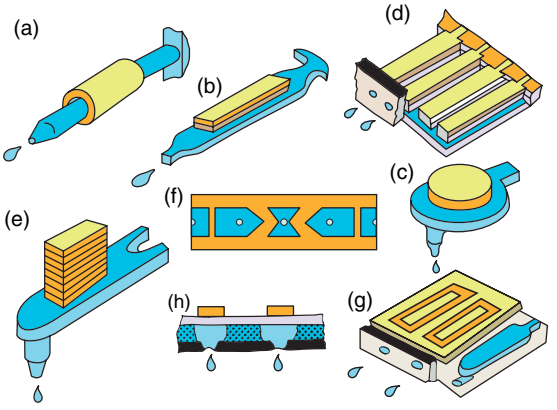


Figure 23 Different configurations of piezoelements and ink systems. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

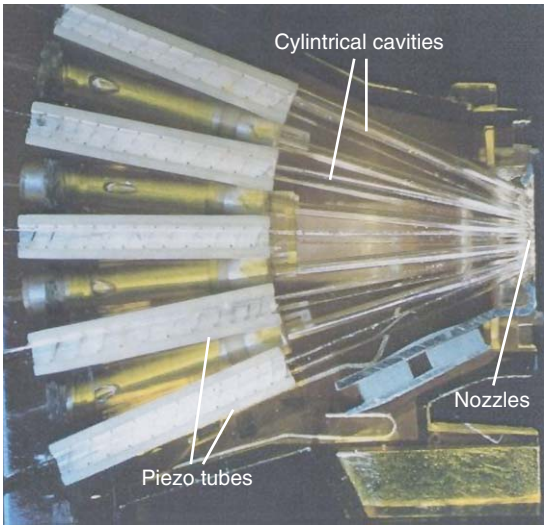


Figure 24 Cross section of the PT80 printhead. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

The bimorph membrane used by Kyser and Sears was long and narrow. The Ink Jet Technology (IJT) printhead shown in **Figure 25** has developed from it, which is an edgeshooter. The cavity structure consists of glass ceramic. To the covering above the cavities, single strips of piezoceramic are glued. Early Epson printheads are similarly designed. **Figure 26** shows two of them, the bodies formed of glass with square or round bimorph actuators. Their ink channels are differently bent to achieve an equal acoustic length. Modern Epson printheads again use lengthy bimorphs. They have succeeded in building the cavities from zirconium oxide, which can be sintered together with the piezoceramic that is formed in thick-film technology. **Figure 27** shows the printhead of the Epson Stylus 64. **Figure 28**

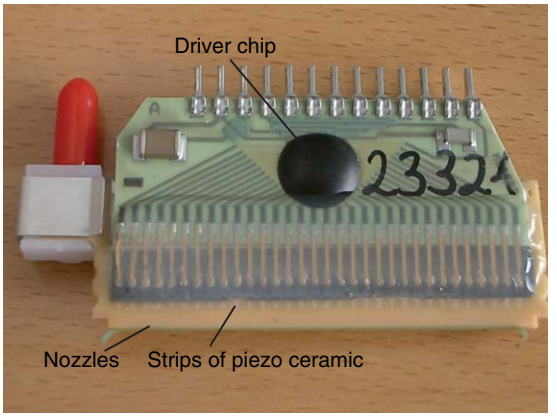


Figure 25 Ink Jet Technology (IJT) printhead. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

shows the actuators after the contacts have been removed. One end of the channel beneath the piezoceramic leads to the manifold, the other is connected to the nozzle through the substrate. **Figure 29** shows a scanning electron microscope (SEM) shot of an opened ink cavity. The PZT layer on top has a thickness of up to 10 µm. The PZT is about 2.2 mm long, and up to 100 µm wide. Altogether the volume of the PZT is 0.003 mm^3. This is less than 0.1% of the PZT of the PT80 tube. This design is realized in the present Epson desktop printers. It is not exchanged together with the disposable ink cartridge (Zhang 2005).

A whole series of printheads uses piezocombs whose teeth are fixed to the membranes covering the cavities and which are acting as push pistons. For this purpose, Epson used 20-µm multilayer ceramic produced by Philips, the Netherlands, in earlier printheads. **Figure 30** shows the design. The cavity structure is built of silicon and covered with plastic material on which the tips of the teeth are glued (Kitahara 1997, Usui *et al.* 1991). Some of these printheads are still in use for large-format printers. Instead of the longitudinal effect used by the multilayer, the diagonal effect can be used, with electrodes arranged along the teeth.

First experiments with printheads containing cantilever beams were made by Kenth Nilsson at Siemens Elema, Sweden. The tips of the teeth of a comb sawn out of a bimorph were placed directly opposite the nozzles and were bent toward them. As the electrodes were in contact with the ink, insulating inks were used. Crosstalk was a problem as well. Recently, this principle has been taken up by Tally (Beurer *et al.* 2003), and small partition walls between the cantilever beams or the teeth of the comb were introduced in MEMS technology. **Figure 31** shows an SEM picture of the inside of a prototype. The advantage of this design is the short distance between actuator and nozzle, which makes the system indifferent toward air bubbles in the ink. Nevertheless, this system did not make it to the market.

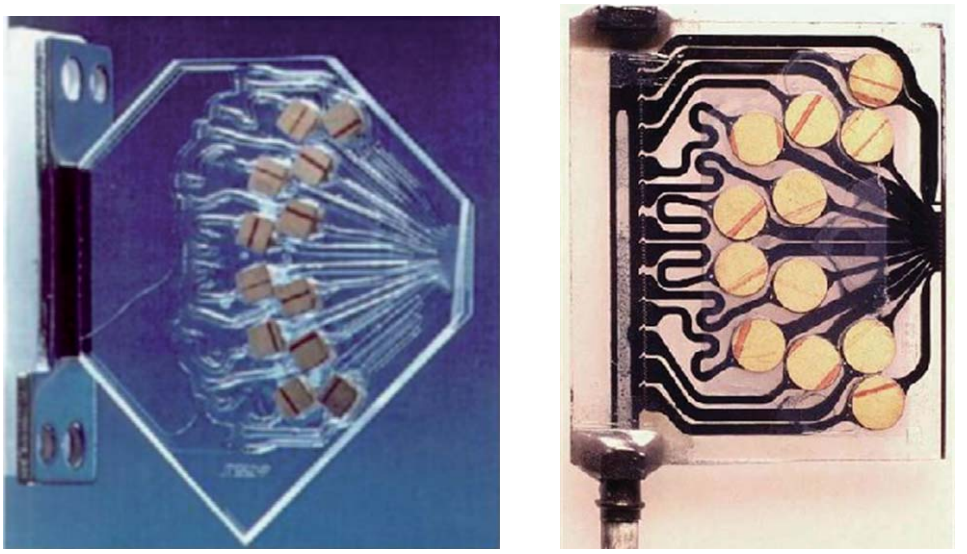


Figure 26 Epson's earlier printheads. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

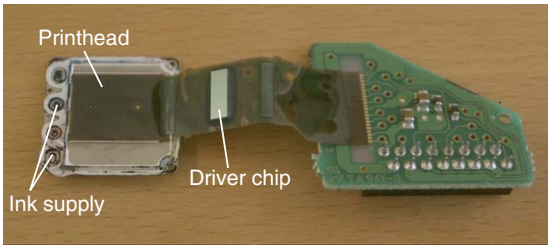


Figure 27 Printhead of the Epson Stylus C64. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

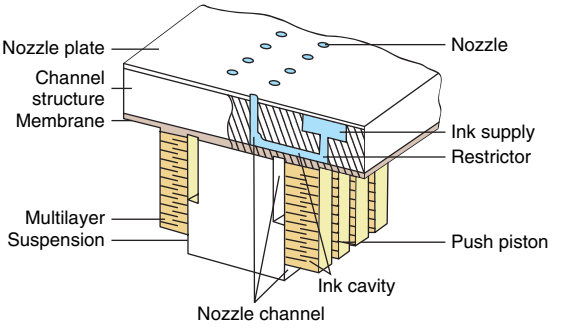


Figure 30 Epson's printhead with push pistons according to EP 1 208 983. (Source: Europäisches Patentamt)

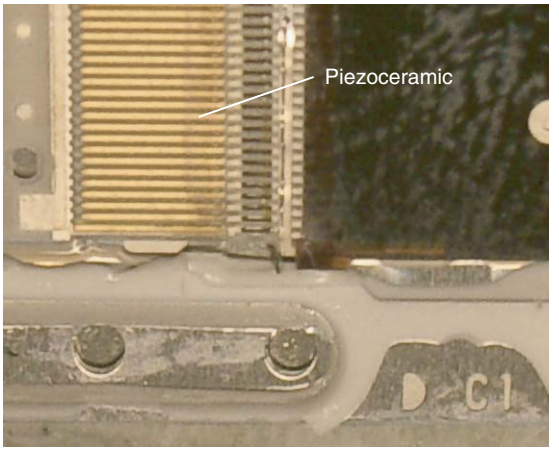


Figure 28 Thick-film piezo of an Epson Stylus C64. (Source: Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

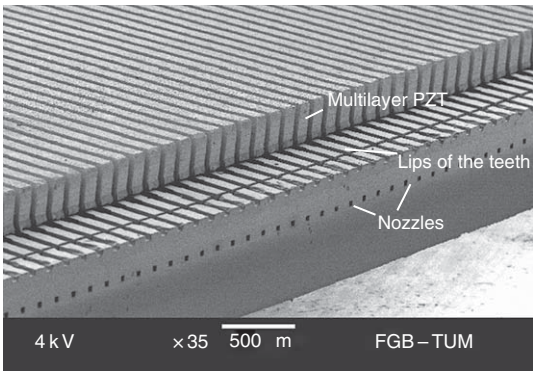


Figure 31 Scanning electron microscope (SEM) shot of the cantilever beams. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

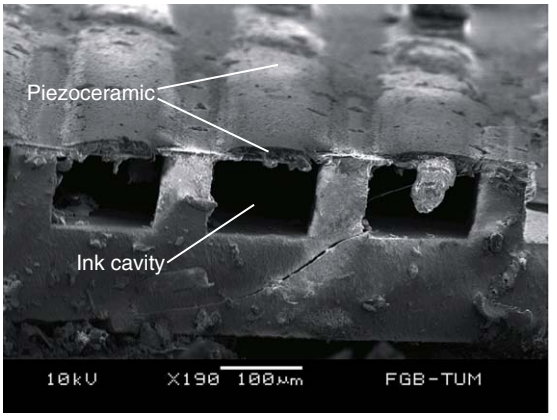


Figure 29 Scanning electron microscope (SEM) shot of an opened cavity (Reproduced with permission from Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

A piezoceramic cavity structure can be obtained by sawing bulk material. XAAR has developed a technology that joins two such structures. Bulk material is polarized before sawing. After sawing, the electrodes used for polarization are removed, and the straight walls of the cavities are provided with new electrodes and are contacted. In this way, the walls can be tilted by the shear effect, thus altering the volume in the channel. Nozzles are formed by laser ablation in a plastic strip that is fixed to the ends of the channels. Because narrowing one channel means widening its two neighbors, every third nozzle can only be triggered simultaneously. The others are triggered with time delay. **Figure 32** shows this design.

The monomorph with electrodes only on one side makes a somewhat different use of the shear effect, as described in Section 3.11.2.5. Spectra uses this principle in its printheads. The piezolayer covers the ink

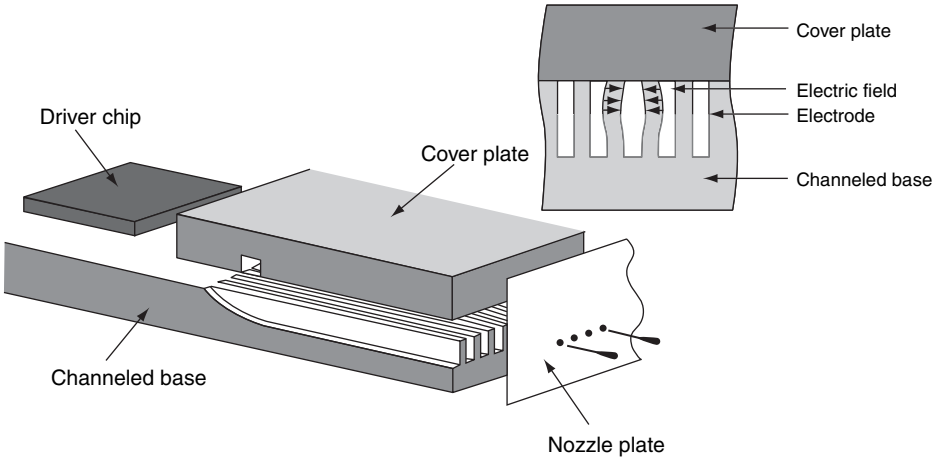


Figure 32 Design of the Xaar printheads. (Source: Xaar plc.)

channels. Above the single channels, the electrodes are separated from the surrounding electrodes by a gap. Only this rim is exposed to the shearing effect and deforms the monomorph thereby ejecting ink. Spectra attaches two rows of channels back to back, interlacing them in order to form a single row of nozzles. Spectra's most recent printheads are produced in MEMS technology offering improved resolution.

Recently Brother has announced a printhead with diamond-shaped cavities and arranged in a honeycomb pattern (Katayama 2004). The cavities are covered with a so-called monomode piezoceramic consisting of a number of layers where only the electrodes are structured.

3.11.4 Print Quality, Colors, and Halftoning

The first ink-jet printers had a pattern structure like that of needle printers. Only 9–12 nozzles composed the letters and the graphs. The distance between the nozzles was 250 µm. Along the lines, the pattern could be refined by introducing quarter steps, whereas the distance between the two neighboring dots is at least one full step. Nowadays it represents 400 × 100 dpi. Thus the slanting lines in particular could be improved substantially. The dot size was suited to the paper in a way that a black field without gaps was created when all positions of the pattern were printed. Since then, ink-jet printing has reached photo quality by increasing the resolution, by the use of three, four, or even six different colors and by applying new halftone strategies.

3.11.4.1 Resolution and Throughput

The resolutions in SOHO printers today range up to 4800 × 1200 dpi or even 5760 × 1440 dpi. The high resolution between the pattern lines is achieved by multipass printing. The printhead crosses the paper several times, each time filling more of the remaining gaps in the writing. Here it makes sense to differentiate between dpi or possible dot positions per inch on the paper (dpi) and nozzles per inch (npi) on the printhead. So a printhead with 300 npi and two rows of nozzles produces 1200 dpi between the raster lines and admitting quarter steps in line direction reaches 4800 dpi when passing the paper four times. Neighboring droplets in the same raster line are not printed at the same pass, thereby coagulation can be avoided. So the resolution in line direction at a single pass is 600 dpi. Hence, in the above example, the basic resolution is 600 × 600 dpi.

The throughput results from the number of nozzles per color multiplied by the maximum drop frequency, divided by the basic resolution, and the number of passes. So a printhead with 300 nozzles per color, a maximum frequency of 8 kHz, and 600 × 600 basic raster doing four passes can print $2''^2\ s^{-1}$ or $12.9\,cm^2\ s^{-1}$. As the printer carriage travels to and fro and needs time for changing direction, this throughput does not come to bear fully. By multipassing, the printhead can print on both ways, forward and back. The position of a dot on the paper depends on the distance from the printhead

to the paper, on the flying speed of the drops, and on the direction in which the carriage moves. Therefore, dots in the same raster line are always printed with the carriage traveling in one direction.

If a nozzle prints with a resolution of 600 dpi along the raster line and with a drop frequency of 8 kHz, it travels at a speed of 0.34 ms^{-1}. The drops fly at a speed of 6–8 ms^{-1}. Because of that the droplet and its long tail do not influence the printing quality. All parts of the droplet meet again in the same dot on the paper.

3.11.4.2 Halftones and Color

Sharp contours and clear letters are obtained by the use of quarter steps. Round and slanty contours are distinctly rendered. The difference between the possible dot positions along the lines and between the lines cannot be perceived with the naked eye, even if flat arcs follow the raster line. As for the contours, photo quality has been obtained. Halftones are more difficult to obtain in photo quality.

There are a number of ways to render halftones in a raster. You can repeatedly trigger small droplets in rapid sequence, vary the size of the drops, use inks of different degrees of saturation, and use halftoning with a fixed drop size using clusters or raster cells (Wild 1990). You can alter the distances between the droplets using error diffusion. Part of these strategies can be interpreted as amplitude modulation (AM) and others as frequency modulation (FM).

It would be ideal to produce droplets in various sizes in order to get a linear gray scale of 16 or 32 steps. Unfortunately, the variation in drop size is difficult to bring about. No more than three different drop sizes have been achieved so far. It is much easier to put together different dot sizes by superimposing varying numbers of small drops. Up to eight microdrops are in use.

If you use only one drop size, you can produce half tones by clustering, and by laying down which dots of the cluster are to be printed in every step of the gray scale. Typical strategies are Digital Halftone and Ordered Dither. Digital Halftone uses neighboring dots in the cluster, and this leads to a greater granularity. Ordered Dither separates the dots within the cluster as far as possible. The transitions in the highlights are especially critical with Ordered Dither.

Error Diffusion does without predetermined clusters. The desired levels of gray are added up along the line. When a threshold is crossed, the next dot is printed and the addition restarts. If in this strategy you only calculate along one raster line, wavy lines may cross the raster lines. Therefore you have to calculate several raster lines at a time and to distribute the dots in such a way that they are staggered in neighboring raster lines.

Normally white paper is used for ink-jet printing. The application of colored inks leads to subtractive color mixing. The colors cyan, magenta, and yellow (CMY) are used. As far as possible, dots of different colors should aggregrate on the paper without overlap. With dark colors this is only possible to a certain extent. The three basic colors can render a big part of the color gamut (Petschik 1994). However, the superimposing of the three colors does not lead to a satisfactory black, but to a brownish gray. Therefore, black has to be used as a fourth ink (cyan magenta yellow black (CMYK)). Because black is used most frequently in the printing of texts, many printers usually provide more nozzles for black than for the other three colors.

There is a major difference between color printing with ink-jet on the one hand and offset printing and laser printing on the other. This concerns registering, which is how the prints of the different colors fit together. Ink-jet printheads carry all the colors and print them in one run of the paper. Offset and laser print the colors in successive runs, thus registering is much more difficult. To avoid moiré, the color screens are slightly twisted, this leads to circular patterns and rosettes for laser printing, especially in light areas.

The difference can be seen clearly in **Figure 33**. The sample on the left side is printed with the HP Laserprinter Color LaserJet 4600, the sample in the middle is printed with the ink-jet printer HP

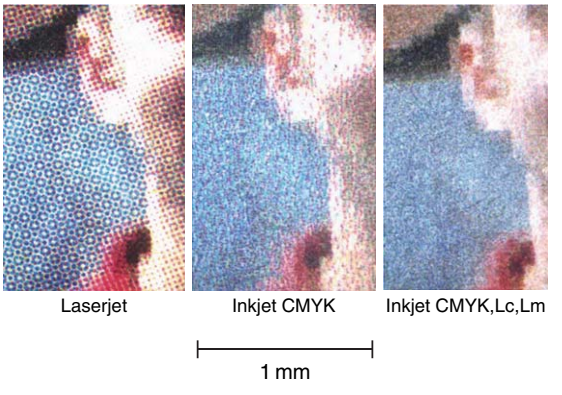


Figure 33 Enlarged samples printed by laser and four or six color ink-jet. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

Officejet 6210. Both samples used the same data. The magnification shows the effect of the different raster strategies. Error diffusion was used in the ink-jet sample and avoids the artifacts that can be seen in the laser print. A further improvement can be achieved by two additional colors. With light cyan (Lc), light magenta (Lm), and CMYK, the granularity and the roughness can be reduced. The sample on the right-hand side is printed with the same ink-jet printer, but using six colors (Allen 1997, 1998). Increasing the number of nozzles by 25% brings on smooth color transitions and low image grain.

Even on plain paper acceptable color prints can be achieved. For photo quality, special coated papers are necessary. They are available on the market in formats from 10 cm × 15 cm up to A4 and for large format as rolls with a breadth up to 60″ or 1524 mm.

3.11.5 Inspection of the Piezoceramic

In this section some instruments and methods are described that can be used to test the function of piezoactuators. The measuring of impedance, the laser Doppler vibrometer, and the SEM are useful in developing an actuator application. The registration of step responses is a powerful tool for quality control.

3.11.5.1 Impedance Measuring

Mechanical stress within the piezoceramic forms instantaneously the moment the field is applied. The transient reaction that relieves the mechanical stress starts after that. Unless hindered, it proceeds at the speed of sound. Depending on the type of ceramic, this sound speed ranges between 3500 and 5000 ms^{-1}. Therefore, the eigenfrequencies of the oscillating ceramic can be predicted easily and measured. The ceramic then acts as a $\lambda/2$ or $\lambda/4$ oscillator. The setup for measuring the electrical impedance applies a sinusoidal alternating voltage to the electrodes of the polarized ceramic and shifts through the frequency range in question. Thus, the ceramic acts as a capacitor. As soon as it enters an eigenfrequency, the amplitude response deviates, even more so does the phase response, because energy is oscillating in the ceramic.

These responses are typical for the magnitude of efficiency and damping. **Figure 34** shows the setup that can measure up to 50 MHz with low-level signal and the resonance peek on the display.

Figure 34 Setup for measuring impedance. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

A tube fixed in the middle will show two eigenfrequencies, lengthwise it is a $\lambda/2$ oscillator, within the wall it is a $\lambda/4$ oscillator. Eigenvalues of 200 kHz and 2 MHz correspond to the length of 10 mm and the wall thickness of 0.5 mm at a sound speed of 4000 ms^{-1}.

Even after the ceramic has been fitted into the printhead, eigenvalues can be measured. Of course, the eigenvalues of a bimorph cannot be calculated easily as the ceramic is fixed to a second layer. This layer, the ceramic, and the glues in between form the oscillating system. The viscosity and the thickness of the glue are difficult to predict. Although the setup measures with low-level signals, a difference can be seen between a full and an empty ink cavity. It is difficult to make these measurements on the actuators of a ready-made printhead since in most cases a serial-to-parallel converter is integrated and whose contacts would have to be uncovered first.

3.11.5.2 The Laser Doppler Vibrometer

The displacement and the motion of the ceramic can be determined by means of the laser Doppler vibrometer. **Figure 35** shows the setup. Fiber optics is coupled into the microscope. Typical deviations of a bimorph range from 10 to 40 nm. The vibrometer resolves motions as small as 2 nm of the piezoelements in one dimension.

3.11.5.3 Measurements Using the SEM

Especially with multilayer ceramic, it is of importance to test whether all electrodes are in contact or if

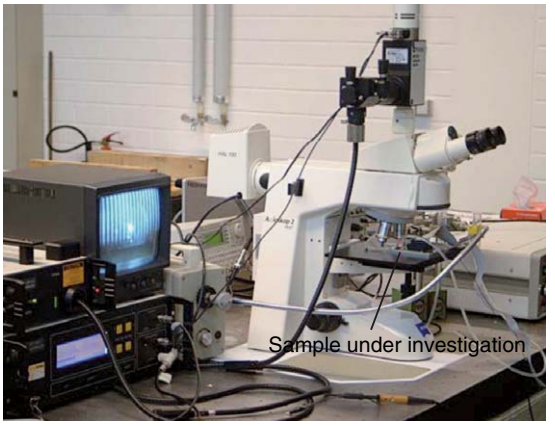


Figure 35 Laser Doppler vibrometer. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich)

electrodes between the ceramic layers are interrupted or broken. Anyone working with the SEM knows about the frustrating phenomenon of insulating materials getting charged and flaring the picture. You can profit from this phenomenon if you provide contacts to control from outside the electrodes in the SEM. **Figure 36** shows the electric field applied to the multilayer ceramic from outside the SEM. By this method, misfunctions of the electrodes can be detected.

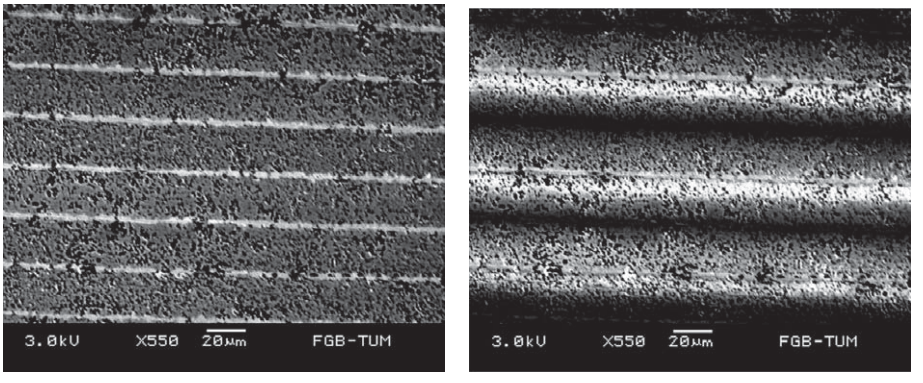
3.11.5.4 Registration of the Step Responses

The first piezotube printheads had to be provided with a potentiometer for each actuator in order to compensate the efficiency. Such an individual adjustment is no longer necessary in today's printheads. The quality of the ceramic and its processing lead to a fairly uniform function of the actuators. At most, an attenuation of the driving pulses of the entire printhead is provided. For quality control, the examination of every single actuator is recommended. This should happen before the serial-to-parallel converter is connected and before ink is filled in. For this purpose, the piezoelectric actuator can be used as a sensor. The actuator is charged slowly. After a sudden discharge, a step response can be seen at the contacts if the electrodes are disconnected from the control. The actuator oscillates and the step response shows its eigenvalues, its efficiency, and its degree of damping. Within some seconds, hundreds of actuators can be tested and their step responses can be registered (Schindler 2005). **Figure 37** shows the printout of such a protocol.

3.11.6 Examination of the Drop Formation

For the ultimate examination, a printhead prints test patterns after being installed in the printer. The patterns that are used show very quickly whether each nozzle prints as required. Helpful test patterns are also installed for the user and can be invoked on the printer. So misfire or missing nozzles can be detected easily.

To examine the formation and the flight of the drops more closely an observation setup like that in **Figure 38** is recommended. The printhead is



Ungeladen Geladen

Figure 36 Charges made visible in the scanning electron microscope (SEM). (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

fixed adjustably and single nozzles are observed. This is done with two telescopic microscopes in stroboscopic light. One of these microscopes looks at the drops in profile, the other looks at the nozzle orifice sideways. Both pictures can be seen on the monitor and can be stored.

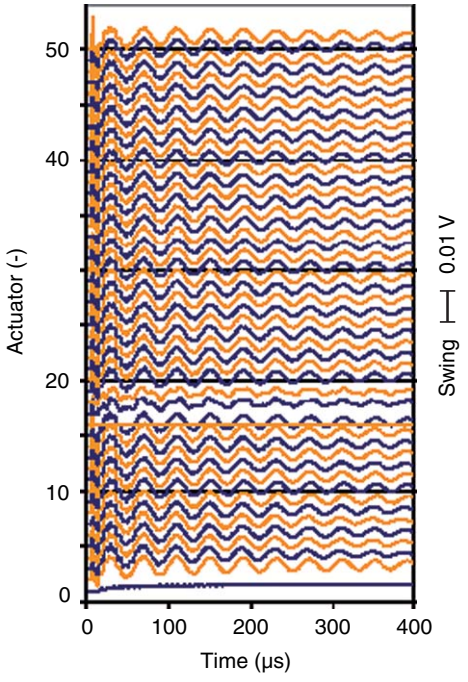


Figure 37 Step responses. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

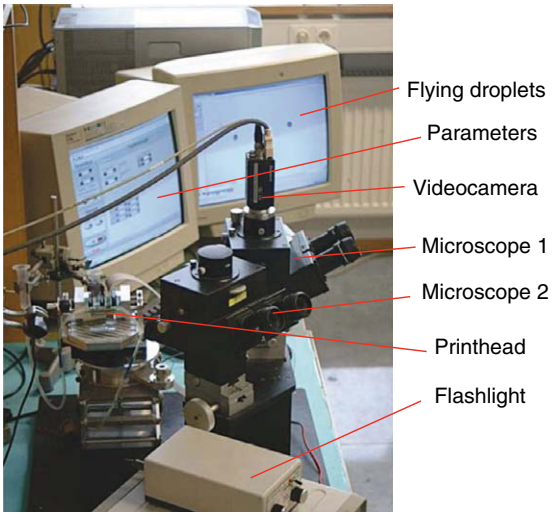


Figure 38 Setup for drop observation. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

3.11.6.1 Flight of the Drop

The profile shows the formation and the disintegration of the drops. If the frequency of the stroboscopic flash is set lower than that of the drops, one observes the drops emerge from the nozzle, disconnect, and fly on. Mostly, drops draw a long tail that disintegrates into small droplets flying at different speeds. Figure 39 shows a typical droplet on the flight. From these pictures you can conclude what the dots will look like on paper if you take into account the distance to the paper and the printing speed. Very high speeds can be achieved by arranging several printheads in series. This is interesting in applications where the printheads are stationary and only the paper moves. The more the printing speed approaches the flying speed of the drops, the more distinctly the drops with the ensuing tails show on the paper. These applications call for high-speed drops that draw in their tails on the flight.

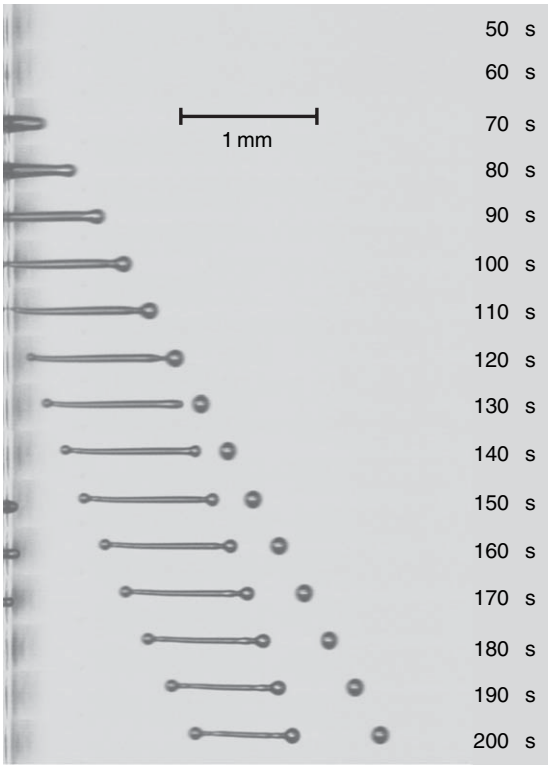


Figure 39 Flying drops. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

3.11.6.2 Formation of the Drop

The second microscope of the setup observes the nozzle. If drops that deviate from their path or slow down are detected, this second microscope starts functioning. It shows whether the flow separation is concentric to the nozzle or whether it starts at the rim. You can detect dust particles or uneven wetting. If the drop is ejected irregularly or at a low speed, air or gas bubbles inside the nozzle can be the reason. You can also check how far the meniscus is pulled back into the nozzle before ejecting the drop and how long it takes to refill the nozzle. Should the meniscus be pulled back too far, air can enter the nozzle.

3.11.6.3 Frequency Response

If you synchronize the stroboscope with the drop frequency, you get a standing picture of the drop. If you shift the time delay, you can choose in which position you want to view the drop. You can, for example, select the position in which the drop would hit the paper. If you change the frequency of the drop sequence, you see how the drops change their shape and speed in relation to the frequency of their ejection. As long as the frequency is low, the influence of the frequency can hardly be detected. The meniscus in the nozzle comes back to its starting position after every drop ejection, and its vibration dies away long before the next drop is triggered. If the sequence of the drops is increased, this steady-state position can no longer be reached completely. The vibration of the meniscus influences the following drop. The drops are faster or slower, depending on the phase of the fading vibration in which they start. If you increase the frequency further, these changes in speed increase. Eventually the drop generation breaks down, the nozzle can no longer be supplied with ink, and air enters the nozzle. If you skip these frequencies, you can find higher-frequency ranges in which drop generation works again.

For printing, a continuous frequency band is not needed, and because the drops are normally printed in a defined pattern, only discrete frequencies are needed. To make use of the full speed, you can select the required frequencies from the frequency response.

3.11.6.4 The First Drop of a Sequence and Crosstalk

In order to examine the usable frequencies in more detail, special patterns of drive pulses can be used. Most instructive are periodic patterns. In a sequence of eight drops, for example, you eject five drops in sequence and then leave a gap. You synchronize the stroboscope to the second drop and, with this pattern of five out of eight, you run through the frequency range. The distances between the second and the following drops behave as expected. Only the first drop moves to and fro, and these deviations cannot be detected from the simple frequency response, because at every frequency the meniscus has reached another stable position when the next drop comes along. The short interval within the above pattern takes it back to its stable rest position or at least near to it. Therefore, the speed of the first drop deviates from the speed of the following. In addition, it must be said that small drops are influenced by air resistance. The following drops fly in the wake of the first, so to say.

Depending on the design of the cavities and the actuators, neighboring channels can interact (McDonald and Zhou 1999). In systems with cantilever beams, this crosstalk is especially strong if the cavities are not completely separate and pressure waves can spread to the neighboring cavities. In order to examine the effect, the drop formation of a single nozzle is observed, while the neighboring nozzles are activated synchronously or with time shift throughout the frequency range.

3.11.6.5 Simulation of Drop Generation

Modern software tools allow the simulation of drop generation. For example, with FLOW 3D the motion of the actuator and the motion of the fluid can be calculated. Even the formation of drops and their flight as well as crosstalk can be included into the simulation. Seitz used a combination of two different software tools (Seitz 2003). With a fast laptop it took him more than 24 h to simulate the 250 μs from the start of the driving pulse up to the separation of a drop and the refill of the nozzle. That is about 4×10^8 times real time. He did it for the prototype of a Tally printhead (Seitz 2002, Yu and Sakai 2003).

Scheicher used SIMULINK for the simulation of the same printhead. He had to introduce some simplifications. He took into account the size and the stiffness of the actuator, mass acceleration, and the viscosity of the ink in the nozzle in the restrictors and the newly created surface. The speed and volume of the ejected drops were calculated in this manner. He found that 10 min were required for one shot. This is

in the range of 10^6 times real time (Scheicher 2004, Voigt 2002).

A version of this program was written in VisualBasic (Irlinger 2007). As a result, it took only 50 ms to simulate the 250 µs, i.e., 2×10^2 times real time. Even in this simple model, 18 parameters are involved: length and cross section of the nozzle, of the restrictor, and of the connections to the neighbors; viscosity, density, and surface tension of the ink; size and stiffness of the actuator; voltage, rise time, and length of a first driving pulse, and in some cases of a second. Such a simple model can be easily adapted to the experimental data. Speed and volume of a single drop are known to depend on voltage and length of the driving pulse by measuring as well as through crosstalk. Some of the parameters are known exactly, others approximately, and the rest is fitted to the data of a single drop known from the experiments.

The advantage of the short runs is that one can simulate not only a single shot, but also the frequency response and one can identify the sensitivity of the system from every single parameter. The calculated frequency response shows how far the system behavior agrees to the measured data. This shows that the simplifications are justified. After fixing a set of nominal values, one can vary every single parameter while all the others are kept. **Figure 40** shows a set of results. Every parameter is varied from half the nominal value to twice the value. The blue line represents the speed of the drop and the red one the volume. The scales on the *y*-axis always remain the same.

For the understanding of the behavior and for the optimization of the design in combination with the measuring of prototypes, a fast and simple model is much more helpful than the simulation in commercial fluid dynamic computer software.

3.11.7 Industrial Applications

Wherever data for production are represented in hardware – master plates, masks, molds, printing plates, negatives, etc. – to be used in the production process, sooner or later there will be a direct digital procedure without intermediate storage in hardware. Reliability, speed, and quality of the digital transformation process are decisive here. In this connection, DOD procedures play an important part. This development has only started as yet. Compared to the classical printing technologies like offset, the speed is not high enough so far. Up to now, it has been easier to replace screen printing and dosing technology. And with 3D printing, completely new production technologies arise.

3.11.7.1 Printheads for Industrial Application

Printheads are pivotal in industrial applications. Reliability and lifetime are of vital importance here (Schwartz 2002). So far thermal ink-jets have a limited lifetime and have severe restrictions as to ink. In almost every industrial application, piezoprintheads are gaining ground. Two technologies using the shear effect are to the fore, the printheads of Xaar and of Spectra, i.e., Dimatix. These printheads are available on the market and both companies support OEM applications. Besides, there are printheads from IJT, Trident, Ricoh, PicoJet, PixDro, and many others. To obtain high speed and high throughput on an industrial scale, printheads moving to and fro will be replaced by stationary printheads covering the full printing width, so that only the workpiece has to be moved. Such printheads should contain an enormous number of nozzles (Allen 1996). Obviously, a great number of companies are developing such print beams or line heads in piezotechnology (Zhou 2000, Zhou and Rensch 2001). Time and time again, there are announcements and presentations of such beams in fairs. In 2003 Panasonic announced a 600 dpi line head with 400 nozzles using a 3-µm sputtering process with a lifetime of 10 billion or more drops per nozzle. Recently, Brother has presented a line head that will be discussed later. Of course, such beams of many single printheads can be composed.

Xaar offers a whole series of printheads, which all make use of shear technology on thin walls (Beurer and Kretschmer 1997). The company's OmniDot760 in **Figure 41**, for example, uses 764 nozzles in 360 npi and distinguishes between binary mode and gray-scale mode. Its maximum drop frequency is 7.8 kHz with a maximum drop size of 40 pl and six shades of gray. In the binary mode, the drop size is 8 pl at up to 40 kHz.

The IJT head 64 ID2 in **Figure 25** contains 64 nozzles at 40 npi and it reaches 6 kHz. The piezo-actuators are glued to the layer that covers the channels. Three such printheads, slanting, can be combined for a resolution of 187 dpi and a width of 24.9 mm. In 2003 IJT Inc. was sold to DGI (Digital Graphic Incorporation), Seoul, Korea.

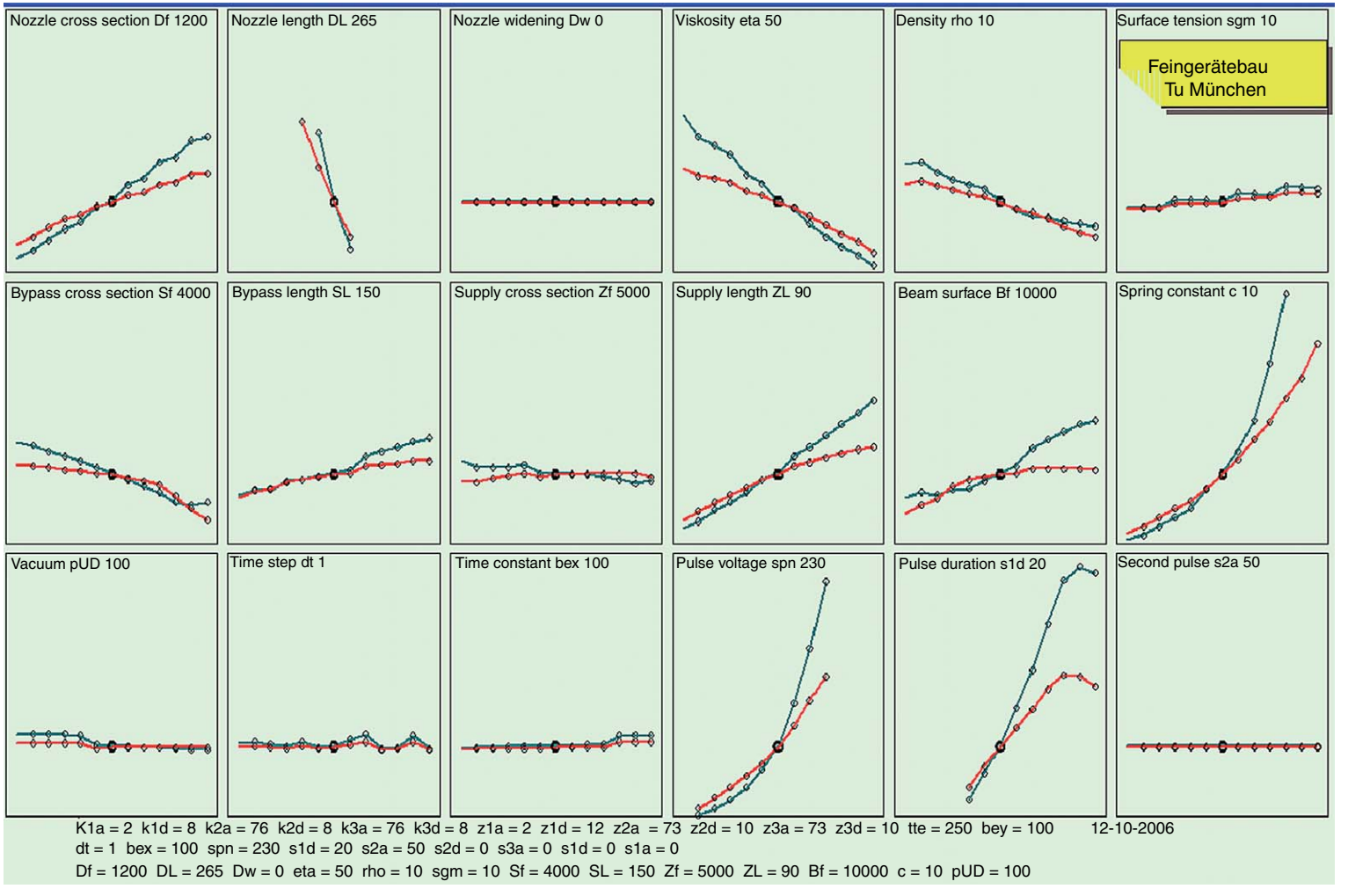


Figure 40 Simulation of sensitivity. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

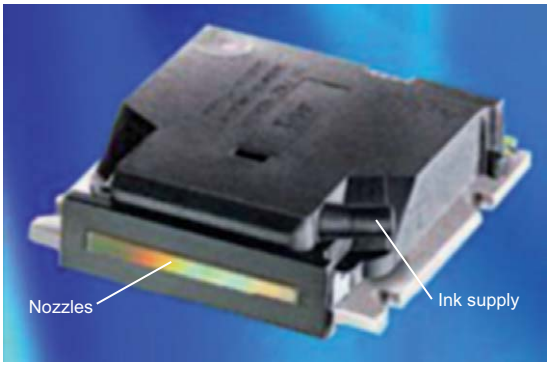


Figure 41 Xaar's OmniDot760 printhead. (Source: Xaar plc.)

Trident offers a particularly robust printhead, the dismountable 768Jet™, as shown in **Figure 42**. It contains 768 nozzles in 256 controllable channels on a length of 100.8 mm. The nozzle plate is built from stainless steel, and push pistons are used as actuators (Chen 2005, 2006, Zhou 2004).

Spectra's printheads use the shear effect. Spectra is a division of Dimatix. The piezolayer covers the ink cavities as a monomorph (Zhou 1997). **Figure 43** shows the SL-128AA with 128 nozzles at 450 npi. Drops of 80 pl fly 8 ms^{-1} at 30 kHz (Duby 2001). There are variants with 50 and 30 pl. Inks with organic solvents and UV-curable inks are used at temperatures of up to 90°C. Systems for water-based inks are also available.

Dimatix has developed monolithic printheads in MEMS technology, the so-called M-Class (Creagh 2005b, Menzel *et al.* 2004, Torrey 2005). The new M-300/10 with 304 nozzles at 180 npi creates 10-pl drops and has a gray-scale mode using multiple

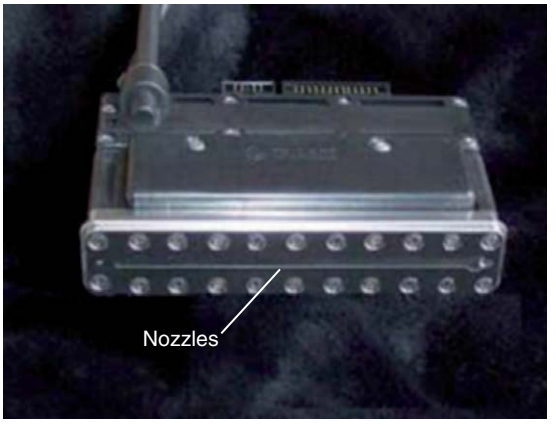


Figure 42 Trident's 768Jet™Printhead. (Source: Trident.)

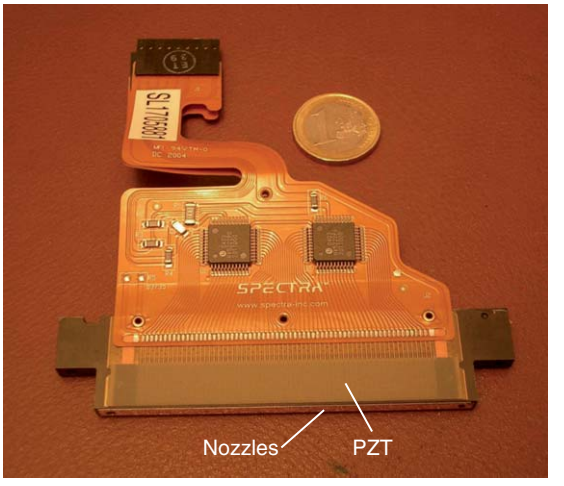


Figure 43 Spectra's SL-128AA printhead. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

pulses. **Figure 44** shows the restrictor at the end of the channel (**Figure 44(a)**), the entrances to the nozzles (**Figure 44(b)**), and a nozzle etched in silicon (**Figure 44(c)**). Dimatix offers this printhead with a complete dosing system Dimatix Material Printer DMP-2800. The printhead can be rotated to adapt to the printing pattern. DMP-2800 is shown in **Figure 45**.

PixDro's printhead PL 128 is also built using MEMS technology (PixDro 2006). As it consists mainly of silicon and glass, even corrosive fluids can be handled. The printhead contains two rows of 64 nozzles each and is piezoactuated. Drops of 6–30 pl fly at a speed of 5–10 ms^{-1} at 25 kHz. Complete systems are working for flat panel display (FPD) manufacturing, plastic electronics, glass and metal etching, etc.

PicoJet offers printheads with 256 nozzles and 300 dpi (Le 2005). The sideshooter contains 16 rows of 60 nozzles each, and is made of layers that are welded together by ultrasonics. The piezoceramic is applied from outside and does not come into contact with the ink.

Brother has presented a novel line head, the Piezo Inkjet Line Head (Hirota and Ishikura 2005, Katayama 2004, Sakaida *et al.* 2003). It was developed in cooperation with Kyocera Corp. On a length of 108 mm, 2656 nozzles are arranged. It is also built from layers of stainless steel and is welded together. The ink cavities are diamond shaped and arranged in trapezoid fields like honeycombs. The sideshooter has dimensions of

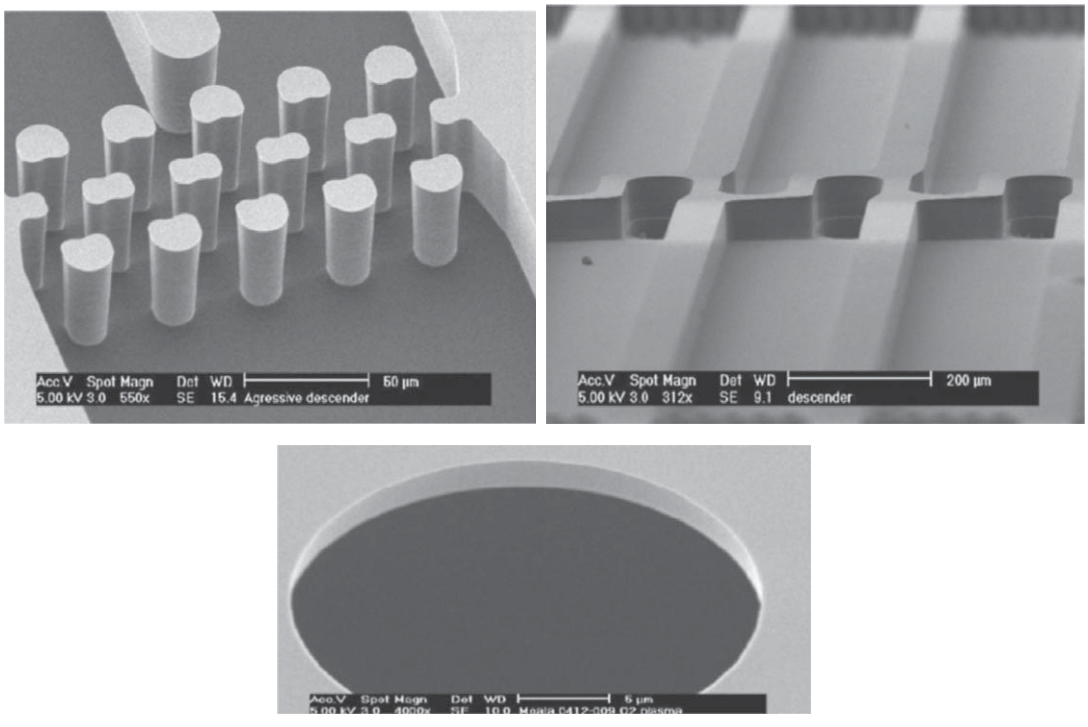


Figure 44 Scanning electron microscope (SEM) shot of the restrictor and the nozzle of M-class. (Photo courtesy: Fujifilm Dimatix.)

Figure 45 Dimatix Material Printer DMP-2800. (Photo courtesy: Fujifilm Dimatix.)

152 mm × 22 mm × 1 mm. Characteristics of 20 kHz, four different drop sizes, and a lifetime of 100 billion drops are predicted. A print speed of 0.847 ms^{-1} at 600 dpi is assumed to be possible.

3.11.7.2 Two-Dimensional Printing

With high production runs, nonimpact printing (NIP) has so far not had a chance. Conventional printing technologies like gravure printing, relief printing, and offset cannot be ignored with respect to cost and speed. Things are different for low runs and personalized printouts. Besides laser printing, color ink-jet printing is especially on the rise. At the beginning of the 1990s, the distribution according to the value of the printed matter was 15% gravure printing, 25% relief printing and flexo, and 40% offset. Following international estimates, a growing percentage passes over from offset to NIP (Kipphan 2000, p. 1019). In absolute numbers, all sectors are growing since the demand of printed matter is increasing steadily.

If ink-jet line heads can be introduced successfully, they will reach a speed of about 1.5 ms^{-1}, compared to sheet offset with 4 ms^{-1} and roll offset with up to 15 ms^{-1}. Even at present, sheet offset is competitive only for batches of 400 or more, and roll offset for batches of 4000 or more as compared to NIP. According to Kipphan (2000, p. 0108) this is true for double-sided four-color prints on A3.

But ink-jet is catching up. The Agfa Dotrix is a modular print system based around ink-jet arrays. It prints 300 dpi with gray scale in CMYK with UV-curable inks. The web width is 0.65 m and the linear

speed ranges from 5 to 24 m min^{-1} (van Daele 2005, Willis 2006). For 2007 Olympus has announced the Wep Printer OP-1 for duplex printing with 12.5″ width and 40 m min^{-1}. Fixed print head arrays print on a drum. A real-time misfire detection will be integrated (Masakawa 2006). Konica Minolta has presented a road map for a single pass printing machine for the years 2006–2008 (Corrall 2006).

Glunz and Jensen have presented an interesting application of ink-jet in offset printing. The PlateWriter™4200 produces press-ready aluminum plates for offset without the use of chemical processing. The patterns are written on the plate with special ink by an ink-jet printhead, the liquid dots are dried and bonded to the plate, and protected with a gum layer in a finishing unit. This technology can make sheet offset profitable for small runs (Beals 2004).

3.11.7.3 Wide-Format Flatbed Printers

In the wide-format market, an increasing number of ink-jet printers is offered even for printing on corrugated cardboard or other materials. On the US market, the number of newly introduced flatbed machines has doubled yearly since 2001, and you can now find more than 50 systems offered by different companies. These machines print at speeds of 40–200 m^2 h^{-1} with resolutions between 300 and 720 dpi. **Figure 46** shows a selection of European devices.

Most of the industrial wide-format printers use UV-curable inks which can produce gloss if desired. Some of them can print white inks to apply the right colors even on colored or metallic substrates. **Figure 47** shows prints on glossy golden paper with (**Figure 47(a)**) and without (**Figure 47(b)**) white underlay.

The machines print on paper, cardboard, corrugated cardboard, wood, vinyl, foam-core board, plastics, acrylic sheets, Plexiglas, glass, metal, canvas, cotton, and T-shirts. In **Figure 48** a 5-mm aluminum panel is printed on. For textile printing, a lot of new inks (Hees *et al.* 2003) and special printheads (Le 2005) are under development, and there are ideas to integrate the weaving and the printing processes (George *et al.* 2005).

In 2005 HP took over Scitex, the leading expert in large format printing. HP Scitex now offers extremely fast printing systems. The HP Scitex TJ8300 in **Figure 49**, for example, offers a multihead, DOD piezoelectric ink-jet printing with six colors (CMYK, Lc, Lm), formats up to 163 cm × 360 cm and a printing speed of 400 m^2 h^{-1} at up to 336 dpi or 100 m^2 h^{-1} at 600 dpi. The printhead Aprion MAGIC (multiple array graphic ink-jet color) is the heart piece of the machines. It is a printhead with ink supply through porous medium (Karlinski 1999, 2000). It is said to achieve 150 kHz in the laboratory.

These printheads have been used in the DReAM textile printer that was introduced in 2002 in Bergamo, Italy, in cooperation with Reggiani Macchine S.p.A., Aprion Digital, and Ciba (Nespaca 2004). Forty-two Aprion Magic PIJ printheads have been used each having 512 nozzles, with each jetting up to 35 kHz. There are 21 504 nozzles or 7.5×10^8

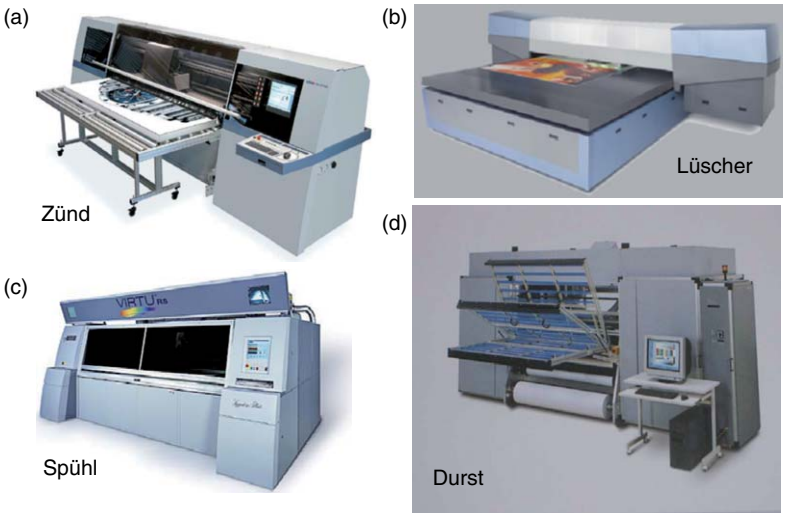


Figure 46 European wide-format flatbed printer. (a) Zünd. (Source: Zünd AG.) (b) Lüscher. (Source: Lüscher AG.) (c) Spühl. (Source: Spühl AG.) (d) Durst (Reproduced with permission from Durst Phototechnik.)

Figure 47 Durst Phototechnik AG prints on golden paper. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

Figure 48 Decoration printing. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

Figure 49 The Hewlett-Packard (HP) Scitex printer TJ 8300. (Source: HP Scitex, Hewlett Packard GmbH.)

total drops per second (TDPS). On a breath of 1.6 m, the machine prints up to 150 m h^{-1} at 600 dpi using special inks developed by Ciba (Cahill and Ujiie 2004).

3.11.7.4 Packaging

Printing on corrugated cardboard opens a lot of possibilities to produce single pieces and small series of cartons and packages, especially in combination with cutting by laser. The speed of ink-jet printing is still far away from the speed of flexoprint, which is 4 m s^{-1}. When printing comparable widths, the fastest flatbed printers reach a mere 12 mm s^{-1} with their printheads running to and fro. This can change as soon as line heads are installed. Up to now UV inks are still much more expensive than flexoprint colors. Nevertheless, DOD promises a lot of advantages over flexoprint machines: smart and small machines, economical productions of small batch sizes, no resetting while patterns are changed, personalization and customization, thereby improved security, the length of prints independent of the size of the cylinder, no flexible printing plates, no waste of board and color, and less environmental pollution (Heinzl 2005, Springford 2004).

For a breadth of 25 mm, Jetrion offers print systems that can be used to print at a speed of 5 m s^{-1} and a resolution of 100 dpi on packaging. They are used for addressing and consist of a staggered arrangement of Spectra printheads in connection with special UV and solvent-based inks.

In 2007 Sun Chemical Digital will launch a digital printing machine Fastjet for corrugated packaging. It contains 480 Spectra printheads with 128 nozzles each. About 1 m^2 can be printed in 0.8 s and will use full-color CMYK plus varnish for UV curing (Gross 2006).

3.11.7.5 Printed Electronics

Nanotechnology allows a lot of applications in plastic electronics. Inks with nanoparticles are used to print conductor paths, to build passive components, diodes, and semiconductors (Bentley 2005, Creagh 2005a, Creagh and Schoeppler 2005, Grooten 2006, Schöppler 2006). There are many areas for applications with integrated circuits (ICs), antennas, radio frequency identification devices (RFIDs), in photovoltaics, and with batteries. In the production of FPDs with organic light-emitting diodes (OLEDs) DOD is already in use (Dirksman *et al.* 2005,

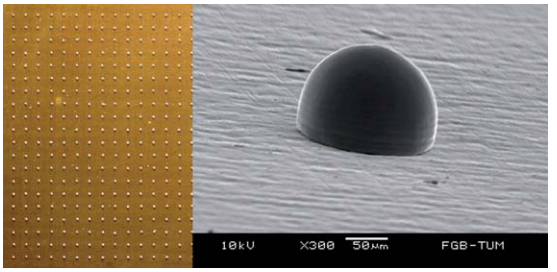


Figure 50 Printed solder bumps. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

McDonald 2003). One can even think of flexible displays and electronic papers (Derbyshine 2004).

Printed electronics will not compete with conventional electronics in the productions of large series. For rapid prototyping and small series, it is advantageous if there is no need for masks or stencil. This is also true for the contacts and the soldering. For single boards the solder bumps can be applied by ink-jet. Lemmermeyer and Wehl have developed a printhead for solder (Lemmermeyer 2006). There is a heater for temperatures higher than 300°C and for piezopush pistons. The tip of the ceramic is not polarized. It pushes against the membrane and separates the polarized ceramic from the hot cavity. Nozzles and contacts are under protective gas. **Figure 50** shows the printed samples.

3.11.7.6 Dosage

Wherever contact-free dosage is needed, DOD systems fit best. That applies for biotechnology and microbiology as well as for medicine, and for analytic as well as for production engineering (Sumerel 2005). The application of DOD in these fields is starting slowly. The Dimatix Material Printer DMP-200 is a good example for a universal device that can be used for applications like rapid prototyping, glass and metal etching, photoresist application, and many more.

3.11.7.7 Three-Dimensional Printing

As in 2D printing, there is a competition in 3D modeling between laser- and drop-based systems. Rapid prototyping, the possibility to produce exemplary embodiments, is developing faster toward rapid manufacturing. Besides stereolithography and laser sintering, 3D printing is gaining ground. In all these technologies the 3D body is built from horizontal layers that are stacked one upon the other and structured by laser beam or drops. In general, drop-based methods are cheaper and more versatile than laser-based ones. They are faster, less restricted as to size, and can print parts in different colors. Apart from fragile embodiments that show the geometry of envisaged construction elements and that serve as models, functional parts in the original materials and with the necessary strength are produced and used (Wohlers 2006). Three-dimensional printing even allows the printing of molds to cast metal and bone substitutes.

Z Corporation offers 3D printers for the printing of design concepts. Spectrum Z™510 offers a building volume of 254 mm × 356 mm × 203 mm with adjustable layers between 0.1 and 0.2 mm. It takes 30 s to print one layer with 1216 nozzles in four printheads. Depending on the powder used, either stiff or elastic structures are obtained (Reid 2005). The colored inks that serve as binders can be mixed and are printed at a resolution of 600 × 540 dpi (Tremblay 2005). **Figures 51–53** show the printer and print samples.

Prometal RCT offers machines that produce cast molds for metals of a size up to 1500 mm × 750 mm × 700 mm. The molding sand is mixed with a curing agent and is brought out layer by layer. The binder is printed after the layer is flattened. After finishing the last layer, the loose sand is sucked off. Then the mold can be mounted and cast. This method can produce far more complex castings than conventional methods, and time, above all, can be saved. The interval between providing CAD data for the mold and the cast is shortened – typically from 20 to 5 days

Figure 51 Z Corporation's 3D color print system Spectrum Z™510. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

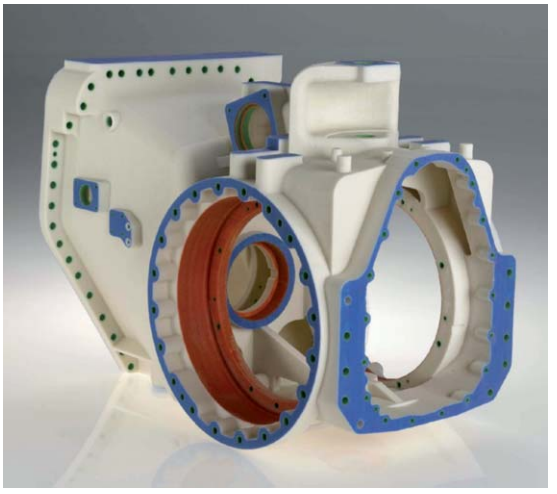


Figure 52 Design model of a housing. (Source: Z Corporation UK Ltd.)

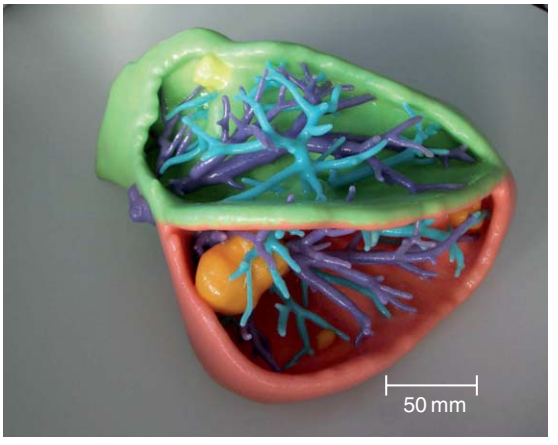


Figure 53 Model of a human liver. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

(Koewius 2005). **Figures 54–56** show the machine and cast examples.

Voxeljet Technology (Ederer 2005) offers a machine that produces plastic parts made from polymethymethacrylate (PMMA). The digital production plant VX 800 prints volumes up to 850 mm × 450 mm × 500 mm. It prints layers of 150 µm at a speed of 36 mm h^{-1}. Modified acrylic glass is infiltrated with epoxy or wax and solvents are used as binders. According to the company's indications, the resistivity is 18 MPa, strain at break is 2.9%, and Young's modulus is 1700 MPa. **Figures 57 and 58** show the machine and a digitally formed plastic part.

Figure 54 ProMetal S15 for rapid castings. (Source: ProMetal RCT GmbH.)

Figure 55 Cast molds and cast. (Source: Voxeljet Technology GmbH.)

Figure 56 Cast mold after sucking off the loose sand. (Source: from Voxeljet Technology GmbH.)

Experiments at the Technische Universität Munich (TUM) have shown that porous bone structures made from hydroxyapatite can be printed as a green pellet with an organic binder (Pelzer 2005). The organic components are burnt out during the sintering. Living cells grow into the structure using it as a scaffold (Zeilhofer 1998). It is envisaged that patients can get personalized implants settled with their own cells and thus can be healed quickly. **Figure 59** shows a scaffold for a lower jaw.

Figure 57 VoxelJet's VX 800. (Source: Voxeljet Technology GmbH.)

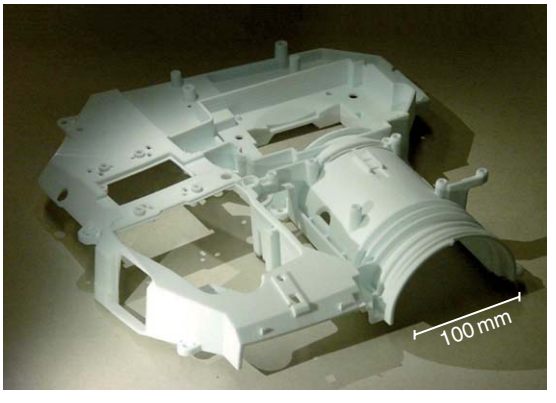


Figure 58 Printed plastic part. (Source: Voxeljet Technology GmbH.)

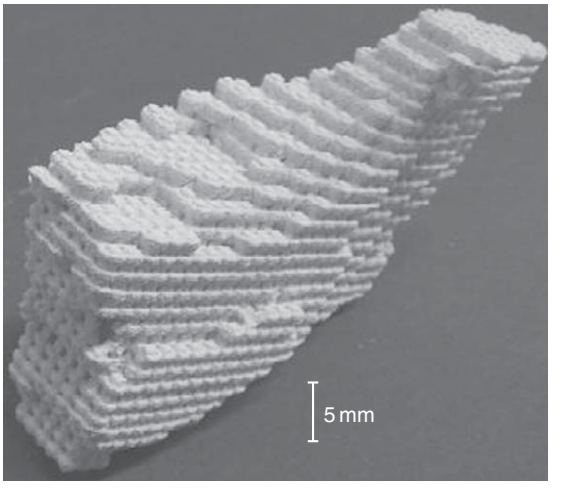


Figure 59 Printed bone structure. (Source: Lehrstuhl für Mikrotechnik und Medizingerätetechnik, Institut für Mechatronik, Technische Universität Munich.)

3.11.8 Summary

DOD ink-jet technology holds an increasingly big market share in SOHO printers. In addition to that, especially piezo-driven printheads offer a great number of further applications. The great variety of designs shows that this technology continues to develop and has not yet come to an end. The mode of functioning of the different actuators and the creation of drops are discussed. Measuring methods and quality checks are described. The chapter gives an overview over the printheads available on the market and of their applications.

References

Allen R R 1996 Ink jet printing with large pagewide arrays: Issues and challenges. *12th Int. Congr. Advances in Non-impact Printing Technologies Proceedings*, San Antonio, TX, USA, p. 43

Allen R R 1997 Black and color resolution enhancement technology used in the Hewlett-Packard DeskJet 850C Printer. In Rezanka I and Eschbach R (eds.) *.RecentProgress in Ink Jet Technologies*, IS&T, Springfield, VA, USA, pp. 147–9

Allen R R 1998 *Photographic Quality Imaging with HP Thermal Ink Jet*. Printing Technology Department, HP Laboratories, Palo Alto, CA

Ascoli E, Pidoux R, Pedrix M 1975 *US Pat.* 3 893 126

Baker J P, La D T, Coverstone R A 1988 *US Pat.* 4 771 295

Bartky W S 1989 *US Pat.* 4 879 568

Beals S 2004 New Advances in Inkjet CtP, Conventional inkjet systems are furthering computer-to-plate solutions. Big picture 2004-08-31, ST Media Group International Inc., Cincinnati, Ohio

Bentley P 2005 The use of inkjet to create direct write conductive features. *IS&T's NIP 21*, Baltimore, MD, USA, p. 246

Beurer G, Kretschmer J 1997 Function and performance of a shear mode piezo printhead. *IS&T's 14th International Conf. Digital Printing Technologies*, Seattle, WA, USA, pp. 621–5

Beurer G, Stempfle J, Schmidt G 2003 Function and performance of a piezo inkjet printhead with bimorph beams. *Proc. 2nd VDE World Microtechnologies Congress*, Munich, Germany, VDI-Verlag, Düsseldorf, pp. 329–34

Berdelle-Hilge P 1990 Elektrodynamische Mikrowandler zur Tropfenerzeugung. Dissertation, TU München, Germany

Boer M 2002 I.T Strategies (Hanover, MA). *Website American Printer*. October 1, 2002, ©Prism Business Media

Boer M 2006 I.T. Strategies (Hanover, MA). *Website American Printer*, October 1, 2002, ©Prism Business Media or IT Strategies 2006 Worldwide Digital Printing Market Report, Boston, Tokyo

Bolmgren J, Nilsson K 1982 *EP* 067 889

Buck R T, Cloutier F L, Erni R E, Low R N, Terry F D 1985 *US Pat*. 4 500 895 (disposable cartridge)

Cahill V, Ujiie H 2004 Digital textile printing 2004. *IS&T's NIP 20*. salt Lake City, UT, USA

Chen T 2005 Recent advances in piezoelectric inkjet print head technology. *IS&T's NIP 21*, Baltimore, MD, USA, p. 268, (Trident-ITW)

Chen T 2006 Piezoelectric ink jet print head technology for precision dispensing application. *Digital Fabrication 2006*, Denver, CO, USA, pp. 66–9, (Trident)

Corrall J 2006 Konica Minolta inkjet technology for high speed 1-pass printing. *IMI Digital Printing Conference 2006*, Barcelona, Spain

Creagh L T 2005a Novel 3-D MEMS approach to digital printing electronics. 2005 DPP, Amsterdam, The Netherlands, pp. 134–6 (Spectra)

Creagh L T 2005b Industrial inkjet printing of single layer electronics. *SGIA Membrane Switch Symposium*. Indianapolis, IN, USA, June 2005

Creagh L T, Schoeppler M W 2005 *Novel 3-D MEMS approach to digital printing electronics*. *DPP 2005*, Amsterdam, The Netherlands (Spectra)

Curie J, Curie P 1880 *Developpement par compression de l'électricité polaire dans les cristeaux hémièdres à faces inclines*. Sorbonne de Paris, France

Derbyshine K 2004 *Developments in Inkjet-Printed Electronics*. Pira Publishing, Pira International Leatherhead, Surrey, UK

Dijksman J F, van Dam D B, Duineveld P C, Rubingh J E 2005 Precision inkjet printing of polymer light emitting diode displays. *DPP 2005*, Amsterdam, The Netherlands, pp. 133, (Philips Lab)

Duby T G 2001 Performance improvements for a commercial piezo printhead. *NIP 17*, Fort Lauderdale, FL, USA, pp. 328–31

Ederer I 2005 *Voxeljet präsentiert erstmals den neuen Plastikprozess. Pressemitteilung vom*. August 27, 2005

Endo I, Sato Y, Saito S, Nagagiri T, Ohno S 1977 Liquid jet recording process and apparatus there for. *GB Pat.* 2007 162 (Canon)

Fischbeck K H 1986 *US Pat*. 4 584 590 (Spectra)

George B R, Ujiie H, Govindaraj M, Wood D, Nandedkar S, Tremere A, Fruscello M 2005 Integration of the weaving and printing processes. *IS&T's NIP 21*, Baltimore, MD, USA, pp. 239–44

Grooten M 2006 Inkjet technologies for printable electronics. *OEC 06*. Frankfurt, Germany

Gross J 2006 High speed digital printing for corrugated packaging. *IMI Digital Printing Conference 2006*. Barcelona, Spain

Hara T 1985 *US Pat.* 4 511 906 (capillarity)

Hees U, Freche M, Kluge M, Provost J, Weiser J 2003 Textile ink jet printing with low viscosity pigment inks. *IS&T's NIP 19*, New Orleans, LA, USA, pp. 626–9

Heider J 1975 *Piezokeramik als Antriebselement in einem Nadeldrucker*. Dissertaion, TU München, Germany

Heinzl J 1977 Piezoelektrisches Antriebselement für Schreibdüsen. *DP* 25 37 767

Heinzl J 2005 Drop on demand inkjet technology – Inkjet printing on corrugated card board. *FEFCO 8th Technical Seminar*, Brussels, Belgium, Nizza

Heinzl J, Hertz C H 1985 Ink-jet printing. *Adv. Electron. Electron. Phys.* **65**, 91

Heinzl J, Kattner E 1975 Piezoelektrisch betriebener Schreibkopf für Tintenmosaikschreibeinrichtungen. *DE* 2 537 767

Heinzl J, Rosenstock G 1977 Silent ink jet printing for printer terminals. *Siemens Rev.* **XLIV(9)**, 402–4

Heinzl J, Kattner E, Rosenstock G 1979 *US Pat.* 4 149 172

Heinzl J, Wehl W, Kattner E, Rosenstock G 1982 Drop-on-demand and acoustic drop shaping.In: *SID 82 Digest.,* San Diego, CA, USA, Vol. 13, p. 152

Heinzl J, Hochwind B, Pötzlberger H, Schlaag H, Steckenborn A 1990 Nach demThermalwandlerprinzip arbeitender, mikromechanisch hergestellter Tintendruckkopf. *EP* 0 530 209

Hirota A, Ishikura S 2005 Development of drop-on-demand piezoelectric line inkjet printhead. *IS&T's NIP 21*, Baltimore, MD, USA, pp. 257–63, (Brother, Kyocera)

Ikeda T 1990 *Fundamentals of Piezoelectricity*. Oxford University Press, Oxford, New York, Tokyo

Irlinger F 2007 Entwicklungsstrategie für innovative Mikrosysteme. Habil. TU München

Irlinger F 2007 *Systemverhalten, Optimierung und Sensibilitätsanalyse mit schnellen Simulationsprogrammen*. Mechatronik, F&M, Carl Hanser Verlag, München, in press

Kahlert H-J 2001 *Ultraviolette Löcher*. Innovation 9, Carl Zeiss, Oberkochen, Germany, ISBN 1431–8040, pp. 16–17

Karlinski H 1999 Ink jet print head with ink supply through porous medium. *US Pat.* 5 940 099

Karlinski H 2000 New DOD printhead for industrial applications. *IS&T's NIP 16*, Vancouver, BC, Canada, pp. 128, (Aprion Digital)

Katayama N 2004 Inkjet head. *EP* 1 559 551 A1 (Brother)

Kipphan H 2000 *Handbuch der Printmedien*. Springer-Verlag, Berlin

Kitahara T 1997 Ink jet head with multi-layer piezoelectric actuator. In Rezanka I and Eschbach R (eds.) *RecentProgress in Ink Jet Technologies*. Springfield, VA, USA, IS&T, pp. 125–8

Koewius A 2005 *Sandgießformen nach dem Digitaldruck-Prinzip*. Aluminium 81, Giesel-Verlag GmbH, Isernhagen, pp. 3–10 (Prometal RCT, Voxeljet Technology)

Kyser E L 1981 *Drive pulse optimization*. In: Gaynor J (ed.) *Advances in Non-Impact Printing Technologies for Computer and Office Applications*, Van Nostrand Reinhold Company, New York, NY, USA, pp. 1175–89

Kyser E L, Sears S B 1976 Method and apparatus for recording with writing fluids and drop projection means therefore. *US Pat.* 3 946 398 (Silonic)

Larsson S G 1983 *US Pat*. 4 367 478 (Xerox, push piston)

Le H P 1998 Progress and trends in ink-jet printing technology. *J. Imaging Sci. Technol.* **42(1)**, 49–62

Le H 2005 Advances in ink jet printhead development for digital textile applications. *DPP 2005*, Amsterdam, The Netherlands, pp. 159–60, (PicoJet)
Lemmermeyer B 2006 Ein hochtemperaturbeständiger Einzeltropfenerzeuger für flüssige Metalle. Dissertation, TU München, Germany
Masakawa Y 2006 New web printer OP-1 with fixed head arrays for industrial applications. *IMI Digital Printing Conference 2006*, Barcelona, Spain
McDonald M 2003 Manufacture of flat panel displays using piezoelectric drop-on-demand ink jet. *SID Digest 2003, Session 39.3*, Baltimore, MD, USA
McDonald M, Zhou Y 1999 Crosstalk study of a high speed shear mode piezo ink jet printhead. *IS&T's NIP 15*, Orlando, FL, USA, pp. 40–3
McWilliams A 2006 Micromechanical systems (MEMS) technology: Current and future markets. *BCC Research Report SMC051B*, pp. 63–70, 143–145. BCC Market Research Reports and Technical Publications, Norwalk, CT, USA
Menzel C, Hoisington P, Bibl A 2004 MEMS solutions for precision micro-fluidic dispensing application. *IS&T's NIP 20; 2004 Int. Conf. Digital Printing Technologies*, Salt Lake City, UT, USA, pp. 169–75
Michaelis A J, Paton A D, Temple S, Bartky W S 1989 *US Pat.* 4 887 100 (Xaar)
Nespaca M 2004 From DReAM to dollar. *IS&T's NIP 20*, Salt Lake City, UT, USA, p. 611
Nexus 2005 *Nexus Market Analysis for MEMS and Microsystem III, 2005–2009*, WTC Wicht Technology Consulting, Munich, Germany
Nielson N J, Bolmgren J 1986 *US Pat.* 4 564 851(Kamm)
Pelzer R 2005 Auslegung von besiedelbaren Knochenimplantaten und deren Herstellung mittels 3D-Druckverfahren. Dissertation, TU München, Germany, Fortschritt-Berichte VDI 17/255, VDI Verlag, Düsseldorf
Petschik B 1994 Wiedergabe farbiger Halbtonbilder mit Non-Impact-Druckern. Dissertation, TU München, Germany, Fortschritt-Berichte VDI 10/288, VDI Verlag, Düsseldorf
PixDro Ind 2006 P350 Industrial inkjet printer for plastic electronics. *Datasheet OEC 06 Organic Electronics Conference and Exhibition 2006*. Frankfurt, Germany
Pöppel J 1990 *Sensor- und Aktoreigenschaften von Bubble-Jet-Heizelementen in Tintenschreibwerken*. Dissertation, TU München, Germany
Rayleigh F R S 1878 On the instability of jets. *Proc. London Math. Soc* **10(4)**, 4–13
Reid B 2005 *Product Overview – Z Corporation 3D Printers.* November 30, 2005. Z-Corporation, Burlington, MA, USA
Reiländer U 2003 Das Großsignalverhalten piezoelektrischer Aktoren. Dissertation, TU München, VDI 8 / 985, VDI Verlag, Düsseldorf
Runge W 1993 Berechnungsmodell thermischer Tintenschreibwerke. Dissertation, TU München, VDI 1/219
Sakaida A, Shinkai Y, Asano T, Hirota A 2003 Ink-jet head and its manufacture method, ink-jet printer and method for manufacturing. *EP* 1 717 034 A2 (Brother)
Scheicher R 2004 Kompaktmodell zur Systemsimulation eines Tropfenerzeugers mit Piezobiegewandlern. Dissertation, TU München, Germany, Fortschritt-Berichte VDI 8/1053, VDI Verlag, Düsseldorf
Scheppler M 2006 Simple tools for ink jet printing on flexible substrates. *USDC's 5th Annu. Flexible Displays Microelectron. Conf*. Phoenix, AZ, USA
Schindler M 2005 Produktionsnahe Qualitätssicherung für Piezobiegeaktoren. Disseration, TU München, Germany, Fortschritt-Berichte VDI 9/375,VDI Verlag, Düsseldorf
Schwartz N R 2002 Ink jet printing in industrial and commercial applications. *IS&T's NIP 18*, San Diego, CA, USA, pp. 118–21
Seitz H 2002 *Modellierung und numerische Simulation des Übersprechverhaltens in einem Tropfenerzeuger mit Piezo-Biegewandleraktoren.* Dissertation, TU München, Germany.
Seitz H 2003 Modeling and numerical simulation of the crosstalk behavior of a DOD printhead. *IS&T's NIP 19*, New Orleans, LA, USA, pp. 343–7
Springford C 2004 Development in Inkjet-Printed Packaging. *Pira Publishing, Pira International*, Leatherhead, Surrey, UK
Sumerel J 2005 Digital printing of bioinks. *SE-SW Regional Meeting of the American Chemical Society*, Memphis, TN, USA, November 2005
Torrey M K 2005 Impact of silicon MEMS on future ink jet printhead architectures. *DPP 2005*, Amsterdam, The Netherlands, pp. 146–7 (Spectra)
Tremblay M R 2005 Realizing potential of color in 3D printing. *28th RP Symposium of Japan*, Tokyo, Japan, June 21 (Z Corporation)
Usui M, Koto H, Nakamura H, Shimada Y, Abe T 1991 Drop-on-demand ink-jet printing head. *EP* 1 208 983
Van Daele J 2005 Inkjet as change agent for printing industry: a vision of digital printwork, DPP 2005. *Int. Conf. on Digital Production Printing and Industrial Application*, Amsterdam, The Netherlands
Vaught J L, Cloutier F L, Donald D K, Meyer J D, Tacklind C A, Taub HH 1984 Thermal ink-jet printer. *US Pat.* 4 490 728 (Hewlett-Packard)
Voigt P 2002 Compact modeling of microsystems. Dissertation, TU München, Germany
Wehl W 1984 Akustik und Fluidmechanik in Kanälen und Düsen von Tintenschreibwerken. Dissertation, TU München, Germany
Wehl W 1995 *Tintendrucktechnologie: Paradigma und Motor der Mikrosystemtechnik.* Carl-Hanser Verlag, Munich, F&M, 103, pp. 318–24 and 486–91
Wild J 1990 *Systematik und Bewertung nichtadaptiver Raster zur Farb- und Halbtondarstellung.* Dissertation, TU München, Germany
Willis M 2006 *The wonderful world of ink jet. IMI Digital Printing Conference 2006*, Barcelona, Spain
Wohlers T 2006 *Wolters Report 2004, Rapid Prototyping, Tooling and Manufacturing State of the Industry*. Wohlers Association, Fort Collins, CO, USA
Yu J-D, Sakai S 2003 Piezo ink jet simulations using the finite difference level set method and equivalent circuit. *IS&T's NIP 19*, New Orleans, LA, USA, pp. 319–22
Zeilhofer H F 1998 *RP-Modelle für die Operationsplanung in der Mund-,Kiefer- und Gesichtschirurgie*. Habil, TU München, Germany
Zhang J 2005 New developments in Epson's inkjet head technology. *IS&T's NIP 21*, Baltimore, MD, USA, pp. 269–72
Zhou Y 1997 A coupled field analysis of an ink jet printhead with shear mode piezo transducer. *IS&T's NIP 13*, Seattle, WA, USA, pp. 809–12 (Spectra)
Zhou Y 2000 The application of ink jet technology to high-speed DOD single pass printing. *IS&T's NIP 16*, Vancouver, BC, Canada, pp. 28–31
Zhou Y 2004 Recent advances in highly durable piezoelectric inkjet print head technology. *IS&T's NIP 20*, Salt Lake City, UT, USA, pp. 855–8
Zhou Y, Rensch C 2001 Applications of page wide piezo ink jet printing to commercial and industrial market. *DPP 2001: Eye on the Future*, Antwerp, Belgium, pp. 50–3
Zollner A 1995 Design aspects for a smart thermal inkjet printhead. *Proc. IS&T/SPIE 95*, San Jose 24, CA, USA, pp. 13–24
Zoltan S L 1974 Pulse droplet ejection system. *US Pat*. 3 683 212 (Clevite)

Biography

Born in 1940, Joachim Heinzl received his diploma in mechanical engineering in 1965 and his doctoral degree in 1970 from the Technische Universität Munich, Germany. In 1968 he joined the Central Laboratory of Siemens AG. Among other projects he conceived ink-jet printing in vacuum technology driven by piezotubes and developed it right to its line production as Printer Terminal PT 80. He held the Chair for Feingerätebau und Mikrotechnik (Precision-Engineering and Micro Technology) from 1978 until 2005. He focused on the fields of research of aerostatic bearings and ultra-precision technology, microactuators, dispensing technologies, rapid prototyping for microsystems, laser metrology, and laser processing. From 1995 to 2002 he was the First Vice President of the TUM, and at present is President of the Bayerische Forschungsstiftung (Bavarian Research Foundation). Professor Heinzl has published more than 50 papers as an author or coauthor. He has filed 90 patents at the German Patent Office as an inventor or coinventor, 44 of them have been filed internationally. In 2003 he was awarded the Aachener and Münchener Prize in recognition of his pioneering research and inventions that have led to the introduction of ink-jet printing worldwide. In 2004 he became member of Acatech (the Convent of Engineering Sciences at the union of German Academies of Sciences). In 2006 he received a honorary doctorate by the Universität Hannover in recognition of his achievements in national and international cooperations. In 2007 he was appointed Foreign Associate of the National Academy of Engineering, Washington, DC, for contributions to the worldwide introduction and the use of DOD ink-jet printers.

3.12 Automotive Applications

Nobuaki Kawahara, Research Laboratories, Denso Corporation, Aichi, Japan

3.12.1 Automotive Control Systems

3.12.1.1 Introduction to Automotive Control Systems

The basic function of automobiles is to transport passengers and objects; however, they are now expected to satisfy various additional demands. Besides meeting such demands that include ever-diversifying user needs, improvements in safety features, and environmental friendliness, automobiles are required to improve their basic functions, such as driving, stopping, and turning, to be more convenient, to have enhanced safety characteristics in the case of accidents, to preclude traffic accidents, to help relieve traffic congestion, to be more energy-efficient, and to realize clean gas emissions. To satisfy these demands, automobiles are equipped with various newly developed control systems. These automotive control systems are classified into four categories:

(1) Powertrain control systems
(2) Vehicle-body control systems
(3) Vehicle control systems (for driving and safety)
(4) Information and telecommunications control systems

Conventionally control systems were mechanical; however, with the recent advances in electronic

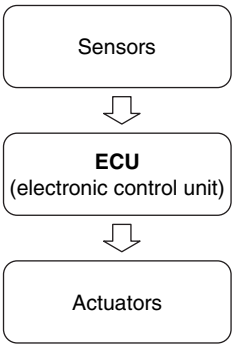


Figure 1 Structure of control systems.

technologies, the performance of these systems has been improving, providing innovative functions. Today, a control system basically consists of sensors, an electronic control unit (ECU), and actuators, as shown in Figure 1. The sensors detect many conditions and send information to the ECU. On the basis of the delivered information, the ECU performs computation to place the function into the best condition and controls the actuators. In the figure, the arrows show the direction of flow of information.

Microsystem technology has contributed to improved performance, reliability, and a lower cost for sensors. Future goals of the automobile industry must be the development of environmentally friendly cars that are capable of obtaining higher gas mileage and are safer and more comfortable. Also, these cars should be capable of receiving and transmitting the necessary information. The automotive industry is currently focusing its R&D on ultralow emission vehicles and intelligent transport systems (ITS). Microsystem technology is expected to play an extremely important role in future R&D in the automotive field, particularly in achieving higher levels of safety. The integration of microsystem technology in various devices is essential so that the system maintains a high level of reliability and mass productivity. This section discusses how microsystem technologies are used in current vehicles and will be used in future vehicles.

3.12.1.2 Microsystem Technology for Automotive Control Systems

Microsystem technology has made major progress in the past few years because of the advancement in miniature machining techniques, drastic improvement in performance due to miniaturization, and realization of new functions that could be attained only by miniaturization. With progress in electronics, microelectromechanical systems (MEMS) sensors have replaced mechanical sensors in many applications. MEMS sensors have the following advantages:

(1) Deterioration-free and durable for a long period
(2) Low cost (suitable for mass production)
(3) Small in size and easy for installation
(4) Good dynamic characteristics (high response and sensitivity)
(5) Superior impact resistance
(6) Low power consumption
(7) Self-diagnosis function

Due to these overwhelming merits, use of MEMS sensors has been naturally increasing in scale. These sensors permit sophisticated control, contributing to the improvement in the basic performance of various kinds of equipment. MEMS sensors are commonly made from silicon, which is better than mechanical sensors, due not only to the superior physical properties of semiconductor silicon but also to their compactness and light weight (essential requirements for automotive parts); in addition they have high environmental resistance and reliability.

Table 1 compares the physical properties of iron and silicon as the substrate material for MEMS sensors. Silicon is equivalent to steel in its physical properties, is precisely machined into the desired shape by semiconductor processes, has various effects for converting change in physical properties to electrical signals (piezoresistance effect, photoelectric effect, Hall effect, etc.), and enables integration in electrical circuits.

3.12.1.3 Automotive Sensors Used in Current Vehicles

The survey by the Japan Electronics and Information Technology Industries Association reveals that approximately 310 million automotive sensors were

Table 1 Physical properties of silicon and iron

	Si	*Fe*
Density ($g\,cm^{-3}$)	2.33	7.86
Young's modulus (GPa)	190(111), 130(100)	210
Melt point (°C)	1412	1534
Specific heat ($J\,g^{-1}\,K^{-1}$)	0.76	0.64
Thermal expansion coefficient ($\times 10^{-6}\,K^{-1}$)	2.6	11.8
Heat conduction ($W\,m^{-1}\,K^{-1}$)	168	48

produced in Japan during 2002, and approximately 10.26 million motor vehicles in the same year. This means that on an average about 30 sensors are installed in each vehicle. It is difficult to obtain accurate figures since the specific number of sensors installed per vehicle differs depending on the vehicle model, e.g., a recent luxury car contains approximately 100 sensors of about 70 types (Fujikawa 2004).

Table 2 lists typical automotive sensors used in current vehicles. As for the basic control functions of drive, stop, and turn, they are the largest in number, of which many are of the semiconductor/MEMS type. The detection parameters are temperature, pressure, acceleration, position/angle, rotation, flow, liquid level, vibration, and gas. The number of MEMS sensors is increasing due to the advantages mentioned in the table. In the table, the bold-faced type is a MEMS sensor.

Table 3 compares the specification for sensors used in different applications. Automotive sensors are used under severe environmental conditions: a wide temperature range, bumpy conditions, an unstable power supply, and a strong electromagnetic field.

3.12.1.4 Engine Control Systems for Automobiles

This section describes how the above-mentioned sensors are used in actual automobiles and their functions. The engine control system controls the engine power, gas emissions, energy efficiency, and other engine conditions to enable optimal operation conditions. Electronic engine control emerged in the 1970s and became common in fuel injection systems (Betta *et al.* 2004, Hirooka *et al.* 2004, Kishi *et al.* 1998, Nishida *et al.* 2005, Nishizawa *et al.* 1997, Shimoda *et al.* 2006). Recently, electronic engine control has evolved to include control of the fuel ignition system and subsystems, such as the ignition system and electronic throttle control system, sharing sensors for a single ECU.

A gasoline engine control system consists of a fuel injection system, an ignition system, and a diagnostic

Table 2 Automotive Sensors

	Application		
	Drive, stop, and turn	*Safety, conformity, and environment*	*Information/communication*
Measurement Temperature	Water temperature sensor, intake air temperature sensor exhaust temperature sensor	Oil temperature sensor, room temperature sensor, air temperature sensor, refrigerant temperature sensor	
Pressure	**Intake air pressure**, **fuel tank pressure**, **atmospheric pressure**, **combustion pressure**	**Air-conditioner coolant**, **pressure sensor**, **brake pressure sensor**	
Acceleration		**Acceleration sensor for air bag**, **ABS**, **VSC**	
Position/angle	Throttle sensor, **steering angle sensor**	**Car height sensor**	
Rotation	**Cam angle sensor**, **Crank angle sensor**, **Vehicle speed sensor**	**Wheel speed sensor**	
Flow	**Airflow sensor**		
Level	Fuel sensor, Engine oil sensor	Brake oil sensor	
Vibration	Knock sensor		
Gas	O_2 sensor, A/F sensor	Smoke sensor, CO sensor, humidity sensor	
Light/electromagnetic wave		**Solar sensor**, **autolight sensor**, clearance sonar	**Laser radar**, **millimeter wave radar**, **satellite communication antenna**, **infrared camera**, **car recognition camera**, **other cameras**
Other		**Gyroscope**	Geomagnetic sensor, magnetic nail

Table 3 Comparison of Specification for Sensors

	System			
	Automobile	*Household appliance*	*Measuring instrument*	*Airplane*
Accuracy	1–5%	~5%	0.1–1%	0.1–1%
Operation temperature	−40–120°C	−10–50°C	−0–40°C	−55–70°C
Impact resistance	2–25 G	~5 G	~1 G	0.1–10 G
Fluctuation of battery voltage	±50%	±10%	±10%	±10%
Electromagnetic environment	Bad	Good	Good	Good
Price	$1~10	$1~10	$10~100	$10~100

monitoring system as major functions, complete with subordinate functions including an electronic throttle control system, a variable cam timing system, and a knock control system (KCS). In actual operation, a gasoline engine control system measures the amount of air intake into the engine using an air flow sensor and the crank angle using a rotation angle sensor, determines whether the engine is free from knocking via a knock sensor, and examines exhaust gas conditions using an O_2 sensor, controlling the rate of fuel injection, timing of injection, timing of ignition, and timing of valve opening and closing, as shown in Figure 2. The controlled functions are thus adjusted to maintain engine power, emissions, and noise under optimal conditions. Today, fuel-efficient direct injection gasoline engines are mass-produced. For use with these engines, systems that control the direct fuel injection into cylinders are emerging. Figure 3 shows sensors used in a gasoline engine control system.

3.12.1.4.1 Fuel injection system

A gasoline engine supplies a mixture of air and fuel (gasoline) into cylinders, produces combustion, and obtains power from the thermal energy generated by combustion. The fuel injection system measures the amount of air intake into the engine and supplies the amount of fuel required for combustion. The fuel injection rate is controlled by an ECU, which calculates the amount of fuel to be injected based on the signals from the sensor that detects the amount of air intake and controls the length of time to keep the fuel injection valves (injectors) open.

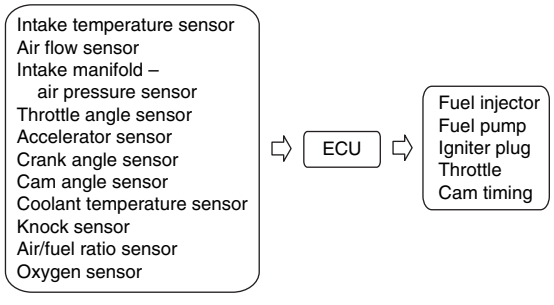


Figure 2 Structure of engine control.

There are two representative methods for detecting the amount of air intake: the mass flow method and the calculation method. The mass flow method directly measures the amount of air intake into the engine by means of an air flow meter. The thermal air flow meter, which is a typical sensor that uses the mass flow method, is able to measure the mass of air accurately even when the amount of air intake varies due to changes in air pressure and/or temperature or due to exhaust gas recirculation (EGR). The calculation method estimates the amount of air intake based on the speed of engine rotation, air temperature, and the pressure detected at the air intake pipe by the intake air pressure sensor, i.e., the manifold absolute pressure sensor (MAPS). Nevertheless, this method can measure the amount of air intake into the engine with superior response characteristics even during a period of transition, because the intake air pressure sensor is placed downstream from the throttle.

3.12.1.4.2 Air–fuel ratio feedback control

The ideal ratio by mass of air to fuel for complete combustion is 14.7:1. The lambda (λ) value (an alternative way to represent air–fuel (A/F) ratio) is 1, which is known as stoichiometric ratio. When $\lambda > 1$, i.e., when air is excessive, NO_x gases increase. When $\lambda < 1$, i.e., when air is insufficient, engine gases HC and CO, which are hazardous substances, increase in the exhaust gas. To reduce the hazardous substances (HC, CO, and NO_x), emissions are cleaned by a posttreatment process using a catalyst. Among the various catalysts that are available for this process, three-way catalysts are the most common today. A three-way catalyst simultaneously performs the oxidation reaction of CO and HC and the reduction reaction of NO_x, converting these substances into harmless substances (CO_2, H_2O,

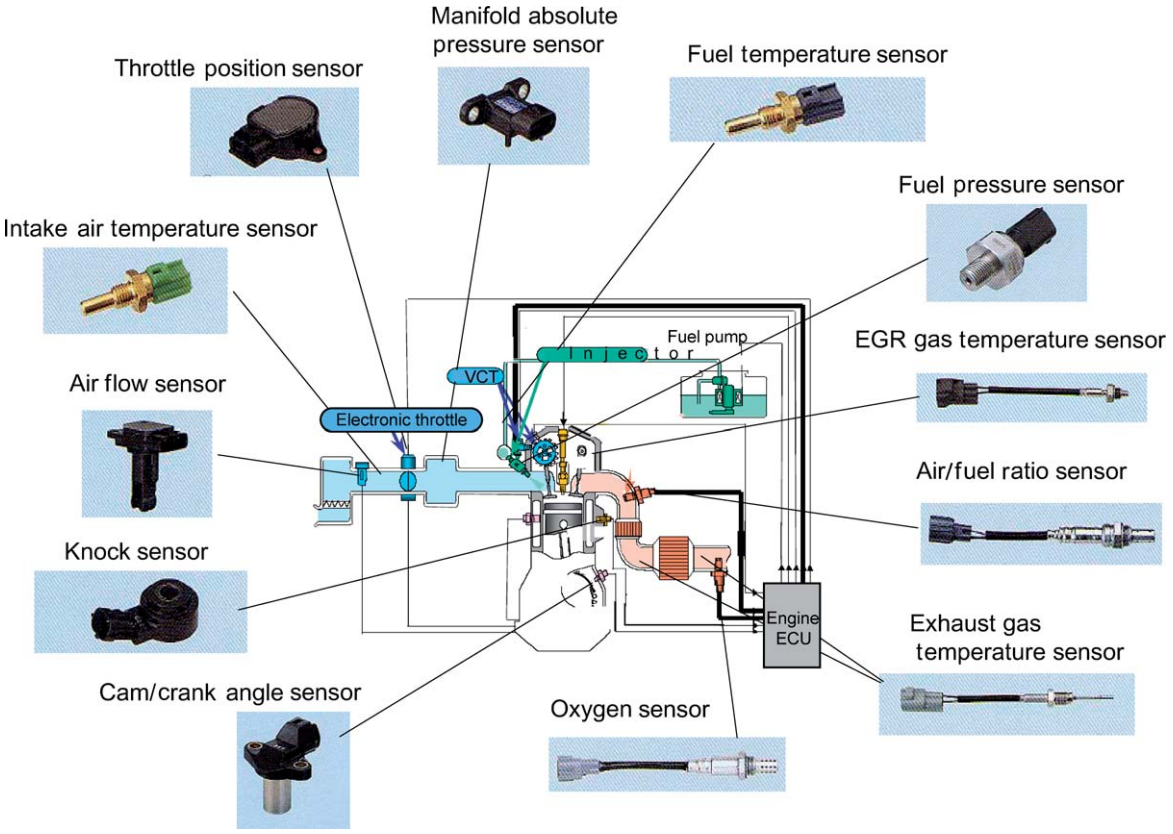


Figure 3 Sensors for engine control.

and N_2). A three-way catalyst achieves an extremely high cleaning rate when λ is controlled within a narrow range of the stoichiometric ratio. To retain the A/F ratio at $\lambda = 1$ under varying driving conditions, the A/F ratio feedback control is used. This control obtains A/F ratio information from an O_2 sensor installed at the exhaust pipe of the engine and corrects the fuel injection rate based on the signals received from the sensor.

3.12.1.4.3 Electronic spark advance system

The electronic spark advance (ESA) system calculates duration for keeping the electric power on and the timing of ignition, and outputs an ignition signal depending on the crank angle. The ESA system detects the angular position of each cylinder based on the signal of the crank angle sensor. The ECU calculates the optimal ignition timing depending on the driving conditions and sends an ignition signal to the igniter. The igniter produces high-voltage electricity in response to the ignition signal, applies it to the electrode of the spark plug, produces a spark, and burns the A/F mixture in the cylinder.

3.12.1.4.4 Knock control system

The KCS has a knock sensor installed at the cylinder block to detect engine knocking, which represents anomaly burning, as mechanical vibration. When the sensor detects vibration, it outputs electric signals. The output signal passes through a filter, is converted into voltage by an analog–digital converter, and is detected in the form of voltage. The KCS determines whether the engine is knocking or not based on the difference in the voltage level between when the engine is knocking and when it is not. Accordingly, the KCS delays or advances the timing of ignition. Because a single knock sensor is used to check multiple cylinders for knocking, the sensor is positioned at an optimal location.

3.12.1.4.5 Electronic throttle control system

The electronic throttle control system determines the throttle opening by means of the engine control computer (i.e., ECU) based on the degree to which the driver depresses the accelerator pedal (i.e., according to the accelerator opening) and

controls the throttle valve to the determined opening. The control system determines the amount of air intake into the engine by controlling the throttle opening. The throttle valve is operated by means of a motor based on the signals from the ECU, and the data of throttle opening are fed back to the ECU by the throttle sensor. The electronic throttle control system can directly control engine output (the amount of air intake) by controlling the throttle opening. The system takes advantage of this feature, and, accordingly, has many functions, such as engine power control based on accelerator opening (determining the throttle opening by considering acceleration based on the accelerator opening specified by the driver), idling speed control (control of throttle valve opening to realize the ideal idling speed, which is determined in advance based on the engine's driving conditions, such as warm-up conditions and accessories' load-carrying conditions), control of torque produced at the time of automatic transmission gear change (engine power control to reduce shock or an abrupt change of torque at the time of gear change), maximum speed limit control (controlling vehicle speed increase beyond a previously set speed by closing the throttle valve to lower engine power), cruise control (controlling engine power to enable the vehicle to continue at a constant speed after the release of the accelerator pedal), traction control (engine power reduction for preventing the slipping of driving wheels), and vehicle stability control (VSC) (engine power reduction for preventing the slipping of driving wheels).

3.12.1.4.6 Diagnosis system

Vehicles are expected to have increasingly sophisticated failure resistance and safety performance; therefore accurate diagnosis is becoming increasingly important. The purpose of the diagnosis is to ensure vehicle safety and to prevent the emission of gas whose quality has deteriorated because of a failure. A diagnosis system detects a fuel leak to preclude a vehicle fire and indicates a deteriorated catalyst, thus preventing aggravated gas emissions. The diagnosis system has a computer (ECU) on the vehicle that self-diagnoses the vehicle in real time. The system checks that the catalyst has not deteriorated in cleaning performance; how frequently an ignition failure occurs, such as the firing failure of the A/F mixture; in what cylinder or cylinders such an ignition failure occurs; that the O_2 sensor has not deteriorated or is functionally abnormal; that the feedback of the A/F ratio is not hindered for any reason including low fuel pressure; that other sensors and actuators are normal; and that any circuit is not disconnected or short-circuited.

3.12.1.5 Air-Conditioning System

There are various onboard heating, ventilating, and air-conditioning (HVAC) systems, such as heater, cooler, manual air conditioner, and automatic air conditioner. This section focuses on the automatic air conditioner, which enables the most advanced control. Today's automatic air conditioner uses a microcomputer, and a passenger can obtain a more comfortable environment by a simple operation, while the conditioner actually performs all highly complicated adjustments (Kataoka *et al.* 2005, Sugi and Yamanaka 2005).

The basic functions of the automatic air conditioner are cabin temperature control (output air temperature control), fan speed control, air distribution control, fresh or recirculated air control, and humidity control (dehumidifies the air for demist). To realize these functions, a number of sensors are used to provide input to the ECU, as shown in **Figure 4**.

3.12.1.5.1 Sensors in air-conditioning system

The temperature sensors include an ambient temperature sensor, a cabin temperature sensor, an evaporator temperature sensor, and a water temperature sensor. They are usually thermistor devices. The

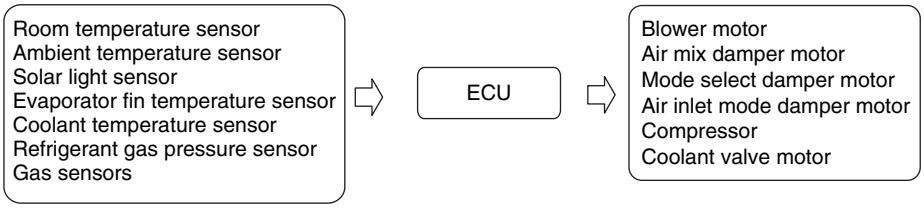


Figure 4 Structure of air-conditioner system.

cabin temperature sensor is used for detecting the average air temperature in the cabin and for allowing an accurate reading; a small fan is used to lead sample air to the sensor. The evaporator sensor is used for detecting the cooling capacity. The water temperature sensor is installed at the engine cooling water outlet for detecting the maximum heating capacity. It is also used for warm-up control to prevent the output of cold air when the engine cooling water, which is used as the heat source at the time of starting up the engine in winter, is not sufficiently warm. The intake air is cooled by means of an evaporator. The blower then adjusts the air volume, and the air mix damper adjusts the ratio of the air that is allowed through the heater core and the air that is not to control air temperature. The heater–cooler switch damper controls the air output opening. To raise the temperature in the cabin area, the air is output from the opening at the passenger foot level; to lower the temperature, the air is output from the opening at the passenger face level. In addition, the differential switch damper uses the air conditioner to ensure clear vision.

A solar light sensor that consists of a segmented photodiode and a window is located on a dashboard and detects the direction and strength of sunlight. When the sunlight shines directly on the driver, the ECU tries to increase the cooling capacity on the driver's side, i.e., tries to increase the air to the driver's face vents and to lower the temperature of the wind. This requires independent controls for the driver's side and for the passenger's side. Control systems that allow the temperature of the air conditioner to be set separately between the driver's seat and the passenger's seat are becoming common.

Gas sensors detecting CO and HC are located outside of the cabin. When these gases are detected and fresh air is introduced into the cabin, the ECU switches fresh/recirculation damper. An air quality sensor that detects diesel exhaust fumes emitted from trucks and the like is installed at the outside air intake of the air conditioner, and when exhaust fumes are detected, the fresh/recirculation switching damper is activated, shifting the fresh air mode to the recirculation mode (Honda *et al.* 2001). New systems that control humidity besides controlling temperature are also emerging. They improve the thermal comfort in the cabin while improving energy efficiency.

The refrigeration cycle of a vehicle air conditioner consists of a compressor, a condenser, an expansion valve, and an evaporator. Refrigerant is circulated in this sealed circuit (cooling cycle), thus transferring heat from the low-heat source (air in the cabin) to the high-heat source (outside air). This mechanism is called vapor compression refrigeration system, which is the most common refrigeration system used in many applications, such as room air conditioners and compact refrigerators. As part of compressor control, this system has the function of detecting excessively high or low pressure by means of the refrigerant pressure sensor and by turning off the compressor. To ensure comfort, the compressor of an automatic air conditioner is always on. Some automatic air conditioners automatically turn off the compressor when the cooling function is not in use, realizing energy saving and economical compressor control.

3.12.1.6 Air Bag System

Besides seat belts, air bags are the most common passenger protection equipment. Air bags are stored in the steering wheel and in the instrument panel. They instantaneously inflate at the time of a frontal collision, offering additional protection for passengers in combination with seat belts.

An air bag consists of a sensor that detects a frontal collision, an ECU that decides whether to deploy the air bag or not based on the signal from the sensor, an inflator that produces gas, as shown in **Figure 5**, and a bag that inflates to protect the passenger and the driver. When the vehicle decelerates due to a collision, the sensor detects the deceleration. This sensor is an electronic accelerometer. The ECU decides whether to inflate the air bag or not based on the waveform obtained at the time of the collision. A sensor may detect vehicle acceleration whose

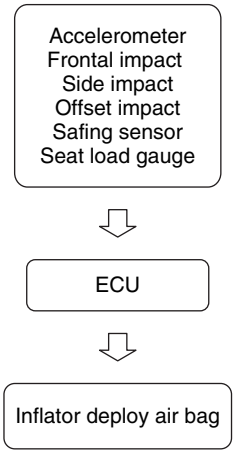


Figure 5 Structure of air bag system.

impact compares to that of a collision, such as when the vehicle collides against a curb, when the vehicle is trapped by a hole on the road, or when the vehicle is hit by a hammer in a service shop. To prevent erroneous air bag inflation in these cases, whether to inflate the bag or not is decided based on the waveform of the signal instead of the threshold value of a sensor signal. It prevents the erroneous inflation of the air bag because this may block the driver's vision and induce panic, possibly causing an accident. The waveform used to make this decision at the time of a collision is determined individually for different vehicle models, because the waveform is influenced by body strength and equipment.

When the ECU decides that an air bag should be inflated, ignition current is supplied to the inflator, the igniter ignites the gas generator material, and the generated nitrogen gas inflates the bag. This process from collision to bag inflation should be completed in tens of milliseconds before the driver's upper body hits the steering wheel. After the air bag has been inflated fully, gas is slightly removed to reduce surface pressure to absorb the shock that the driver and the passenger receives. In case the ECU is not supplied with power by the battery due to the disconnection resulting from the collision, the ECU is provided with a backup capacitor.

Safety features are increasingly in high demand in the market nowadays, and to satisfy such demands, various air bags have been developed in addition to the air bags for frontal collisions (Kamiji 2003). For side collisions, side air bags and curtain air bags are available. A side air bag protects the chest of the passenger, while the curtain air bag protects the head. The side air bag is inflated from the side of the seat, and the curtain air bag is inflated between the side window and the head of the passenger in the front seat (Summers *et al.* 2005). Curtain air bags for rear seats are also in practical use. Knee air bags, which inflate from below the instrument panel to protect the passenger's legs, are also available.

As for accidents involving children seated in the front passenger seat, the child may become injured due to bag inflation pressure when the child is not restrained by a seat belt, when the child's seat is not correctly installed, or in other situations. In some cases the air bag has blown away the child and the child's seat. As an improvement measure to prevent such accidents, some air bags are designed to inflate more slowly when a collision occurs while the vehicle is being driven at a low or medium speed. As another measure, the air bag is designed not to inflate when a child or a child's seat is detected. Among the various methods that have been publicized to detect a child or a child's seat, the method of providing the seats with a load meter and measuring the weight to distinguish a child from an adult is the most common. It is expected that the cabin will be provided with a camera or another sensor in the future, to enable more delicate control.

3.12.1.7 Antilock Brake System

The antilock brake system (ABS) ensures the vehicle's steering control, shortens the braking distance, and prevents abnormal tire wear by preventing the wheels from locking during braking (Hou and Zhuoping 2003). When the brakes are applied while the vehicle is in motion, the tire's rotation speed (wheel speed) decreases, decelerating the vehicle. Normally there is no difference between the vehicle speed and the wheel speed, but on a slippery road surface, a difference may be generated. This phenomenon is called a slip, and the value that represents the size of the slip is called slip ratio, which is calculated by dividing the difference between the vehicle speed and the wheel speed by the vehicle speed, namely ((vehicle speed – wheel speed)/vehicle speed). The ABS controls the brakes to maximize the friction coefficient using the slip ratio.

An ABS consists of wheel speed sensors provided for individual wheels, an ECU that obtains the slip ratio and calculates the signals that will adjust the slip ratio to an adequate value, and an actuator that sets the brake pressure based on the signals from the ECU, as shown in **Figure 6**. Wheel speed sensors are normally magnetic sensors. A sensor with a built-in magnet is placed in opposition to the teeth of the rotor of the axle, detecting the projections and depressions of the rotor teeth. Rotation speed can be obtained depending on the numbers of the

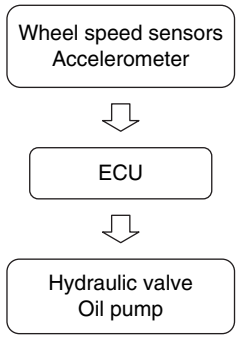


Figure 6 Structure of ABS.

projections and depressions detected during a unit period of time. The ECU calculates a virtual vehicle speed, namely, an ideal deceleration pattern, based on the signals received from wheel speed sensors. Because the simultaneous locking of all four wheels should be prevented, some systems have an accelerometer to correct vehicle speed by measuring deceleration. Because each wheel is provided with a wheel speed sensor, the slip ratio of each wheel can be obtained separately. The ECU sends signals to the actuators to realize the slip ratio that will maximize the friction coefficient. More specifically, if the slip ratio is large, the ECU sends signals to reduce the hydraulic pressure of the brakes, lowering the slip ratio of wheels. If the slip ratio does not lower sufficiently because of the residual hydraulic pressure of the brakes, the ECU sends rapid deceleration signals. When the wheel speed has lowered almost equal to the vehicle speed and the slip ratio has also sufficiently lowered, then the ECU sends pressure-increasing signals. Simply speaking, the ABS performs pumping braking at high speed in the place of the driver. The actuator consists of a valve controlling hydraulic pressure, a reservoir, a pump, and accessories. The ABS is becoming common, as an increasing number of people are becoming conscious about safety (Lu and Chan 2004). An ABS is provided in almost 100% of vehicles and is widely recognized as a common safety feature.

3.12.1.8 Stability Control

In addition to the ABS, which ensures stability when the vehicle is under braking, a traction control system (TRC) is the system that is used mainly to ensure stability when the vehicle is under acceleration after the accelerator has been applied. These are VSC systems in the longitudinal direction. These days, there is a stability control system that ensures vehicle stability while it is cornering via the steering operation (Aga 2004). Normally, a vehicle makes a turn stably at a curve as the driver steers the wheel, but when the road surface is slippery, when the vehicle speed is extremely high, or when the vehicle is making a sharp turn, the front wheels or the rear wheels may skid. In such a case, the stability control automatically controls the engine power and brakes, relieving the skid and ensuring stability. Unlike the ABS, the stability control system does not have a single designation, and can be called a VSC system, a vehicle dynamics control (VDC), or an electronic stability program (ESP) (Liebermann 2004). These systems are developed into a more sophisticated system, which controls complicated vehicle dynamics and realizes safer driving.

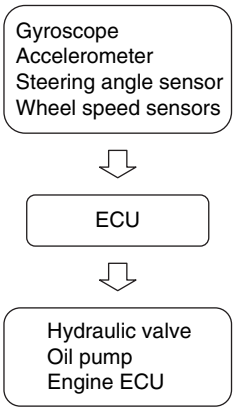


Figure 7 Structure of VSC.

The VSC system consists of a gyroscope, an accelerometer, a steering angle sensor, a wheel speed sensor, an ECU, and a brake actuator as shown in Figure 7. The vehicles skid on a curve in two modes: vehicle spin and drift. The ECU determines whether the skid occurs based on the signals from the sensors in either mode. The ECU initially calculates the angular velocity of the vehicle under no-skid conditions and the lateral acceleration generated by a centrifugal force and applied to the vehicle, based on the signals received from the wheel speed sensor and steering angle sensors. The ECU then determines whether spinning and/or drifting occurs or not, receiving the signals of the vehicle's actual angular velocity from the gyroscope and the signals of the centrifugal force the vehicle is actually exposed to via the accelerometer. When the vehicle has a strong tendency to spin, the system applies the brakes to the front and rear wheels running on the outside of the corner, producing yaw moment toward the outside of the vehicle, thus depressing spinning. When the vehicle drifts, the system limits engine power and applies the brakes to the front and rear wheels to slow down the vehicle, restoring grip to the front wheels. The construction of the brake actuator is almost the same as that of the ABS. The VSC changes the balance of the braking force between the right-hand side and left-hand side. According to some literature, the VSC is highly effective in reducing the number of serious accidents. Apart from air bags and ABSs, VSCs are expected to become a standard safety component in all vehicles.

3.12.1.9 Adaptive Cruise Control System

The cruise control is a system that maintains vehicle speed at a given level. The adaptive cruise control (ACC) system enables the vehicle to maintain a certain distance from the vehicle directly in front. This system is a representative system that will realize the automatic operation of vehicles in the future.

The ACC consists of a radar sensor that measures the distance between the ACC-built vehicle and the vehicle in front, the wheel speed sensor that determines the speed of the ACC-built vehicle, and the ECU. The ACC controls the engine, transmission, and brakes, but the ACC does not have its own actuator and so it controls them indirectly, sending signals to the ECUs of each component as shown in **Figure 8**. The radar sensor of the ACC can be a laser or a microwave radar (Furui *et al.* 1998, Tokoro *et al.* 2003). The measurement principle of laser radar is that the radar emits a laser beam from the front of the vehicle, and the optical reflection from the object in front is received by a photodiode. The time of flight (TOF) of the laser beam is measured to determine the distance between vehicles. The laser beam has a wavelength in the near-infrared region and is invisible to the human eye. An exclusive integrated circuit is provided to measure the extremely short period of time in the nanosecond range between beam emission and reception. The laser beam scans it both horizontally and vertically. The scan angles and periods are calculated to detect any vehicle in front, the position of the vehicle (in the same lane or adjacent lane), the distance up to the vehicle, and the relative speed. The principle of the microwave radar is basically the same as that of the laser radar. The wavelength of microwaves is about 77 GHz, and the target object that is to be detected by the microwaves is the body of the vehicle in front. In the case of the microwave radar, the emitted microwaves do not scan the object; they are emitted at a wide angle and a phased-array antenna functions as the receiver of the reflection, thus realizing angular resolution. Microwave radars are superior in bad weather conditions like fog or rains and laser radars are more economical.

The current ACC is for cruise driving, but a new radar-based system has been developed to follow vehicles in a traffic jam. In addition, the radar-based precrash safety (PCS) system has been developed to reduce damage resulting from a collision. A system equipped with both a radar and a camera has been publicized recently, enabling sensing of integrated image information and distance information. These new systems are expected to significantly improve the convenience and the safety of automobiles.

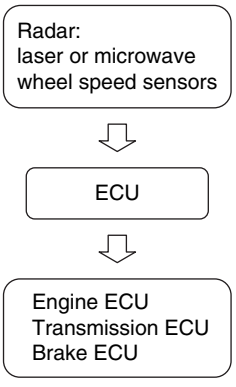


Figure 8 Structure of ACC.

3.12.2 Automotive Sensors

There were no electronic sensors in vehicles in the 1950s. All the components were mechanically controlled. MEMS sensors are very useful for satisfying automobile requirements, which realize new functions such as high-speed operation, fast response, self-diagnosis, compact and lightweight devices, low power consumption, long life, and cost efficiency. Hence, MEMS sensors were first used on vehicles in the 1980s, and have since become common. This section describes typical automotive sensors, such as the accelerometer, pressure sensor, gyroscope, and air flow sensor, as well as the most recent air quality sensor and laser radar.

3.12.2.1 Accelerometer

An accelerometer detects the impact to which the vehicle is exposed and the inclination of the vehicle as it accelerates. More specifically, an accelerometer is used with many control systems such as air bag systems, ABSs, stability control systems, navigation systems, and security systems. Concerning air bag systems, the acceleration at the time of a collision differs between a frontal and a side collision, and depends on the rigidity of the vehicle body. Therefore a sensor is required to sense gravitational acceleration (g) in the range of tens to hundreds of g ($1\,g = 9.8\,\mathrm{m\,s^{-2}}$). As for navigation systems and security systems, the inclination of the vehicle is

identified as changes in gravity, so that the sensor is required to be able to detect a range below 1 *g*. Accelerometers are now becoming common with the dramatic spread of air bag systems.

There are two typical detection methods for MEMS accelerometers: the piezogauge method and the capacitance method. Accelerometers using the piezogauge method started to be used on automobiles in the late 1980s (Payne and Dinsmore 1991, Sulouff 1991). As a result of the evolution of fabrication technologies, accelerometers using the capacitance method were first used on automobiles in the late 1990s. **Figure 9** shows the transition of typical accelerometers actually used on automobiles. In 1989, the sensor was large, with the piezogauge sensor element and circuit packaged in a can. As fabrication technologies advanced, it became possible to reduce the chip size, and the sensor size was drastically reduced. To reduce cost, ceramic packages emerged. Since the late 1990s, fabrication technologies advanced further; for example, the dry process replaced the wet process and surface micromachining technology was introduced. As a result, it became possible to reduce the chip size even further, enabling the size reduction of sensors. The detection method also changed from the piezogauge method to the capacitance method (Marek *et al.* 2003, Offenberg *et al.* 1996). Compared with the piezogauge method, the most important advantage of the capacitance method is that it has a self-diagnosis function that generates virtual acceleration, realizing extremely high reliability. Many accelerometer-equipped systems are safety systems, such as air bags, ABS, and stability control. A failure of the accelerometer means that these systems do not operate in an emergency, resulting in critical failure, i.e., failure in protecting passengers and accidents. For this reason, sensors using the capacitance method with a self-diagnosis function have become common, replacing those with the piezogauge method. Sections 3.12.2.1.1 and 3.12.2.1.2 explain the piezogauge method and the capacitance method, respectively.

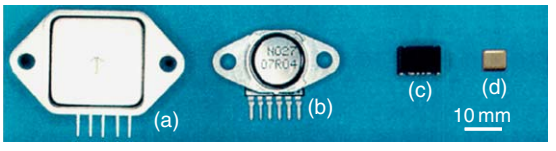


Figure 9 Transition of accelerometer; (a), (b), (c) are a piezo-gauge type, and (d) is a capacitance type.

3.12.2.1.1 Piezogauge-type accelerometer

Semiconductor materials change electrical resistance significantly depending on the strain applied. Among such materials, single-crystal silicon is advantageous because it is superior in reliability and because piezoresistors can be easily formed in the silicon by means of diffused resistance technology, which is used to fabricate integrated circuits. **Figure 10** outlines the structure of the sensing element used in the piezogauge method. The silicon substrate is fabricated to form a mass, and four beams support the mass, the beam being provided with piezoresistors. When acceleration is applied in the direction perpendicular to the substrate, the force of inertia displaces the mass. The amount of strain on the beam due to mass displacement is detected by measuring the changes in the piezoresistors. One of the piezoresistors is provided on the mass side of the beam and the other on the substrate side of the beam. As an example, if the mass is displaced upward from the substrate, the piezoresistor on the mass side receives tensile strain and the piezoresistor on the substrate side receives compressive strain. The tensile strain and the compressive strain increase resistance in one direction and decrease resistance in the opposite

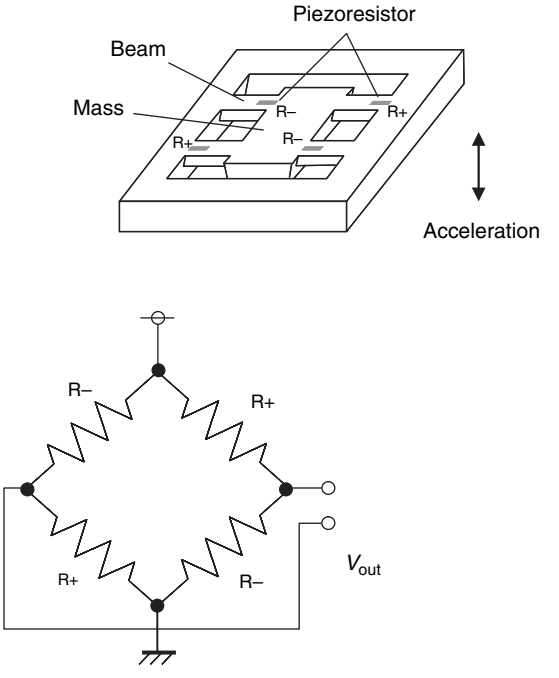


Figure 10 Outline of sensing element and circuit of piezo type accelerometer.

direction. Therefore, a bridge circuit configured as shown in **Figure 10** enables the detection of the change in resistance by doubling it. One problem with respect to piezoresistance detection is that piezoresistance is susceptible to changes in temperature. However, in principle a bridge circuit as shown in **Figure 10** enables piezoresistance detection independently of changes in temperature.

The following summarizes the production process of this structure. In a single-crystal silicon substrate, impurity ions are implanted, piezoresistors and wiring are formed by activation annealing, and aluminum pads are formed by sputtering. The pattern of the beam and the mass is formed on the front surface of the substrate by dry etching.The beam and the mass are formed by anisotropic etching on the backside of the substrate as the final process. Because the third power of the beam's thickness is proportional to sensitivity of this structure, careful control of the thickness is critical. Therefore, etching should be extremely precise to meet the tolerance of thickness.

3.12.2.1.2 Capacitance-type accelerometer

The advantages of the capacitance method are that it is, in principle, independent of temperature change and that it has an actuator function that produces electrostatic force when voltage is applied. **Figure 11** shows the structure of the sensing element using the capacitance method and also with the cross-sectional view of the element. The mass and the beam that support it are formed in the silicon substrate. The mass is provided with comb electrodes that have the function of capacitance. The comb electrode consists of a movable electrode connected to the mass and a fixed electrode connected to the substrate. The two electrodes are placed in opposite directions, producing capacitance. When acceleration is applied in the direction parallel to the substrate, the mass is displaced, whose displacement is determined depending on the spring constant of the beam. As a result of mass displacement, the distance between the movable and the fixed electrodes varies, changing the capacitance and detecting acceleration. Regarding fabrication dimensions, while the wet process used for the piezogauge method is in the range of 10 μm, the dry process has enabled fabrication in the extremely fine range of a few microns. The beam, the mass, the movable electrode, and the fixed electrode are formed in one body of single-crystal silicon or polycrystalline silicon. The silicon operates both as the physical structure and as the electrical wiring. By applying voltage between the fixed electrode and the movable electrode to this structure, the electrostatic force actuates the mass. The displacement is measured by the same electrodes as the change in capacitance, thus performing self-diagnosis. If the beam is damaged or either of the electrodes becomes disconnected, the capacitance changes differently from the normal change in response to a voltage application, indicating a failure.

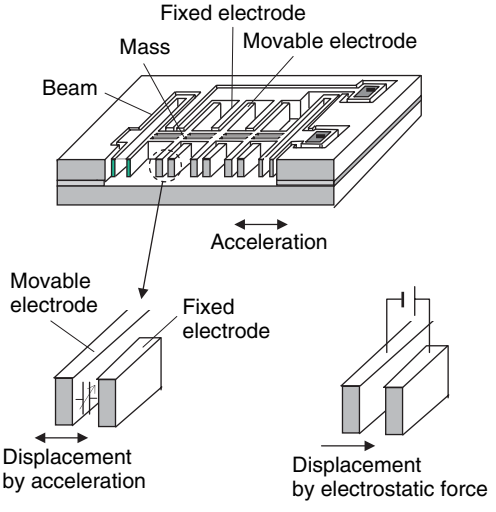


Figure 11 Outline of sensing element and self-test of capacitance type accelerometer.

The following summarizes the production process of this structure. A silicon dioxide film, which functions as the sacrificial layer, is deposited on the silicon substrate. Polysilicon, which forms the structure, is formed for producing a film of a predetermined thickness. The product fabricated by the processes described so far can be substituted by a silicon-on-insulator (SOI) substrate. In this case, single-crystal silicon should be used to produce the structure because it is more reliable. Electrode pads are produced in the next step. The pattern of the beam, the mass, and the electrode is formed at the same time by dry etching the polysilicon or single-crystal silicon. Lastly, the silicon dioxide film, which is the sacrificial layer, is removed by etching, and the beam, the mass, and the electrode are released from the silicon substrate to be movable. If sacrificial layer etching is performed in the liquid phase, the movable section may adhere to the fixed section due to surface tension produced during drying. To prevent this, the movable section should be kept immovable during drying, or etching can be performed in the gas phase. Another fabrication technology that has realized the

capacitance method is high aspect ratio fabrication. In the capacitance method, the greater the capacitance, the higher the sensitivity. To obtain greater capacitance, it is necessary to enlarge the electrodes placed in opposite directions. With high aspect ratio fabrication, the area of the electrodes can be enlarged in depth. As a result, sensitivity has been improved without enlarging the chip size. A new dry etching machine has enabled this technology: inductively coupled plasma system (Kuhl *et al.* 1998, Larmer *et al.* 1999). This machine has evolved further and today it can fabricate a trench of an aspect ratio of more than 50 (Ohara *et al.* 2001).

3.12.2.2 Pressure Sensor

Semiconductor-based piezoresistive pressure sensors were first inbuilt in automobiles in the early 1980s. Later, these sensors were integrated with IC technologies, and an integrated pressure sensor was developed, which consisted of an amplifier circuit, a regulator circuit, and a sensing element on a single chip. An onboard pressure sensor has a wide variety of functions, including the measurement of engine air intake pressure (manifold absolute pressure), fuel injection pressure, fuel tank internal pressure, air conditioner refrigerant pressure, brake pressure, and tire air pressure (Marek *et al.* 2003, Otake *et al.* 1998, Suzuki *et al.* 2001). Accordingly, the sensor is provided with an extremely wide pressure measurement range from 5 kPa to 200 MPa for common rail pressure sensors that detect fuel pressure in the diesel engine common rail system. **Figure 12** shows a MAPS and a refrigerant pressure sensor.

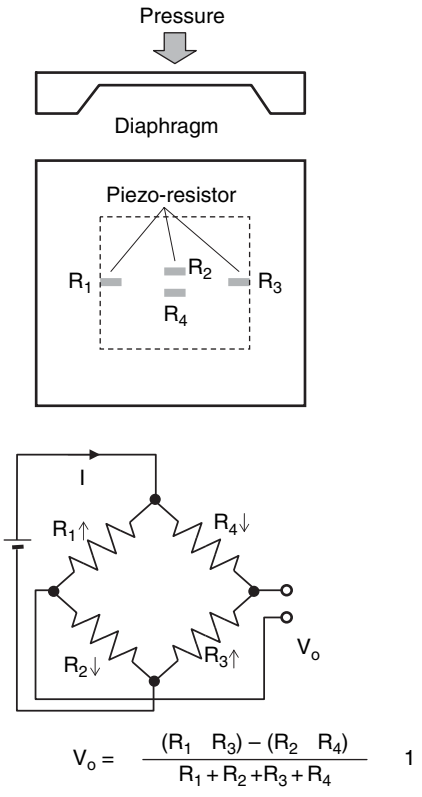
Manifold absolute pressure sensor

Refrigerant pressure sensor

Figure 12 Manifold absolute pressure sensor and refrigerant pressure sensor.

Most pressure sensors use the piezogauge method, which uses the trait of semiconductor materials to considerably change electrical resistance depending on the strain applied. **Figure 13** shows the outline of the sensing element and the circuit of a piezogauge-method sensor. The silicon substrate is processed to form a diaphragm, which is provided with piezoresistance. When exposed to pressure applied from above, the diaphragm sags. The amount of the diaphragm deformation is detected based on the changes in piezoresistance. The piezoresistors are at the center and end of the diaphragm. When the diaphragm deforms downward, the piezoresistor at the end of the diaphragm receives tensile strain while the piezoresistor at the center of the diaphragm receives compressive strain. The tensile strain and the compressive strain increase and decrease the resistance in opposite directions. Changes in the resistance can therefore be doubled in size for detection by configuring a bridge circuit. In addition, though one problem concerning detection using piezoresistance is that the resistance is influenced by temperature, so a bridge circuit compensates temperature dependence in principle.

The following summarizes the production process of this structure. In a single-crystal silicon substrate, impurity ions are implanted, piezoresistors and leads are formed by activation annealing, and aluminum pads are formed by sputtering. A diaphragm is produced using anisotropic etching from the back surface of the substrate as the final process. The diaphragm is normally 20 μm or greater in thickness, although it varies depending on the pressure to be measured. Some high-pressure sensors have a metal diaphragm connected to a silicon chip on which piezoresistors are formed (Tanaka *et al.* 2004). Like the thickness of the beam in accelerometers, the



$$V_o = \frac{(R_1 \quad R_3) - (R_2 \quad R_4)}{R_1 + R_2 + R_3 + R_4} \qquad 1$$

Figure 13 Outline of the sensing element and the circuit of a pressure sensor.

thickness of the diaphragm significantly influences sensor sensitivity. To ensure high thickness precision, some pressure sensors use electrochemical etch stop for anisotropic etching. Further, it has been reported that a satisfactorily etched surface has been successfully realized by controlling the impurity concentration of the potassium hydroxide (KOH) solution used for anisotropic etching in the parts per billion range (Inoue *et al.* 2000, Tanaka *et al.* 1999). The etching condition has been established and compiled in a database. Simulation technology has also been developed for this application. **Figure 14** shows the calculated shapes of etched silicon diaphragms.

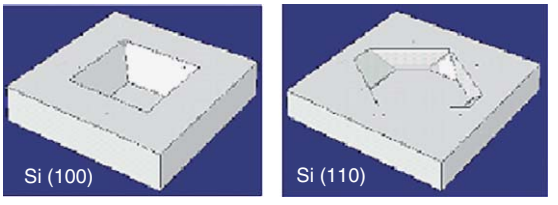


Figure 14 Calculated diaphragm shape of Si(100) and Si(110).

Many used a silicon (110) substrate on which an octagonal diaphragm can be formed. An octagonal diaphragm permits thermal stress distribution well balanced with the piezoresistive gauges, making it possible to minimize distortion of a sensor's offset voltage temperature characteristic.

3.12.2.3 Gyroscope

The gyroscope is commonly used to detect the turning. This turning angle is called yaw angle in the automobile field; therefore, the gyroscope is sometimes called yaw rate sensor. A gyroscope is typically used for stability control and navigation. It is also used in convertibles to detect a rollover rate. When the gyroscope on a convertible vehicle detects the possibility of a rollover, the roll bar rises to protect the heads of passengers (Marek *et al.* 2003). Most onboard gyroscopes use the Coriolis force in their detection mechanism. Vehicles having a high-precision ring laser gyroscope using the Sagnac effect were seen at one time, but today this type of gyroscope is not common due to cost and size (Barbour and Schmidt 2001). Then, Coriolis force-based gyroscopes were mostly of the tuning fork type (Putty and Najafi 1994), triangle pole type (Takeshi 1993), or ring type (Putty and Najafi 1994). These gyroscopes had a piezoelectric film on a metal structure, which was vibrated by the film. The gyroscope detected the deformation of the structure produced by Coriolis force by means of the film. Through the advancement of MEMS technology, these components of gyroscopes are today made from silicon materials. Some newly developed gyroscopes use electrostatic force to excite vibration (Howe *et al.* 1996).

Figure 15 shows the outline of the sensing element of a gyroscope using the capacitance. In the figure, metal leads are omitted. A movable mass is supported by the beams and by the frame. The mass is provided with comb electrodes to excite vibration by electrostatic force. The frame is equipped with comb electrodes to sense capacitance. The specific process of sensing is as follows. The gyroscope allows the exciting comb electrode to produce electrostatic force, and the mass to produce vibration in the x-axis direction. If rotation occurs around the z-axis direction at this moment, Coriolis force is produced on the mass, and the mass and the frame are displaced in the y-axis direction. This displacement is detected from changes in the capacitance of the comb electrode with the frame. The mass is resonated so as to obtain a sufficient degree of vibration speed. The Coriolis

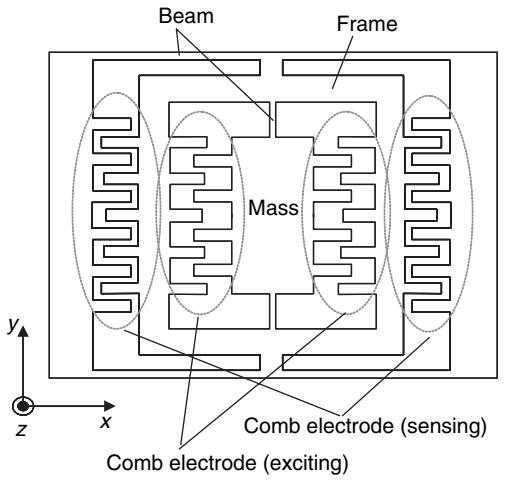


Figure 15 Plain figure of MEMS type gyroscope.

Figure 16 Air flow sensor.

force is proportional to the vibration speed of the mass. In this sensing mechanism, the biggest noise is caused by the vibration of the vehicle (acceleration). To solve this problem, two masses are used and vibrated in different phases whose difference is 180°. Though the displacement of two masses due to acceleration is in the same phase, Coriolis force of the two masses occur in opposite phases, so that the two masses are displaced in the opposite directions along the *y*-axis. By subtracting the displacement signal from another displacement signal, the noise due to the acceleration is eliminated. A gyroscope that detects both angular velocity and acceleration is now emerging (Nagao *et al.* 2004). The production process of gyroscopes is quite similar to the production process of the accelerometer as shown in Section 3.12.2.1.

3.12.2.4 Air Flow Sensor

The air flow sensor measures the amount of air intake into the engine. On the basis of the signals of the air flow sensor, the ECU determines an adequate rate of fuel injection. In the 1970s, a vane-type air flow sensor, which used the phenomenon of plate displacement that occurred depending on the amount of air, was developed. In the 1990s, the hot wire sensor was developed as shown in **Figure 16** (Kowatari *et al.* 1998). This sensor has the heater placed in the intake manifold, and it senses the changes in heat radiation from the heater to air, which occurs depending on the air velocity. The sensor was therefore able to measure the mass flow of air (Cavalloni and Berg 2002).

The hot wire sensor consists of a heater formed by winding platinum wire around a ceramic bobbin (Kowatari *et al.* 1998), a few millimeters in length as shown in **Figure 17**. The sensor is provided with a thermometer because heat radiation is influenced by changes in intake air temperature. In the late 1990s, a thin-film heater sensor was developed using MEMS technology (Ashauer *et al.* 1998, Moser and Baltes 1993, Moser *et al.* 1992, Neda *et al.* 1996, Parameswaran *et al.* 1990). The thin-film heater sensor using MEMS technology was developed to satisfy the needs of an air flow sensor that could handle high-speed changes in air intake resulting from the sophisticated engine control. The former bobbin heater with a millimeter range has too large a heat capacity to measure fast changes in air intake. The thin-film heater with a micrometer range having small heat capacity solves the problem.

Figure 18 shows a typical structure of the sensing element of the thin-film heater sensor. A membrane is formed in the silicon substrate, and a thin-film heater and a thermometer used for flow rate measurement are provided in the membrane. Another

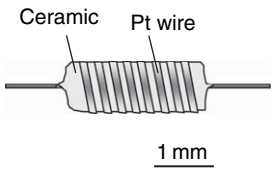


Figure 17 Bobbin heater.

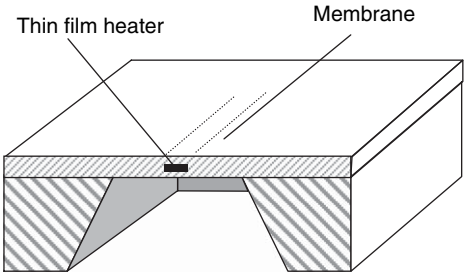


Figure 18 Cross section of MEMS type heater.

thermometer used for air temperature measurement is placed on the silicon substrate side. The membrane size is generally hundreds of square micrometers and a few micrometers in thickness, reducing the heat capacity of the heater while thermally insulating it. Because of these arrangements, the sensor can perform fast-response, power-saving sensing. The sensing process proceeds as follows. Air temperature is measured as the first step, and the heater temperature is retained at a temperature higher than the air temperature by a specific difference. Because the difference between air temperature and heater temperature is unchanged, the variation in heat radiation attributed to changes in air temperature can be corrected and only heat radiation from the air flow can be detected. The amount of electric power applied to the heater represents heat radiation, namely, the flow rate. To measure the air temperature, changes in electrical resistance, which varies with temperature, are used. The sensitivity of the sensor depends on how correctly the electrical resistance can be detected; therefore, a highly sensitive, stable material, such as platinum, is used to produce the heater. The heater can also be used as the thermometer of the heater itself. Some sensors have thermometers before and after the heater to measure the temperature distribution in the membrane, thus measuring both the direction of flow (Moser *et al.* 1992, Moser and Baltes 1993).

The following summarizes the production process of this structure. On a silicon substrate, a silicon dioxide film and a silicon nitride film, both of which serve as membrane materials, and a thin heater film are formed, and then patterning is performed. From the backside of the substrate, the silicon undergoes anisotropic etching, which stops automatically when it reaches the silicon dioxide film or the silicon nitride film. The silicon dioxide film and the silicon nitride film have stresses, which may cause buckling or damage to the membrane. To avoid this, the stress of each material should be identified and adjustments should be made taking the entire structure into consideration.

Figure 16 shows the entire body of the air flow sensor. The sensing element is not placed directly in the intake manifold but in the bypass flow channel. The entire bypass flow channel is placed in the intake manifold. The bypass flow channel plays an important role in the sensor, improving sensor sensitivity and removing disturbance.

3.12.2.5 Air Quality Sensor

Concerning the air-conditioning system for the cabin area, smell of exhaust fumes makes passengers uncomfortable. In an advanced air-conditioning system, exhaust gases are detected, the system switches fresh/recirculation dumper (Frers 2001). This indoor air/outdoor air automatic switching system first emerged in Europe, and is used for high-class vehicles. This system was initially equipped with a sensor that detected NO_x, and later with another sensor that detects CO was developed for use with this system. Though there are many methods for gas detection, most sensors use changes in the electrical resistance of a metal oxide to detect gas (Clifford 1983, Fung *et al.* 1996, Lee *et al.* 1995, Liu 1992, Seiyama *et al.* 1962). In the air, oxygen is trapped on the surface of the metal oxide, producing an electrical depletion layer, which provides the metal oxide with high electrical resistance. When combustible gases such as NO_x exist in the air, the effect of catalyst on the surface of the metal oxide produces oxidation reaction with oxygen, removing the oxygen from the surface of the metal oxide. The electrical depletion layer disappears and lowers electrical resistance (Clifford 1983). **Figure 19** outlines the structure of the air quality sensor's sensing element. The sensor consists of a heater and a thin film; the film is made from metal oxides such as SnO_2, which are usually formed on the ceramic substrate. The heater plays the role of heating the metal oxide and accelerates the reaction. The metal oxide is provided with a

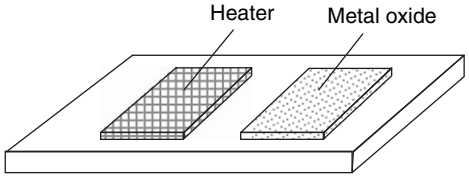


Figure 19 Outline of air quality sensor.

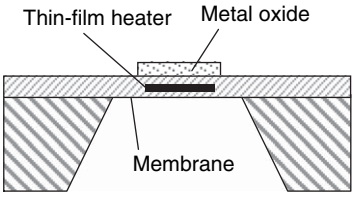


Figure 20 Cross section of MEMS type air quality sensor.

catalyst made of, for example, platinum. As this method detects all combusting gases in principle, it is designed to detect only a specific gas by formulating the materials of the metal oxide and the catalyst, as well as heating temperature. Air quality sensors have not been integrated with the latest MEMS technology (Fung *et al.* 1996, Lee *et al.* 1995, Liu 1992); however, those using the thin-film heater like the previously mentioned air flow sensors have been developed because such a heater is advantageous in various respects, such as power dissipation as shown in **Figure 20**. To improve sensitivity further, many efforts have been made concerning the formulation of the metal oxide, including research on film-forming methods and porous film. The air quality sensor integrated with MEMS technology is expected to spread rapidly in the future.

3.12.2.6 Laser Radar

The laser radar is used in the ACC system (Furui *et al.* 1998, Miyakoshi *et al.* 1999, Osugi *et al.* 1999). The principle of the distance measurement is as follows. A laser beam is emitted from the front of the vehicle in pulses, and the reflection from the target object is received by the photodiode. It was first used in passenger vehicles in the mid-1990s.

Figure 21 shows the laser radar. The laser diode, as shown in **Figure 22** (Abe *et al.* 2001, Atsumi *et al.* 2004, Kimura *et al.* 2000), emits a pulsed laser beam. The laser beam emitted from the laser diode is first reflected by the reflective mirror, scanned laterally and longitudinally by a rotating polygon mirror (2D scanning), and emitted through the emission opening. The beams reflected by the object in front are collected in the photodiode by means of the lens, and converted into electrical signals. These signals are processed by the time-measuring integrated circuit (Watanabe *et al.* 1993, Watanabe *et al.* 1998) and by obtaining the distance between the individual scanned positions. The reason why scanning is performed is that when a vehicle is moving in the horizontal direction the vehicle in front should not be lost on a curve in the road and when a vehicle is moving in the vertical direction the vehicle in front should not be lost in an upward or a downward slope. The radar's maximum measurement distance is currently 120 m, and its distance resolution is 1 m. The radar is useful not only in the ACC system but also in precollision safety (PCS) systems and for supporting parking (Alessandretti *et al.* 2002, Fuerstenberg *et al.* 2002). Installation space is limited in automobiles, and to increase sales of the laser radar, size reduction is essential. Therefore it is necessary that MEMS technology-integrated radars are earmarked for size reduction.

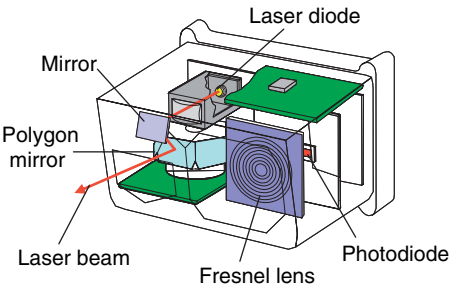


Figure 21 Review and structure of laser radar.

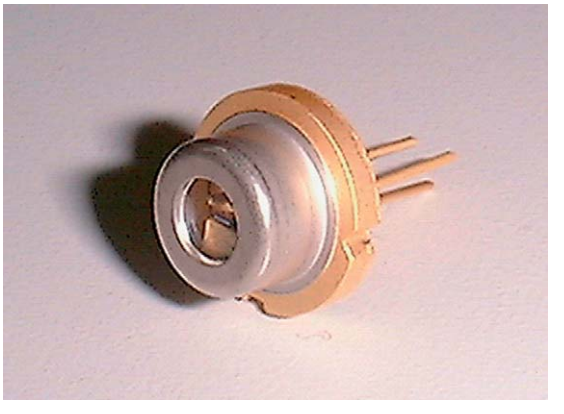


Figure 22 Laser diode for laser radar.

3.12.3 Future Direction of the Automotive Industry

One hundred years have passed since the invention of the modern automobile. While the automobile has provided us with great convenience, one cannot deny the problems it has caused: environmental pollution from fuel emissions in addition to an increase in the number of deaths and injuries due to traffic accidents. Accordingly, future methods of manufacturing cars must target environmentally friendly cars that obtain higher gas mileage ratings and safer cars capable of, if not preventing accidents, reducing the injurious effects of accidents on passengers and pedestrians. In the early 20th century, the basic functions of the automobile were to run, stop, and turn. Recently, in addition to these basic functions, automobiles have a new function, i.e., communication. Advances in electronic technology have contributed to this new function. Through various media, such as satellites, cellular phones, and broadcasts, we have acquired the ability to drive safely and more comfortably. We call this telematics, which is derived from the words telecommunication and informatics. Automotive telematics includes the transmission of information to and from vehicles using advanced telecommunication technologies. In the future, the world is scheduled to evolve into ITS by dedicated short-range communications (DSRC) between roads and vehicles as well as between vehicles. ITS is based on the principle that humans, roads, and vehicles mutually interact and cooperate by sharing information, as shown in **Figure 23**. New systems, such as the advanced cruise-assist highway system (AHS), the advanced safety vehicle (ASV) system, the human–machine interface (HMI) system as well as others, are key technologies for realizing this principle. After the realization of these technologies a new relationship between humans, roads, and vehicles will be created.

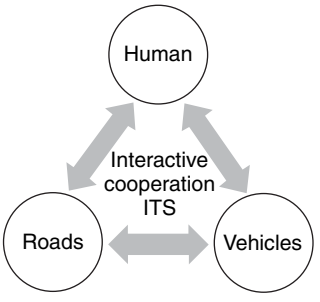


Figure 23 Basic concept of ITS.

Automotive manufacturing technologies for the future must be aimed at creating vehicles that are fuel-efficient and environmentally friendly, as well as safe, i.e., collision-free or minimal in impact to pedestrians and vehicle occupants in the unfortunate event of a collision. It is also important to incorporate advanced information and communication technologies that are coordinated with road infrastructure systems into vehicles, with the aim of ensuring efficient vehicle traffic and for manufacturing safer and more comfortable vehicles in which drivers can share necessary information with other people anywhere and anytime. In future, vehicles are expected to evolve along the three parameters: ecology, safety, and comfort (Hara 2002). It is therefore necessary to develop and manufacture vehicles and automotive systems that comply with these parameters. Of the three parameters, safety should be given the highest priority since it relates to human lives. Microsystem technology is expected to play an immense role in realizing automotive safety.

3.12.3.1 Concept of Advanced Driving Assistance

Taking this viewpoint into consideration, the automotive industry has been implementing R&D with the focus on ITS. A typical example is the development of AHS for which the Ministry of Land, Infrastructure, and Transport has taken the initiative in Japan. Through the AHS, the ministry intends to provide seven basic services that are considered necessary for realizing a safe motorized society in future (ITARDA 2005). The seven services that have been proposed to provide support are as follows:

(1) Prevention of collisions with forward obstacles
(2) Prevention of overshooting on a curve
(3) Prevention of lane drift
(4) Prevention of crossing collisions
(5) Prevention of right-turn collisions
(6) Prevention of collisions with pedestrians on a crossing
(7) Road surface information for maintaining an appropriate headway distance.

The support services for points (1)–(3) are being realized by a radar system (Osugi *et al.* 2001) and a lane keep system by camera, which have already been installed in some vehicles.

At present, however, these systems are costly and affordable only on some luxury cars. It is necessary to reduce the size and cost of these systems further by

using microsystem technology. If these systems are to follow the same course of evolution as automotive sensors, the reduction in size and cost will make the systems and the safety secured by them affordable on general-class vehicles as well. Hence, there are high expectations from microsystem technology.

The support services for points (4)–(7) cannot be realized by on-vehicle equipment alone. Coordination with infrastructure systems is indispensable. In the United States, development of a system that supports a driver's decisions at intersections is under way in a project led by the University of California. A device installed at each intersection identifies vehicles approaching the intersection and advises each driver of the vehicle whether he/she may make a left turn (right turn in Japan) (Shladover 2003). Safety-related support for pedestrians and vehicles at intersections is highly desired for the future. One of the effective and promising measures for such support is a wearable HMI device manufactured by micromachining technology. The HMI device would be embedded in a tag, which is to be worn by vulnerable road users, i.e., by pedestrians. A sensor installed at each intersection would detect the tag information and inform vehicle drivers about the presence of pedestrians. The tag should be manufactured using microsystem technology so that it is extremely small in size and should preferably be provided with, say, elderly certificates, security labels on bicycles, name tags worn by elementary school children, etc. In these tags, a self-power supply system by Power MEMS is more desirable than batteries. At each intersection, an appropriate infrastructure system would inform vehicle drivers of approaching pedestrians and bicycles to help the drivers avoid recognition mistakes and resultant unfortunate collisions.

Many possibilities can be imagined when contemplating the contributions of microsystem technology to the needs of advanced driving assistant systems. Many microsystem technologies, such as a microsensor, a microactuator, an optical MEMS, a RF MEMS, a power MEMS, and an auditory MEMS, are currently in the research stage and are regarded as promising technologies for these systems. It is not an overstatement that advanced driving assistant systems are realizable with integration of various microsystem technologies. Here, the word integration does not mean the simple gathering of device functions. Instead, it means the creation of new standards for production ability and robustness.

3.12.4 Conclusion

The above paragraphs describe some practical automotive application examples of MEMS sensors, and the evolutionary trend of microsystem technology for the future. Microsystem is highly regarded as a key technology with the potential to meet the requirements of the coming ITS-based society and ubiquitous information network society and to create new systems with new functions. Microsystem technology is expected to play an extremely important role in the future direction of R&D in the automotive industry, particularly for achieving new levels of safety. In performing this role, a high degree of integration of microsystem devices is essential so that these systems maintain a high level of reliability and mass productivity. However, a microsystem is basically a manufacturing technology. For further progress in this technology, therefore, it is vital and indispensable to develop a wide variety of core technologies, which is time-consuming.

References

Abe K, Kimura Y, Atsumi K, Gotoh H 2001 High-power pulsed laser diode for an automotive laser radar sensor. *DENSO Tech. Rev.* **6(1)**, 62–7

Aga M 2004 Characteristics of vehicle stability controls effectiveness derived from the analysis of traffic accident data statistics. *Convergence Int. Congr. Exposit. Transportation Electronics*, Detroit, MI, USA, October 2004, Session: Digital Driving Increases Consumer Value, Paper No. 2004-21-0074

Alessandretti G, Baraud P, Domsch Chr, Sala G 2002 A European activity on pre-crash application: The Chameleon project. *Proc. e-Safety 2002.* Lyon, France

Ashauer M, Glosch H, Hedrich F, Hey N, Sandmaier H, Lang W 1998 Thermal flow sensor for liquids and gases. *Proc. MEMS'98,* Heidelberg, Germany, pp. 351–5

Atsumi K, Abe K, Kato H, Michiyama K 2004 Highly reliable high power diode laser for laser radar application. SAE Paper No. 2004-01-1594

Barbour N, Schmidt G 2001 Inertial sensor technology trends. *IEEE Sens. J.* **1(4)**, 332–9

Betta R D, Sheridan D, Cizeron J 2004 Fuel processor enabled NO_x adsorber after treatment system for diesel engine emissions control. *DEER Conference,* Coronado, CA, USA

Cavalloni C, Berg J V 2002 Measurement-pressure sensing in automotive applications. In: Krueger S and Gessner W (eds.) *Advanced Microsystems for Automotive Applications 2002.* Springer-Verlag, Berlin, pp. 232–42

Clifford P K 1983 Homogeneous semiconducting gas sensor: A comprehensive model. *Proc. Int. Meet. Chemical Sensors*, Fukuoka, Japan, Vol. A113, p. 135

Frers K D 2001 Air quality sensor for air condition units. In: Krueger S and Gessner W (eds.) *Advanced Microsystems for Automotive Applications 2001.* Springer-Verlag, Berlin, pp. 167–75

Fuerstenberg K Ch, Dietmayer K C J, Stephan E, Volker W 2002 Multilayer laser scanner for robust object tracking and classification in urban traffic scenes. *Proc. ITS 2002, 9th World Congr. Intelligent Transport Systems,* Chicago, IL, USA

Fujikawa T 2004 Microsystem for automotive electronics systems. *Proceedings of the 10th International Micromachine/Nanotech Symposium,* Tokyo, Japan, November 11, 2004, p. 14

Fung S K, Tang Z, Chan P C, Sin J K, Cheung P W 1996 Thermal analysis and design of a micro-hotplate for integrated gas-sensor applications. *Sens. Actuators A* **54**, 482–7

Furui N, Miyakoshi H, Noda M, Miyauchi K 1998 Development of a scanning laser radar for ACC. SAE SP-1332 No. 980615, pp. 71–6

Hara K 2002 Micromachine strategy in automotive industry. *Proceeding of the 8th World Micromachine Summit,* Maastricht, The Netherlands

Hirooka H, Mori S, Shimizu R 2004 Effects of high turbulence flow on knock characteristics. SAE Paper No. 2004-01-0977

Honda Y, Ninomiya H, Sassa Y 2001 Development of cabin air quality system. *SAE 2001 World Congress*, Detroit, MI, USA, March 2001, Session: Climate Control Systems (Part A), Paper No. 2001-01-0292

Hou Y, Zhuoping Y 2003 The evaluation of ABS performance. *SAE 2003 World Congress & Exhibition*, Detroit, MI, USA, March 2003, Session: Anti-Lock Braking/Traction Control (ABS/TCS) Systems

Howe R T, Boser B E, Pisano A P 1996 Polysilicon integrated microsystems: Technologies and applications. *Sens. Actuators A* **56**, 167–77

Inoue K, Kosaka S, Kawai Y, Tanaka H, Abe Y, Yoneyama T 2000 *Electrochem. Soc. Proc. Vol. Clean. Technol. in Semicond. Device Manuf. VI*, p. 491

ITARDA URL 2005, http://www.itarda.or.jp/info33/info33_1.htm#part1

Japan Electronics and Information Technology Industries Association (JEITA) 2004 Report of Sensor Production XVII, p. 72

Kamiji K 2003 Present technologies and future of occupant sensing system. *J. Soc. Automot. Engrs Jpn* **57(2)**, 35–8

Kataoka T, Kato K, Kumada T, Mizuno 2005 New technology of climate control comfort in passenger compartments. *DENSO Tech. Rev.* **10(1)**, 58–63

Kimura Y, Matsushita N, Kato H, Abe K, Atsumi K 2000 High power pulsed diode laser for automotive scanning radar sensor. *Proc. SPIE* **3888**, 759–66

Kishi N, Kikuchi S, Seki Y, Kato A, Fujimori K 1998 Development of the high performance L4 engine ULEV system. *International Congress & Exposition*, February 1998, Detroit, MI, USA, Session: Advanced Converter Concepts For Emission Control. SAE Paper No. 980415

Kowatari T, Arai N, Kobayashi C 1998 Signal processing for error reduction in a hot-wire air flow sensor. *SAE Technical Paper Series #980265*, pp. 23–7

Kuhl K, Vogel S, Schaber U, Schafflik R, Hillerich B 1998 *Advanced silicon trenchetching in MEMS applications. Part of the SPIE Conference on Micromachining and Microfabrication Process Technology IV*, SPIE, San Diego, CA, pp. 97–105

Larmer F, Schip A, Funk K, Offenberg M 1999 Bosch deep silicon etching: Improving uniformity and etch rate for advanced MEMS applications. *Technical Digest MEMS'99*. IEEE Robotics and Automation Society, Orlando, FL, USA, pp. 211–16

Lee K, Yun D H, Park H S, Shin H W, Hong H K, Kwon C H, Kim S T 1995 Thin-film micro-gas sensor for detecting CH_3SH. *Sens. Mater.* **7(1)**, 35–41

Liebermann E 2004 Safety and performance enhancement: The Bosch electronic stability control (Esp). *Convergence Int. Congr. Exposit. Transportation Electronics*, Detroit, MI, USA, October 2004, Session: Active Safety, paper No. 2004-21-0060

Liu C C 1992 Development of chemical sensors using microfabrication and micromachining techniques. *Proc. Int. Meet. Chemical Sensors,* Tokyo, Japan, pp. 2–5

Lu Z, Chan P 2004 Out-of-position airbag load sensitivity study. *SAE 2004 World Congress & Exhibition*, Detroit, MI, USA, March 2004, Session: Air Bags (Part 1 of 3)

Marek J, Trah H-P, Suzuki Y, Yokomori I 2003 *Sensors for Automotive Technology*. Wiley-VCH, Weinheim

Miyakoshi H, Furui N, Miyakoshi T, Noda M, Osugi K, Miyauchi K 1999 Development of adaptive cruise control system. *TOYOTA Tech. Rev.* **48(2)**, 88–93

Moser D, Baltes H 1993 A high sensitivity CMOS gas flow sensor on a thin dielectric membrane. *Sens. Actuators A* **37–8**, 33–7

Moser D, Lenggenhagar R, Wachutka G, Baltes H 1992 Fabrication and modeling of CMOS microbridge gas-flow sensors. *Sens. Actuators B* **6**, 165–9

Nagao M, Watanabe H, Hashimoto M, Nakatani E, Shirai K, Aoyama K 2004 A silicon micromachined gyroscope and accelerometer for vehicle stability control system. *SAE 2004 World Congress & Exhibition*, Detroit, MI, USA, March 2004, Session: Vehicle Sensors & Actuators (Part 1 of 2), 2004-01-1113

Neda T, Nakamura K, Takumi T 1996 A polysilicon flow sensor for gas flow meters. *Sens. Actuators A* **54**, 626–31

Nishida M, Isobe R, Goto T, Hanzawa T, Aiga S 2005 The new 2.3L direct injection turbo gasoline engine from Mazda. *14th Aachen Colloq. Automobile and Engine Technology,* Aachen, Germany, October 5–6, 2005

Nishizawa K, Ishizuka Y, Yamada T, Inoue T 1997 Technologies for reducing cold-start emissions of V6 ULEVs. *International Congress & Exposition*, Detroit, MI, USA, February 1997, Session: Advanced Emission Control Technologies. SAE Paper No. 9701022

Offenberg M, Munzel H, Schubert D, Schatz O, Larmer F, Muller E, Maihofer B, Marek J 1996 Acceleration sensor in surface micromachining for airbag applications with high signal/noise ratio. SAE Paper No. 960758

Ohara J, Kano K, Takeuchi Y, Otsuka Y 2001 Improvement of Si/SiO_2 mask etching selectivity in the new D-rie process. *Technical Digest MEMS 2001*. IEEE Robotics and Automation Society, Interlaken, Switzerland, pp. 76–9

Osugi K, Miyauchi K, Furui N, Miyakoshi H 1999 Development of scanning laser radar for ACC system. *JSAE Rev.* **20**, 549–54

Osugi K, Miyauchi K, Furui N, Miyakoshi H 2001 Development of the scanning laser radar for ACC system. *DENSO Tech. Rev.* **6(1)**, 43–8

Otake S, Onoda M, Nagase K 1998 Automotive high pressure sensor. *International Congress & Exposition*, Detroit, MI, USA, February 1998, Session: Sensors and Actuators. SAE Paper No. 980271

Parameswaran M, Robinson A M, Ristic K C, Allegretto W 1990 A CMOS thermally isolated gas flow sensor. *Sens. Mater.* **2(1)**, 17–26

Payne R S, Dinsmore K A 1991 Surface micromachined accelerometer: A technology update. SAE Paper No. 910496

Putty M W, Najafi K 1994 A micromachined vibrating ring gyroscope. *Solid-State Sensor and Actuator Workshop,* Hilton Head, SC, USA, June 13–16, 1994, pp. 213–20

Seiyama T, Kato A, Fujiishi K, Nagatani M 1962 A new detector for gaseous components using semi conductive. *Thin Films Anal. Chem.* **34(11)**, 1502

Shimoda M, Hosoya M, Takenaka Y 2006 Diesel exhaust emissions reduction technology for commercial vehicles in 2010. *J. Soc. Automot. Engrs Jpn* **60(1)**, 49–54

Shladover S 2003 California PATH program, intersection decision support. *Proc. IVI National Meeting and*

Demonstration, Washington DC, USA, June 24–26, 2003, URL: http://www-path.eecs.berkeley.edu/

Sugi H, Yamanaka Y 2005 Present status of thermal systems and possible future trends. *DENSO Tech. Rev.* **10(1)**, 58–63

Sulouff R 1991 Silicon sensors for automotive applications. *Int. Conf. Solid-State Sensors and Actuators Digest of Technical Papers*. IEEE, San Francisco, CA, USA, pp. 170–6

Summers S, Willke D T, Sullivan L K 2005 NHSTA'S crashworthiness rollover research program. *19th ESV Conference*. Paper No. 05-0279

Suzuki Y, Imai M, Mizuno T, Yokomori I 2001 Evolution of automotive semiconductor pressure sensors. *DENSO Tech. Rev.* **6(1)**, 96–100

Takeshi N 1993 Detecting circuit for a vibrating gyroscope. *US Pat.* 5 220 833

Tanaka H, Abe Y, Yoneyama T, Ishikawa J, Takenaka O, Inoue K 1999 Effects of small amount of impurities on etching of silicon in aqueous potassium hydroxide solutions. *Proc. Transducers '99,* Sendai, Japan, p. 539

Tanaka H, Ito O, Suzuki Y 2004 High-pressure sensors for common rail systems. *DENSO Tech. Rev.* **9(2)**, 19–23

Tokoro S, Kuroda K, Kawakubo A, Fujita K, Fujinami H 2003 Electronically scanned millimeter-wave radar for pre-crash safety and adaptive cruise control system. *IV2003 IEEE Intelligent Vehicle Symposium,* Columbus, OH, USA, pp. 304–9

Watanabe T, Makino Y, Ohtsuka Y, Akita S, Hattori T 1993 A CMOS time-to-digital converter LSI with half-nanosecond resolution using a ring gate delay line. *IEICE E* **76-C(12)**, 1774–9

Watanabe T, Isomura H, Agatsuma S, Ohtsuka Y, Akita S, Hattori T 1998 Development of a time-to-digital converter IC for laser radar. *JSAE Rev.* **19**, 161–5

Biography

Nobuaki Kawahara received the BE, ME, and Ph.D. degrees from Kyushu University, Japan, in 1981, 1983, and 1993, respectively. He joined Denso Corporation in 1983 and has been engaged in the research and development of microdevices, microfabrication, and microsystems. He is now general manager of Denso research laboratory and is responsible for the development of future automotive devices and systems. He is a member of the IEEE, the Japan Society, the Society of Mechanical Engineers, and the Society of Automotive Engineers of Japan and is a visiting professor of Nagoya University.

3.13 Environmental Monitoring

Takashi Korenaga, Tokyo Metropolitan University, Tokyo, Japan

Glossary

CCD Charge-Coupled Device
CE Capillary Electrophoresis
FIA Flow Injection Analysis
GC Gas Chromatography
LC Liquid Chromatography
LIGA Lithographie Galvanoformung Abformung
MEMS Microelectromechanical System
MS Mass Spectrometer
μ-TAS Micro Total Analysis System
PAH Polycyclic Aromatic Hydrocarbon
PDMS Polydimethylsiloxane
PMMA Polymethyl Methacrylate
PTFE Polytetrafluoroethylene
VOC Volatile Organic Compound

3.13.1 Introduction

In the field of advanced analytical science and technology, as most of the researches using microsystems have been proposed for analyzing liquid samples, few researches deal with the gas and solid phases. Moreover, few researches have been done on the application of microchip techniques to atmospheric gas analysis because there is no suitable gas sampling method. Because of the large differences between physical parameters of gas and liquid phases, such as viscosity and compressibility, the control of a gas–liquid two-phase flow in a microchannel was relatively difficult. Thus, mostly, gas–liquid two-phase flow on the microfluidic devices is treated as gas–liquid segmented flow. A miniaturized ammonia sensor has been reported. In this system, gas flow and liquid flow were introduced in the reaction chamber, and an electrolyte conductivity sensor was used for the detector. Hibara *et al.* (2005) reported an evaporation utilizing the gas–liquid two-phase laminar flow in the microchannel. However, as this system is dependent on special microchannel structures and surface modification, complicated and advanced fabrication techniques are required. A combination of thermal lens microscopy (TLM) and time-of-flight mass spectrometry helps in ultrasensitive detection, as both the microanalysis systems used in detecting microamounts of samples have excellent sensitivity when compared with conventional analysis systems.

Chemical and biochemical analysis systems using microsystems have many advantages in analytical chemistry. The microsystems realized fast analysis, reduction of reagent consumption, and portability, so that the microsystems have been applied to many analytical researches such as flow injection, solvent extraction, and immunoassay. Furthermore, microsystems can also be used as a microreactor for organic synthesis, and high reaction yields compared to the conventional procedure have been reported. For example, enzyme reaction for horseradish peroxide in a microsystem was compared to the reaction in a cuvette by using TLM, and almost two times higher reaction rate than the bulk scale was reported. It is considered that such advantages of the micro total analysis system (μ-TAS) arise from the unique mass transfer mechanism in microchannels. Therefore, it is important to understand the specific fluidic behavior such as diffusion and mixing of reagents that exists as laminar flows in microchannels of microsystems.

Recently, observation of the fluidic behavior of solutions in a polymer (PMMA, polymethyl methacrylate) microchannel was reported. The mixing and diffusion of dyed solutions was monitored by using a simple commercial charge-coupled device (CCD) camera equipped with an optical microscope, which clearly showed the unique mixing process of liquids at the channel turnings with different shapes and roughness of the inner walls. Microsystems used for μ-TAS are usually made of silicon, polymer, and glass. Among these materials, glass is the most useful material because of its easy processibility. Besides, glass is transparent to visible light, so that it is easy to trace and observe the mixing and diffusion of reagents in microsystems. The specific fluidic behavior of solutions ($KMnO_4$, Basic Blue 3 (BB3), and Oriental Blue (OB)) in a glass microsystem was investigated using the same technique reported for the PMMA microsystem. In the study, however, not only the simple observation of mixing of liquid flows, but also the quasi-steady-state UV-visible absorption spectroscopy was carried out using the microsystem owning to the optical transparency of quartz glass. The spectroscopic results obtained were analyzed based on a convection–diffusion equation in one dimension, and fluidic properties of liquids in different types of microchannels (wet etched and mechanical fabricated) were discussed. It is certain that much more powerful methods will be developed as new microreactors and analytical microsystems for environmental analysis and monitoring are required worldwide (**Table 1**).

3.13.2 Theoretical and Experimental Study on Microfluidics

A microspace such as a microchannel in the microchip system is characterized by a small volume, a short diffusion length, a large specific interface (solid–liquid or liquid–liquid), and small heat capacity. By utilizing these characteristics, micro chemical systems are expected to offer improved performances over conventional systems. Therefore, it is important that microsystems are designed to obey Taylor laminar–convection flow theory in the miniaturized flow space.

The sampling procedure for gas samples can generally be classified into active and passive methods. The active sampling method is performed using a vacuum pump, and has the advantage that adjustment of the amount of sampling is easy. However, as this method needs power, the sampling point is restricted.

Table 1 Required microsystems conditions for environmental analysis and monitoring

Compound	*Sample tested*	*Concentration level*	*Precision (%)*
NO_2	Air, indoor air	1–100 ppbv	±10
SO_2	Air	1–40 ppbv	±20
O_3	Air	60 ppbv	±10
CO	Air, indoor air	0.4 ppmv	±20
NH_3	Air, clean room air	0.1 ppmv	±10
HCHO	Air	2–100 ppbv	±10
Benzene	Air	20 ppbv	±10
Trichloroethylene	Groundwater → air	200 ppbv	±10
Tetrachloroethylene	Groundwater → air	200 ppbv	±10
Boron	Water	0.3 mg l^{-1}	±10
Aluminum	Water	0.2 mg l^{-1}	±10
Fluoride	Water	1.5 mg l^{-1}	±10
Bisphenol A	Water	–	–
Benz[*a*]pyrene	Exhaust particulate, dust	–	–
Polycyclic aromatic compounds	Exhaust particulate, dust	–	–
Polychlorinated biphenyl	Soil, transmittance oil	–	–
Dioxine	Air, water, soil, dust	–	–

On the other hand, the passive sampling method, which samples the object gas by diffusion, is suitable for miniaturization of sampling devices and the multipoint sampling is enabled.

Diffusion denuders, a tube with interior walls coated to function as a sink for the desired gases, have been used since the 1970s for the collection of gases. The original implementation was labor-intensive, wetted denuders in which a flowing liquid is used as the collection surface and their membrane-based counterparts are called diffusion scrubbers. Nowadays, automated continuous collectors are used (Simon and Dasgupta 1993). The first sampling methods based on diffusion to the absorption solution were investigated by Fukui (1963). Although these samplers are light, small, and easily operable, there is a tendency for the absorption rate to change with variations in wind velocity. In order to diminish the effect of wind velocity, a personal (passive) sampler, which has an air column in an acrylic tube, was developed (Palmes *et al.* 1976). Yanagisawa *et al.* (1982) developed a filter badge-type passive sampler, which consisted of an absorption sheet and a diffusion-controlling mat for diminishing the effect of wind velocity. Nowadays, a number of passive samplers adapted for various gases such as SO_2 (Korenaga and Yanagisawa 1995), CO (Lee *et al.* 1992), and O_3 (Helaleh *et al.* 2002) are being developed. However, as the sampling rate in passive samplers is dependent on gaseous diffusion, sufficient sampling time is needed to determine the level. Consequently, it is unsuitable for continuous on-site monitoring of gaseous concentration, which change with the passage of time.

A novel sampling method using the droplet formed on the tip of the capillary has been reported and applied to NH_3, SO_2 (Chang *et al.* 2005, Dasgupta and Kar 1995, Liu and Dasgupta 1995a, b, Yang *et al.* 1997), and nitrogen dioxide (NO_2) (Helaleh *et al.* 2001, Korenaga 1997, Korenaga 2002, Kumemura and Korenaga 2004a, b, Sato *et al.* 2001, Takabayashi *et al.* 2006). In this procedure, gases diffuse from the gas phase into the liquid phase of a droplet. However, as the interface of droplet is in direct contact with the atmosphere, it is easily influenced by humidity and temperature. In such a sampling method using liquid droplets (Kumemura and Korenaga 2004a, b, Kumemura and Korenaga 2006, Kumemura *et al.* 2005), a specific interface area is obtained by a decrease in droplet volume.

3.13.2.1 Downsizing of Flow Systems in Flow Injection Analysis

Downsized flow systems offer a lot of excellent innovations to various fields, such as medical, environmental, biological, and industrial applications. When a microchannel is characterized by a small volume, a short diffusion length, a large specific interface, and small heat capacity, a miniaturized chemical analysis system offers excellent performances over conventional analysis methods. Therefore, it is important that μ-TASs be designed to obey the theory of miniaturized space. For

example, microarray analysis has become a powerful technology for drug screening, disease gene identification, and signaling pathway studies. In order to utilize microchip technologies effectively, it is necessary to grasp the characteristics of microspace and to design based on theory. As mentioned above, microchip devices for analytical chemistry have many advantages compared with the conventional chemistry system. In order to enable broader applications of microchip technologies in the field of chemistry, the selection of suitable materials and fabrication technique is required. Moreover, the detection of low concentrations of analytes in the small volumes of samples present in microfluidic chip requires sensitive analytical techniques. The characteristics of microspace, microfabrication technologies, and detection technologies are described below.

The solute dispersion of an injected sample zone was analyzed by using an experimental apparatus with ideal laminar flow to develop a hydrodynamic model for the downsizing design of sensitive and precise flow injection analysis (FIA) systems (Korenaga 1995, Korenaga and Stewart 1988, Korenaga *et al.* 1988, 1989a–c, 1991a, b, Takahashi *et al.* 1990). The dispersion behavior of the sample slug was first analyzed in detail under the flow systems with different manifold conditions to evaluate the effects of operating conditions such as tube radius, tube length, flow rate, and cell volume to control the molecular diffusion coefficient of the sample solute. The capillary flow properties were also examined for some commercial micropumps to select the most suitable pumping method for the minimized flow injection system. A double-plunger micropump having a linear cam mechanism and a fast, short reciprocation time was proposed to obtain smoother mixing and more stable pumping with good reproducibility (Korenaga *et al.* 1992). The s4-micro flow injection system with a small tube bore, short tube length, small cell volume, and slow flow rate gave regular and reproducible breakthrough peak profiles that reflected stable feeding; other flow systems showed irregular peak profiles at smaller flow rate regions (Korenaga 1992).

The s4-micro flow injection system that drives complete, rapid mixing and low flow rate pumping are strongly desired for reliable downsized flow injection systems for industrial process use. Biosensing devices for protein and enzyme bioassays require lower consumption of valuable reagents (Korenaga *et al.* 1994, Li *et al.* 1995).

Chemical sensors are widely needed and have thus been developed in the field of microelectromechanical systems (MEMS). The substrate material and microfabrication process were performed by applying the semiconductor processing method. Because the method of microfabrication was developed in the semiconductor industry, the analytical miniaturized devices were typically fabricated in silicon substrate in the early days. In recent years, materials such as glass and polymer have been widely used. Some of the attractive features associated with the use of glass include its excellent optical properties, allowing sensitive detection, e.g., fluorescence is used for readout, and also its surface chemical properties, which are well understood and allowed to modify chemistries extensively. Thus, glass substrates are mainly used for microfluidic systems, which are applied to organic extraction process and optical analysis such as spectrophotometric, fluorometric, and chemoluminescence detection methods.

Lithographie galvanoformung abformung (LIGA) process is attracting attention as a fabrication technique for 3D microstructures with a high aspect ratio. The process that has been developed in the field of MEMS is composed of deep X-ray lithography using synchrotron radiation, electroforming, and molding. A resist, e.g., PMMA, is coated on the substrate, and the desired pattern is transferred by X-ray exposure. X-ray lithography on PMMA can produce extremely high aspect ratios due to the small amount of mask undercutting and its ability to penetrate deep into the resist. The exposed resist is removed by development. By using electroforming, Ni is deposited on the bottom of the substrate and a Ni structure is created. A metal structure may be used as a micromachine product in this stage. The resist is removed and a Ni-electroforming molding die is produced. The 3D structures are fabricated in the polymeric materials by molding using a Ni molding die.

Polymeric materials are used widely because of a broad range of materials with different physical properties, flexibility in the choice of the microfabrication method, the ability to produce high aspect ratio microstructures, and potential disposability. Polydimethylsiloxane (PDMS) elastomer and PMMA resin are the commonly used materials now. PDMS elastomer is widely used in microfluidic applications to form components such as channels, valves, and diaphragms. Furthermore, since the absorption in a visible is small and most private fluorescence is not seen as the optical characteristic, it can be adapted also for fluorescence detection widely used in a biochemistry field. One of the advantages of PDMS is that self-adhesiveness over a flat surface is excellent when compared with other

materials. In addition, employing oxygen plasma oxidation of PDMS, irreversible bonds with glass, silicon, and other PDMS structures are achieved. The various microchip fabrication methods using PDMS are studied, as production of the microstructures in PDMS is comparatively easy when compared with that with other materials. PMMA acrylic resin exhibits excellent transparency, and bonding of substrates is achieved easily by thermal adhesive bonding. Other standard polymeric materials are also used such as polyamide (PA), polybutyleneterephthalate (PBT), polycarbonate (PC), polyethylene (PE), polypropylene (PP), polystyrene (PS), and polyethyleneterephthalate (PET).

These microsystem processing methods make use of the new concept of μ-TAS, which is based on the s4-micro FIA system (Korenaga 1992, Korenaga and Stewart 1988) shown in **Figure 1**.

3.13.2.2 Analysis of Diffusion and Mixing Process in Microchannel Laminar Flow

3.13.2.2.1 Liquid flow

The state of the flow in a microspace is distinguished into a laminar flow and the turbulent flow obeying Reynolds number (Re) as follows:

$$Re = \frac{Vd}{v} \qquad [1]$$

where V is the average flow velocity, d is the diameter of the tube, and v is the viscosity. If $Re < 2300$ then the flow is laminar, if >4000 then it is turbulent, and if in between these then it is in the transition zone. In general, in a microsystem, the diameter of the microchannel ranges from several micrometers to several hundred micrometers, the flow rate ranges from several microliters per minute to several hundred microliters per minute, and the aqueous viscosity is $\sim 10^{-6}\,m^2\,s^{-1}$. Therefore, as Re is estimated to be 10^{-6} to 20, the state of flow is stable laminar flow. The dispersion of a solution traveling in a laminar flow along the major axis of the microchannel is described by Taylor dispersion. The flow velocity distribution of laminar flow in the smaller tube radiates when the flow velocity becomes slow as it approaches the wall from the center. However, this is in the case of a cylinder-shaped tube. It is surmised that the flow velocity distribution is distorted partly as a section of microchannel is generally a rectangle or a short form.

3.13.2.2.2 Diffusion in two-phase laminar flows

Mass transfer between two aqueous solutions is dependent on the diffusion at the liquid–liquid interface, because the pressure-driven flow in microchannels is generally laminar. The diffusion time t is expressed with the following formula:

$$t \propto \frac{l^2}{K} \qquad [2]$$

where l is the diffusion distance and K is the diffusion coefficient. Thus, as the diffusion time is proportional to the square of the diffusion distance, it is shown that the diffusion time becomes short in proportion to the square of the miniaturized length.

In addition, the specific interface area, σ, becomes large as the volume of space becomes small.

$$\sigma = \frac{S}{V} \qquad [3]$$

where S is the area of the liquid–liquid interface and V is the volume of the solution. Thus, mass transfer by diffusion is efficiently carried out in microspace.

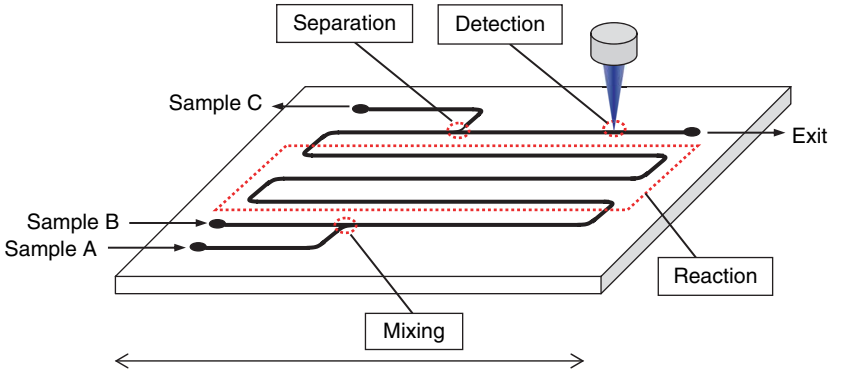


Figure 1 Illustration of a concept of micro total analysis system (μ-TAS).

3.13.2.2.3 Pressure drop

Because fluids are viscous, energy is lost by friction between the flowing fluids and the wall of the microchannel. The effect of the friction shows itself as a pressure loss. The pressure loss is equivalent to the shear stress at the wall P. In a pipe, P is calculated using the Hagen–Poiseuille equation for laminar flow.

$$P = \frac{128\mu QL}{\pi D^4} \quad [4]$$

where L is the distance from the outlet, Q is the flow rate, μ is the viscosity, and D is the diameter. In the case of a microchannel, which has a rectangular profile, the tube diameter D can be replaced using the hydraulic equivalent diameter D_e.

$$D_e = \frac{4A}{U} \quad [5]$$

where A is the cross section and U is the wetted perimeter. Thus, the pressure loss increases with the miniaturization of the microchannel. In order to prevent the breakage of a microsystem and the liquid leak by pressure, this is an important parameter for the design of a microchannel pattern. In this study, in order to perform a gas sampling, porous materials are used. Because these have a low strength when compared with a glass material, it is necessary to take this into consideration for a microsystem design.

The size of the quartz plates used for making microsystems is 50 mm × 26 mm with a thickness of 2 mm. The microchannels in microsystems were fabricated by two methods (wet chemical etching and mechanical fabrication) and different widths of channels having different depths were produced (500 μm × 100 μm, 250 μm × 100 μm, and 100 μm × 40 μm). Detailed procedures for making chemical etching methods have been reported elsewhere. The mechanical fabrication was carried out with the bench micro working machine (Kumakura). In the produced microsystems, two channels for reagents were confluent into the main reaction channel with an incident angle of ∼60°. The top and cross-sectional view of the microsystem (Y-shaped microsystem) used for the experiments are illustrated in **Figure 2(a)** and **2(b)**. The fluidic behavior of reagents was monitored using a commercial digital camera (8-bit resolution) equipped with an inverted optical microscope (Eclipse ME600, Nikon) with a ×10 (NA = 0.3) objective lens (**Figure 2(c)**).

The dyed solutions and counter aqueous solutions were injected into the microsystem using syringe pumps (IC3200, KD Scientific), and syringes were 1 ml gastight (Hamilton). Flow analysis experiments (**Figures 3–5**) were carried out to obtain diffusion phenomena for an aqueous solution of $KMnO_4$ reagent (Wako Pure Chemical), BB3 dye (MP Biomedicals), and arabic gum solution of OB pigment (Holbein Works). The KCl solution of the same density was used as the counter aqueous solution for $KMnO_4$ chemical and BB3 dye and 30% (w/w) arabic gum solution was used for OB pigment solution.

The RGB (red, green, blue) color intensity of the mixing images obtained was monitored and used for the analysis. The laminar flow and molecular diffusion behavior were investigated by visualizing the confluence of some colored and clear solutions. Digital color images from a digital still camera were analyzed resolving the color in the RGB mode. This color contrast is equivalent to the difference in the dye concentration. Numeric data from the RGB color components in the channel images could reveal the diffusion phenomena in the microchannel. τ, a dimensionless number, was used for the analysis of diffusion phenomena (**Figures 6–8**). The microchannel effect of the wet etching process and the mechanically fabricated process are also shown in **Figure 9**.

The concept of lab on a chip is the integration of different steps such as chemical reaction, pretreatment, and detection into a miniaturized space that is a fabricated detailed structure of a micro- or nanoscale on substrates of several centimeters. Such a technology is called μ-TAS or lab on a chip. Recently, various microchip devices have been developed with great progress in microfabrication technology, and advances in microchip analysis have been remarkable. The microchip technology provides a miniaturized analytical system, which leads to downsizing of the experimental equipment as well as high-speed, high-throughput analysis, resulting from on-site monitoring, a small sample volume, and a small experimental space. On the other hand, the studies related to miniaturization of silicon focused on the fabrication of components such as micropumps, microvalves, and chemical sensors.

3.13.3 Phase Transfer Microsystems for Pretreatment of Gas, Liquid, and Solid Samples

The operation of gas absorption is very important in the analytical process as a pretreatment operation for the measurement of gaseous concentration in the atmosphere. The process of gas absorption is used

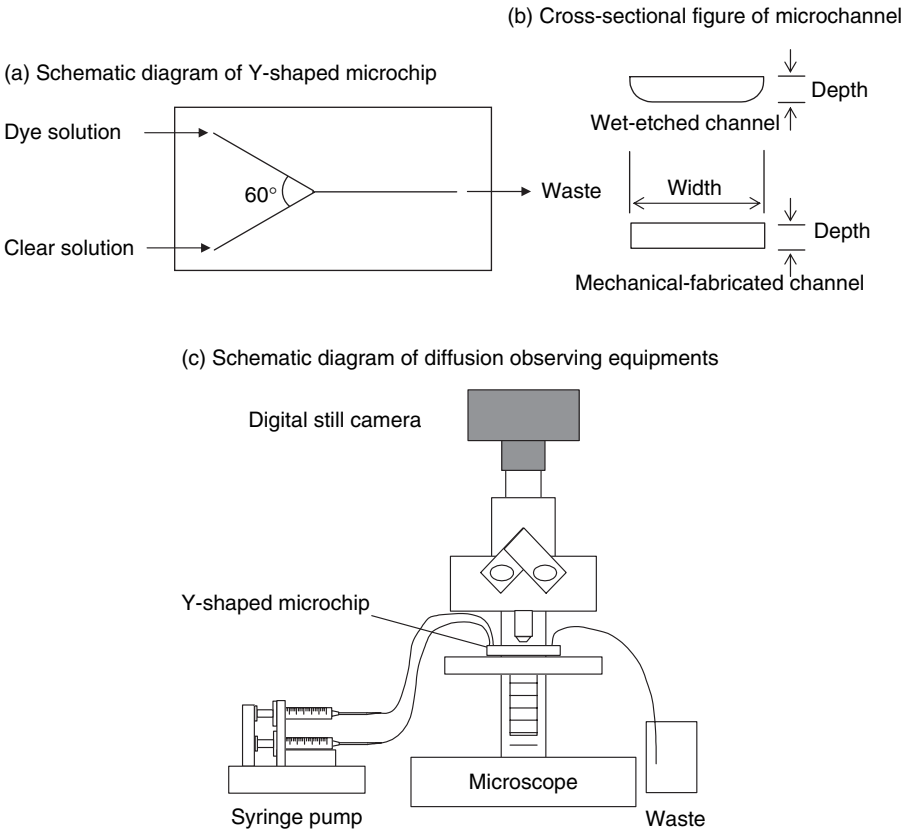


Figure 2 Y-shaped microchip and diffusion observing equipments. (a) Schematic diagram of Y-shaped microchip. (b) Cross-sectional figure of microchannel. (c) Schematic diagram of diffusion observing equipments.

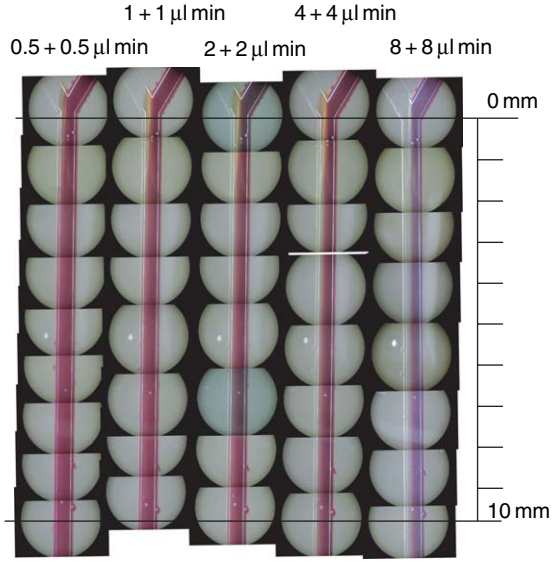


Figure 3 Photographs for diffusion phenomena of $KMnO_4$ reagent in microchannel.

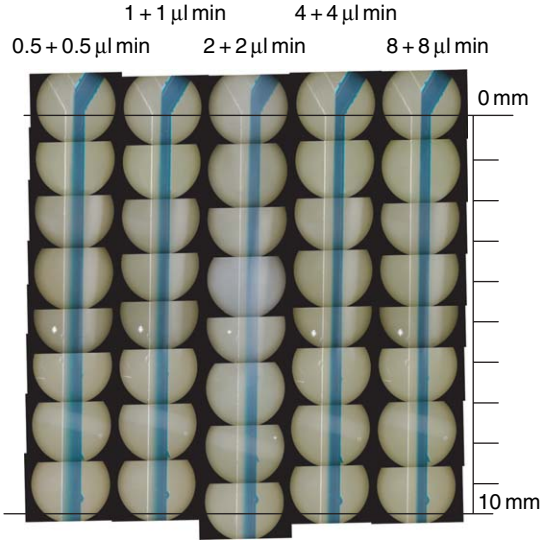


Figure 4 Photographs for diffusion phenomena of Basic Blue 3 (BB3) dye in microchannel.

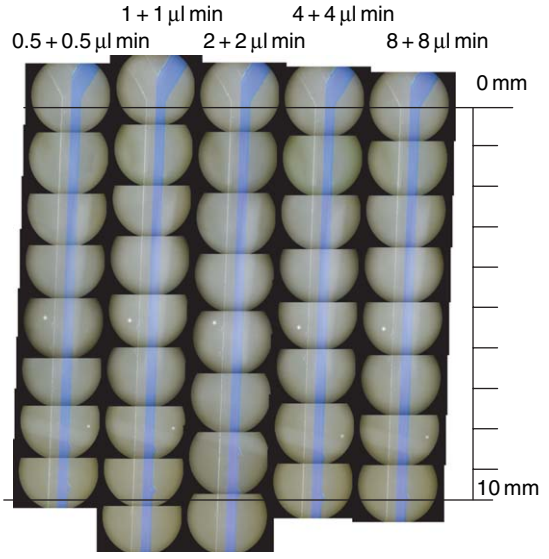


Figure 5 Photographs for diffusion phenomena of Oriental Blue (OB) pigment in microchannel.

in various industries, such as in a chemical plant. Most gas absorption aims at separation of acidic impurities from a mixture of gases. These acidic impurities include carbon dioxide (CO_2), hydrogen sulfide (H_2S), nitrogen oxide (NO_x), hazardous sulfur oxide (SO_x) as SO_2 (Chang *et al.* 2005, Dasgupta and Kar 1995, Liu and Dasgupta 1995a, b, Yang *et al.* 1997), hydrochloric acid (HCl), and organic acid compounds.

NO_2 (Helaleh *et al.* 2001, Korenaga *et al.* 2000, Kumemura and Korenaga 2004a, b, Sato *et al.* 2001, Takabayashi *et al.* 2006), found in the combustion exhaust of automobiles and oil-burning power plants, is a typical example of acutely toxic gases that cause problems such as acid rain, photochemical smog, corrosion, and human respiratory morbidity. The examined value obtained from the Japanese standard ranges from 40 to 60 ppb for the daily average.

A number of methods for measuring NO_2 have been developed in the past few decades. The measurement of NO_2 can be generally classified into wet and dry methods. Incidentally, the continuous analyzers for the concentration of NO_2 in ambient air, which is utilized absorptionmetric method as wet and chemiluminescence method as dry, are prescribed in the Japanese Industrial Standards (JIS).

NO_2 is usually absorbed in a suitable solution first and converted into nitrite; this is followed by the determination of nitrite (Chaube *et al.* 1984). Nitrite is frequently detected using a diazo-coupling reaction, and the spectrophotometric method established

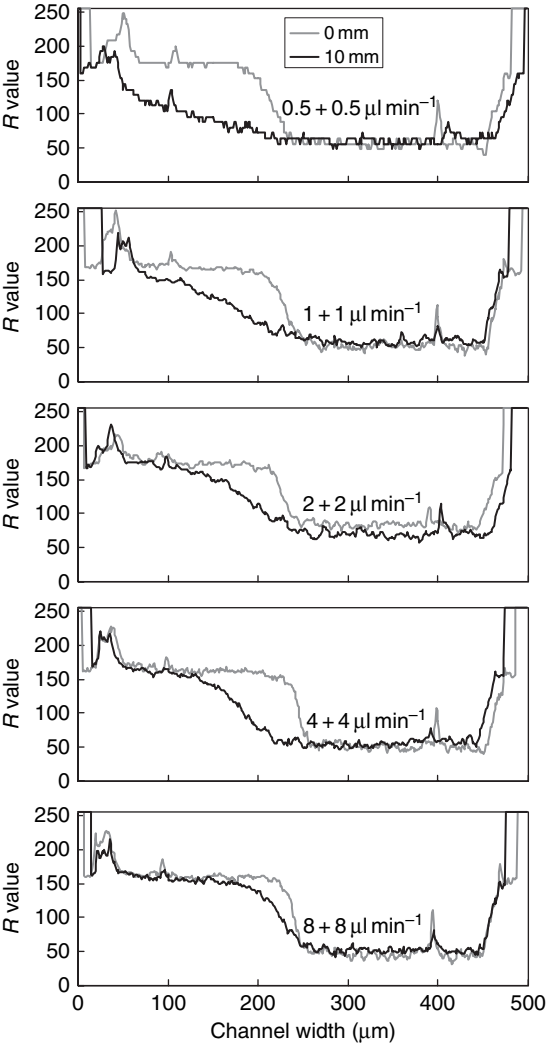


Figure 6 Red, green, blue (RGB) analysis of Basic Blue 3 (BB3) diffusion phenomena in microchannel.

by Saltzman (1954) is used as one of the standard reference methods worldwide. The Saltzman reagent used as a diazo-coupling reagent is a mixture of *N*-1-naphthyl-ethylene diamine dihydrochlorate, sulfanilic acid, and acetic acid. Although the above-mentioned methods are excellent in reproducibility, the sensitivity is poor and they need to extract the azo dye formed to improve the detection limits.

Many methods were developed to detect nitrite ion fluorimetrically, in which fluorescent reagents such as 5-aminofluorescein, 2,6-diaminopyridine, 3-amino-4-hydroxy-coumarin and 3-aminonaphthalene-1,5-disulfonic acid were used (Axelrod and Engel 1975, Dombrowski and Pratt 1972, Motomizu *et al.* 1986, Ohta *et al.* 1986). On the other hand,

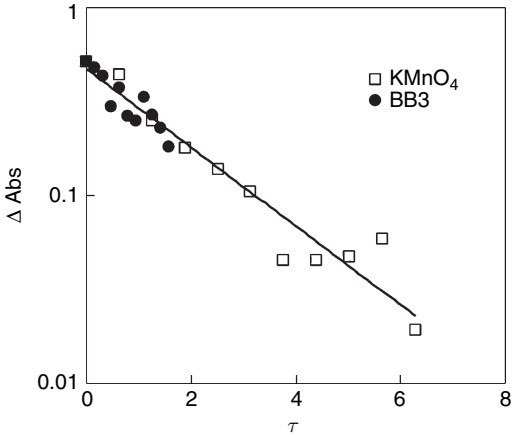


Figure 7 Comparison of $KMnO_4$ and Basic Blue 3 (BB3) by τ where $\tau = tD/R^2$.

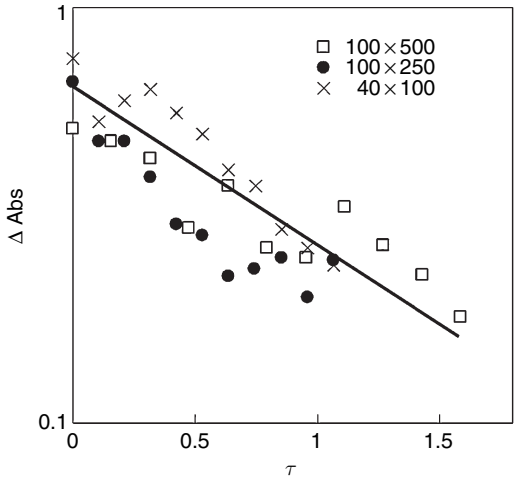


Figure 8 Comparison of diffusion progression by channel size. $\tau = tD/R^2$.

a laser-induced fluorescence detection technique for detection of NO_2 was reported (George and O'Brien 1991). The technique employed a direct determination of NO_2 with high sensitivity (Matsumi *et al.* 2001), but the high cost could prevent it from current monitoring use.

Chemiluminescence methods are also widely applied, which are based on the catalytic or photolytic reduction of NO_2 to NO and subsequent gas-phase reaction with ozone (Fontijin *et al.* 1970, Sigsby *et al.* 1973), or on the chemiluminescence reaction of NO_2 with an alkaline solution of luminol (Maeda *et al.* 1980, Mikuaka and Vecera 1992). These methods are highly sensitive. However, some of the methods encounter interference from species such as SO_2,

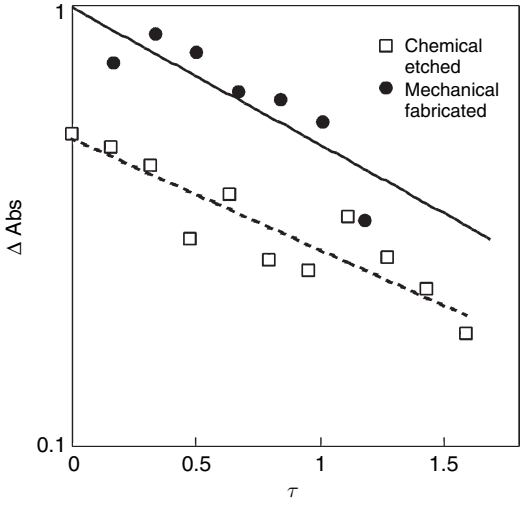


Figure 9 Comparison of diffusion between wet chemical-etched microchannel and mechanical-fabricated microchannel. $\tau = tD/R^2$.

H_2S, CO_2, and O_3. In addition, complex and costly instrumentation is required, and due to the limited stability of the luminal reagent, the systems need to be recalibrated frequently.

A number of small and simple sensors for detecting NO_2 have been reported, such as a solid-state sensor using solid electrolytes (Miura *et al.* 1993), a sensor based on a proton conductor (Albert *et al.* 1996), a sensor detecting the electric resistance using WO_3 as an oxide semiconductor (Akiyama *et al.* 1991, Nelli *et al.* 1996), and a sensor based on a surface acoustic wave (SAW) (Penza and Vasanelli 1997). Tanaka *et al.* 1999 developed a practical NO_2 sensor using a porous glass-doped organic compound, which was a diazo-coupling reagent. However, as the detection system depends on the difference calculus of absorbance, the detection sensitivity is low and it is difficult to adapt it for the measurement of the living environment level. Thus, few sensors suitable for monitoring NO_2 in the atmosphere have been developed as yet because of insufficient sensitivity or gas selectivity.

3.13.3.1 Gas–Liquid Equilibrium and Gas Absorption Pretreatment Microsystems

3.13.3.1.1 The overall mass transfer coefficient

To include the qualitative observations in the correct equations, the example of the gas–liquid interface in **Figure 10** is used. In this case, the gas on the top is

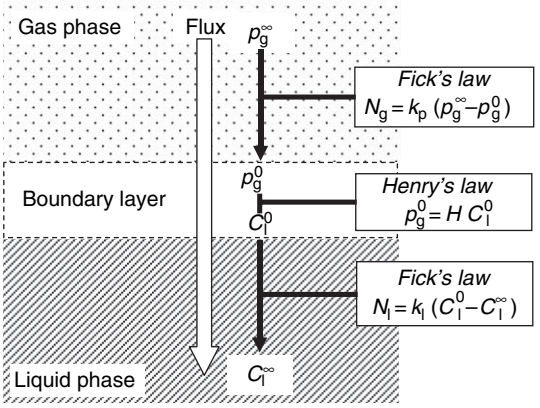


Figure 10 Mass transfer across a gas–liquid interface.

being transferred into the liquid in the bottom. The flux in the gas phase, N, is

$$N = k_p\left(p_g^\infty - p_g^0\right) \qquad [6]$$

where k_p is the gas-phase mass transfer coefficient (typically in $\mathrm{mol\,cm^{-2}\,s^{-1}\,atm^{-1}}$), p_g^∞ is the bulk pressure, and p_g^0 is the interfacial pressure. Because the interfacial region is thin, the flux across is in the steady state and the flux in the gas phase is equal that in the liquid phase. Thus,

$$N = k_l\left(C_l^0 - C_l^\infty\right) \qquad [7]$$

where k_l is the liquid-phase mass transfer coefficient, and C_l^0 and C_l^∞ are the interfacial and bulk concentrations, respectively.

Here, equilibrium exists across the interface.

$$p_g^0 = HC_l^0 \qquad [8]$$

where H is the Henry's constant (here in $\mathrm{cm^3\,atm\,mol^{-1}}$). Combining eqns [6] through [8], the interfacial concentration is found.

$$C_l^0 = \frac{p_g^0}{H} = \frac{k_p p_g^\infty + k_l C_l^\infty}{k_p H + k_l} \qquad [9]$$

And the flux

$$N = \frac{1}{1/k_p + H/k_l}\left(p_g^\infty - HC_l^\infty\right) \qquad [10]$$

k_p is called an overall gas-side mass transfer coefficient and k_l is an overall liquid-side mass transfer coefficient. The expression $1/k_p + H/k_l$ denotes the resistance on the gas absorption, and the reciprocal of that gives the overall mass transfer coefficient on the gas absorption process (Cussler 1997). In this study, the mass transferred because of microsystem-based gas absorption was analyzed based on this theory.

3.13.3.2 Liquid–Liquid Equilibrium and Solvent Extraction Pretreatment Microsystems

In general, three theoretical models can be used when a diffusion phenomenon is employed for analysis. If we consider that a solid chemical dissolves in water, this phenomenon can be analyzed as chemical reaction, mass transfer, and diffusion. The dissolution of a chemical is known to be controlled by the diffusion of the dissolved chemical away from the solid surface of the undissolved materials. If the chemical concentration versus time is measured, the results can be correlated in terms of as few parameters as possible. One way to correlate the dissolution results is to use a mass transfer coefficient. To do this, a mass balance on the solution is expressed as follows:

$$V\frac{dc_1}{dt} = Aj_1 = Ak(c_1(\mathrm{sat}) - c_1) \qquad [11]$$

where V is the volume of solution, A is the total area of the chemical particles, $c_1(\mathrm{sat})$ is the concentration of chemical at saturation and at the solid phase, and c_1 is the concentration in the bulk solution. Integrating this equation allows quantitative fitting of the results with one parameter, the mass transfer coefficient k. This quantity is independent of solubility, the area of the chemical, and the solution volume; however, it does vary with physical properties like stirring rate and solution viscosity. Correlating the effects of these properties turns out to be straightforward. The alternative to mass transfer is diffusion theory, for which the mass balance is given by the following equation:

$$V\frac{dc_1}{dt} = A\left(\frac{D}{l}\right)(c_1(\mathrm{sat}) - c_1) \qquad [12]$$

where l is a known parameter, equal to the average distance across which diffusion occurs. This unknown parameter, called a film or unstirred layer thickness, is a function not only of flow and viscosity but also of the diffusion coefficient itself. Eqns [11] and [12] are equivalent, and they share the same successes and shortcomings. In the former, the mass transfer coefficient must be determined experimentally; in the latter, the thickness l should be determined instead. Those who like a scientific veneer prefer to measure l, for it genuflects toward

the Fick's law of diffusion. Those who are more pragmatic prefer explicitly recognizing the empirical nature of the mass transfer coefficient. The choice between the mass transfer and the diffusion models is thus often a question of taste rather than precision. The diffusion model is more fundamental and is appropriate when concentrations are measured or needed versus both position and time. The mass transfer model is simpler and more approximate and is especially useful when only average concentrations are involved. The way to correlate the results other than the two diffusion models is to assume dissolution is a first-order, reversible chemical reaction. Such a reaction might be described by the following equation:

$$\frac{dc_l}{dt} = kc_l(\text{sat}) - kc_l \qquad [13]$$

In this equation, the quantity $kc_l(\text{sat})$ represents the rate of dissolution, kc_l stands for the rate of precipitation, and k is a rate constant for this process. This equation is mathematically identical with eqns [11] and [12] and so is equally useful. However, the idea of treating dissolution as a chemical reaction is flawed. Because the reaction is hypothetical, the rate constant is a composite of physical factors rather than chemical factors. Therefore, the physical process should be considered in terms of diffusion or a mass transfer model. In this experiment, the mass transfer of Al^{3+}–2,2-dihydroxyazobenzene (DHAB) chelate from continuous phase to TBP droplet was investigated by the mass transfer model.

The droplet volume can be changed by controlling the flow rate in the continuous phase and in the dispersed phase in a T-shaped microchip (**Figure 11**). **Figures 12** and **13** show variation in the extraction amount in different continuous phase

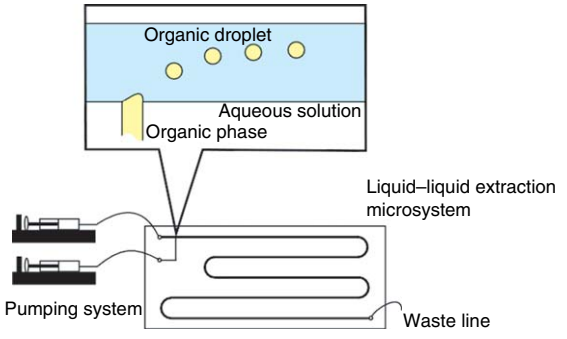


Figure 11 Experimental apparatus for liquid–liquid extraction microsystem from aqueous solution into organic droplet phase.

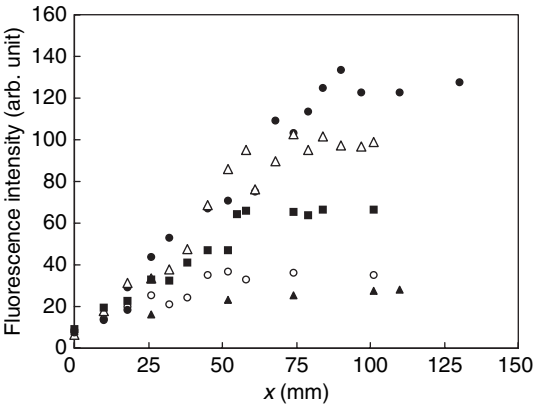


Figure 12 Solvent extraction of Al^{3+}–DHAB chelate at each flow rate of continuous phase in the T-shaped microchannel: 70 μm (wide) × 20 μm (deep) in dispersed phase. x is the distance from the T-junction. Flow rate of continuous phase (mm s^{-1}): ●, 16; △, 26; ■, 37; ○, 47; ▲, 57. Flow rate of continuous phase (mm s^{-1}): 25.

flow rates. The data of **Figure 12** were obtained in the microchip having 70 μm × 20 μm in the dispersed phase and those of **Figure 13** were obtained in the other microchip having 100 μm × 40 μm in the dispersed phase. The plots in both the figures gradually increased and apparently reached equilibrium at each point. The calculated droplet volumes were 0.6, 1.2, 2.7, 5.9, and 11 nl at each flow rate of the continuous phase, which were obtained by substituting the data of **Figure 12**. Smaller droplets could be obtained at faster flow rates in the continuous phase. The

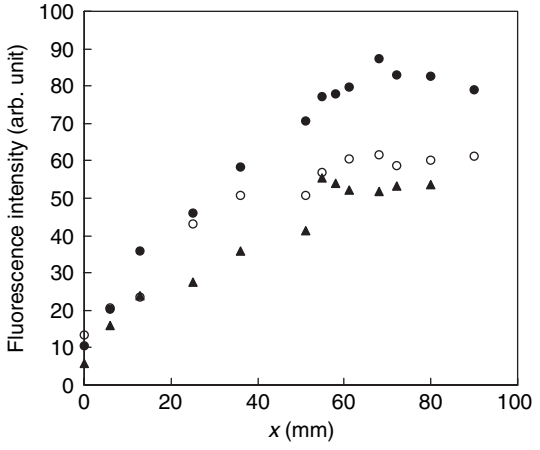


Figure 13 Solvent extraction of Al^{3+}–DHAB chelate at each flow rate of continuous phase in the T-shaped microchannel: 100 μm (wide) × 40 μm (deep) in dispersed phase. x is the distance from the T-junction. Flow rate of continuous phase (mm s^{-1}): ●, 27; ○, 37; ▲, 47. Flow rate of dispersed phase (mm s^{-1}): 53.

fluorescence intensity was increased with increasing distance from the T-junction and reached equilibrium at each flow rate. During the extraction process, only 0.9-s extraction time was sufficient to extract at a continuous-phase flow rate of 57 mm s^{-1}. It was assumed that the difference of fluorescence intensity and the time to reach equilibrium were dependent on the difference in the droplet volume. In order to investigate the mass transfer from aqueous phase to droplet in this experiment, the overall mass transfer coefficient was estimated.

Generally, the equilibrium concentration in the organic phase is given by following equation:

$$C_{o,eq} = kC_{aq,eq} = \frac{kC_{aq,initial}}{1 + k\left(V_o/V_{aq}\right)} \quad [14]$$

where $C_{o,eq}$ is the equilibrium organic-phase concentration, $C_{aq,initial}$ and $C_{aq,eq}$ are the initial and equilibrium aqueous-phase concentrations, V_o and V_{aq} are the organic and aqueous-phase volumes, and k is the distribution coefficient, defined by the following equation:

$$k = \frac{C_{o,eq}}{C_{aq,eq}} \quad [15]$$

When the solute exists in various forms in the aqueous phase, the distribution ratio, D, is expressed as follows:

$$D = \frac{[A_1]_o}{[A_1]_{aq} + [A_2]_{aq} + [A_3]_{aq} + \cdots + [A_n]_{aq}} = \frac{[A_1]_o}{C_{aq}} \quad [16]$$

where $[A_1]_o$ and $[A_n]_{aq}$ are the concentrations of the solute in the organic and aqueous phases, respectively. C_{aq} is the concentration of all forms of the solute in the aqueous phase. When the solute exists as one form in the aqueous phase, D is equal to the distribution coefficient.

$$D = \frac{C_{o,eq}}{C_{aq,eq}} = k \quad [17]$$

Thus, k and $C_{aq,initial}$ must be sufficiently large, and the phase ratio, V_o/V_{aq}, must be reasonably small to avoid detection problems.

The overall mass transfer coefficient, $\bar{\beta}_o$, was obtained from the following equation:

$$\frac{dC_o}{dt} = \frac{A_i}{V_o}\bar{\beta}_o\left(kC_{aq} - C_o\right) \quad [18]$$

where C_o is the concentration of the analyte in the organic phase at time t, A_i is the interfacial area, $\bar{\beta}_o$ is the overall mass transfer coefficient with respect to the organic phase, and C_{aq} is the analyte concentration in the aqueous phase at time t. Assuming rapid transfer across the liquid–liquid interface, the overall mass transfer coefficient can be expressed as follows:

$$\frac{1}{\bar{\beta}_o} = \frac{1}{\beta_o} + \frac{k}{\beta_{aq}} \quad [19]$$

where β_o and β_{aq} are the individual mass transfer coefficients for the organic and aqueous phases, respectively. For compounds with a large diffusion coefficient, k, the mass transfer in the aqueous phase becomes the rate-determining step, and $\beta_{aq} \approx k\bar{\beta}_o$. Thus, β_{aq} can be obtained from the extraction rate curve. The overall mass transfer coefficient was obtained from the fit to the extraction rate curve in **Figure 14**. As a result, the estimated mass transfer coefficients at each volume were not of constant value. There could be two reasons for this. One is the difference in droplet sizes. From the viewpoint of chemical engineering, at a shorter diffusion distance, the concentration gradient became smaller. Namely, for a smaller droplet volume, a molecule is transported faster. Another possibility is the difference in the strength of internal circulation, which occurs because of both the pressure of a fluent shut up in channel wall and the flow of a continuous phase in this experiment. According to the experimental results, the mass transfer coefficient increased with the increase in the flow rate of the continuous phase. The flow rate of dispersed phase was fixed at 25 mm s^{-1}. Therefore, the effect was assumed to occur only because of the flow rate of the continuous phase.

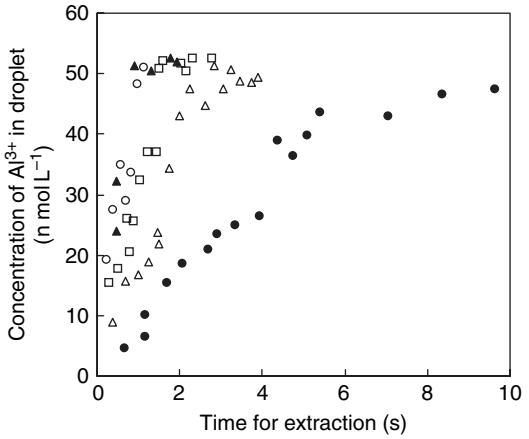


Figure 14 Elimination curve of Al^{3+}–DHAB chelate by this method. Flow rate of continuous phase (mm s^{-1}): ●, 16; △, 26; □, 37; ○: 47; ▲, 57.

3.13.3.3 Solid–Liquid Extraction Pretreatment Microsystems

Polycyclic aromatic hydrocarbons (PAHs) are solid organic compounds with the greatest environmental effects owing to their widespread distribution in the environment; this is because they are generated by natural causes, for example, forest fires and volcanic activity, and by incomplete combustion of fossil fuels and other organic materials. Air pollutants, especially those from diesel exhaust particles, including PAHs, seriously affect human health. Benzo[*a*]pyrene (B[*a*]P), a PAH, has strong mutagenicity and carcinogenicity and is an endocrine disrupter. Moreover, B[*a*]P and benzo[*k*]fluoranthene (B[*k*]F) have caused an immune disturbance that causes allergic diseases, such as asthma and hay fever in developed countries.

The pretreatment microsystem presents an integrated microfluidic system that performs cleanup for PAHs from diesel exhaust particles on silica gel beads in a microchip. A column chromatography phase was constructed by filling the silica gel beads into a microchannel that had a dam structure 25 μm in height. The height of the dam structure was determined according to the rate of wet etching. This work on the cleanup of PAHs from diesel exhaust particles showed that the microchip-based system has the same performance as the conventional method on the solid-phase extraction column and has advantages, such as less reagent consumption and shorter pretreatment time, over the conventional method.

Figure 15 shows a photograph of a microchannel characterized by a dam structure, which shows a magnified view of the dam structure, which was subjected to 3D image processing after images were obtained by a color laser confocal microscope (VK-8500, Keyence). The dam structure at the bottom of the microchannel was clearly seen. The dam structure shown in the image has a height of 25 μm.

Taking into consideration that etching proceeds isotropically, the height of the dam structure was evaluated by using the following equation:

$$h = d - \sqrt{d^2 - \left(\frac{w}{2}\right)^2} \qquad [20]$$

where h is the height of the dam structure, d is the depth of the microchannel, and w is the width of the metal mask not etched. Figure 16 shows a graph of the experimentally determined heights and their theoretical values obtained from the above equation, when the etched depth was 100 μm. As seen from the graph, the experimental values agree well with the theoretical values.

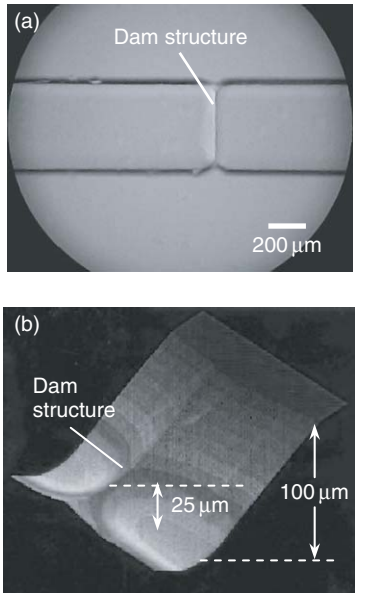


Figure 15 (a) Photograph of the dam structure; (b) 3D image of dam structure.

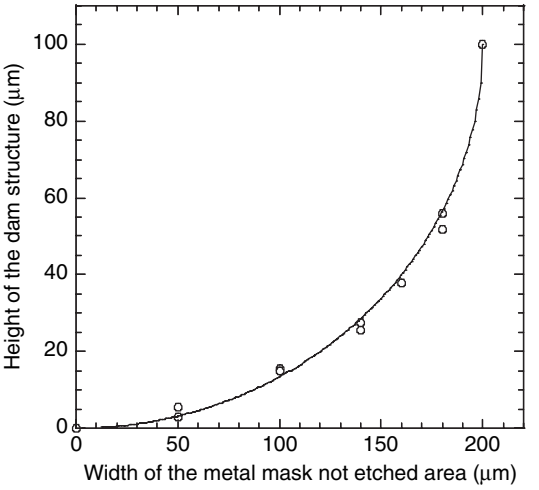


Figure 16 The experimentally determined heights (○) and their theoretical values obtained from the equation (–), when the etched depth was 100 μm.

Consequently, the height of the dam structure can be determined by measuring the width of the metal mask not etched (the width of the masked part in the microchannel at the time when the mask was fabricated) and the etching depth. However, this method for determining the height can be applicable only when the depth of the flow channels is reproducibly

etched, because wet etching processes generally can cause detrimental errors in creating patterns.

3.13.4 Design of Microsystems for Environmental Sample Measurements

The concept of microsystem technology is the integration of different steps such as chemical reaction, pretreatment, and detection into a miniaturized space that is a fabricated detailed structure of a micro- or nanoscale on substrates of several centimeters (Manz *et al.* 1990). Such a technology is called micro total analysis systems (μ-TAS) or lab-on-a-chip system. Recently, various microsystem devices have been developed with great progress being made in microfabrication technology, and advances in microsystem analysis have been remarkable.

The microsystem technology for analytical chemistry in a liquid microspace is especially called a microfluidic system. The concept is that sample treatment, handling, and detection are performed with microflow manifolds by integration of the FIA system. The first realization of microfluidic device was electorophoretic (capillary electrophoresis (CE)) separations in fabricated microchannels on microsystems (Harrison *et al.* 1992). Since this realization of microsystem-based CE separations, many studies dealing with electrically driven separation on a microsystem have been reported, such as CE (Harrison *et al.* 1993), free-flow electrophoresis (Raymond *et al.* 1994), open-channel electrochromatography (Jacobson *et al.* 1994), micellar electrokinetic capillary chromatography (Heeren *et al.* 1996), capillary gel electrophoresis (Effenhauser *et al.* 1994), and sizing of DNA restriction fragments (Jacobson *et al.* 1994). Because electroosmotic flow and electrophoresis are very suitable for manipulation and separation of samples in a microspace such as a microchannel, the study of microfluidic system using electrically driven have been trend, and some of microsystem devices are marketed. However, the electroosmotic driving force can be applied only to a aqueous solution or to a few polar solvents. Moreover, the application of an electric field causes spatial inhomogenization of the components owing to a difference in the mobility. Thus, an electroosmotic driving force is definitely suitable for separation, but not for transportation. Recently, a pressure-driven flow by a pump has been used for transportation in a microfluidic system, and many applications of microfluidic system for general chemical and analytical system, such as solvent extraction (Tokeshi *et al.* 2000), immunoassay (Hadd *et al.* 1999), and organic synthesis (Hisamoto *et al.* 2001a, b), have been presented.

It is expected that microsystem technologies provide technical innovations in various fields, such as medicine, environment, and industry. For example, microarray analysis has become a powerful technology for drug screening, disease gene identification, and signaling pathway studies. In order to utilize microsystem technologies effectively, it is required that the characteristics of microspace be grasped and a design based on theory be created. As mentioned above, microsystem devices for analytical chemistry have many advantages when compared with a conventional chemistry system. In order to enable broader applications of microsystem technologies in the field of chemistry, the selection of suitable materials and fabrication technique is required. Furthermore, the detection of low concentrations of analytes in the small volumes of samples present in microfluidic system requires sensitive analytical techniques. The characteristics of microspace, microfabrication technologies, and detection technologies are described below.

3.13.4.1 Microcolumn with Monolithic Silica for Organic Compounds Separation in Microliquid Chromatography System

A polystyrene–divinylbenzene polymer monolithic (PS–DVB) column was prepared by *in situ* copolymerization of styrene and divinylbenzene as a preconcentration column. First, the wall of the fused silica capillary (250 μm i.d.) was silanized to enable covalent attachment of the monolithic column. After pretreatment, the capillary, which was filled with styrene, divinylbenzene, solvents, and azobisisobutyronitril (AIBN), was heated and washed with methanol. A silica monolithic column was prepared as a separation column by sol–gel reaction. A fused silica capillary (100 μm i.d.) was filled with 1N NaOH and heated at 40°C. Polyethyleneglycol, urea, acetic acid, and tetramethoxysilane were mixed at 0°C. After pretreatment, the capillary, which was filled with the mixture, was heated at 40°C. Then, the capillary was filled with urea and heated. Moreover, in order to improve its separability, a silica monolithic column was used for surface octadecylation. The column prepared above was filled with dimethyloctadecylchlorosilane and heated at 80°C. After washing with tetrahydrofuran (THF),

the column was filled with trimethylchlorosilane and heated at 80°C and washed with methanol. The image of the monolithic silica is shown in Figure 17.

Backpressure indicates the performance of the column, because for FIA, it is necessary to use a pressure pump suitable for the backpressure of the column. If the backpressure is high, a high-pressure pump will be required. Additionally, considering the pressure at the connection point of a microchip and a pump (via a capillary tube), a low-pressure-driven flow is desirable. The backpressure versus the flow rate of sample solution was related to the concentration columns (Chang and Korenaga 2006).

The performance of the separation column was investigated for use of downsized liquid chromatography (LC) with a monolithic silica column (Figure 18). The LC device with a miniatured column requires about 3 ml elution solvent a day in comparison with 500 ml a day for conventional high-performance LC instruments.

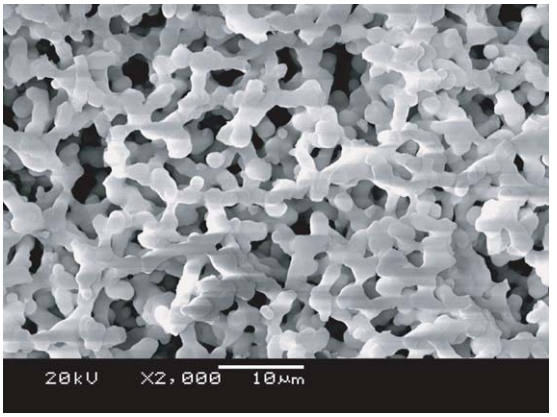


Figure 17 Monolithic silica packing material.

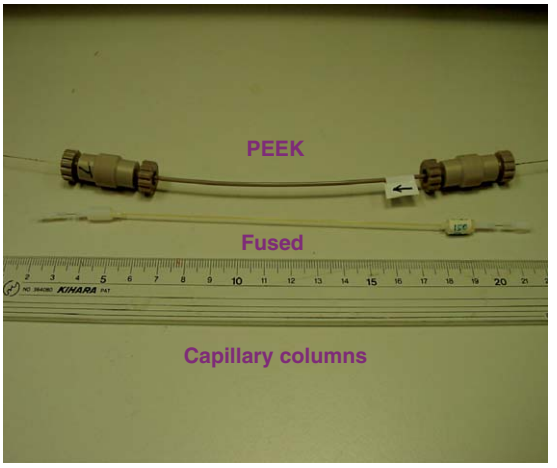


Figure 18 A fused silica capillary and PEEK columns packed with monolithic silica.

3.13.4.2 Microsystem for Inorganic Air Pollutants and Volatile Organic Compounds Monitoring in Gas Samples

A gas absorption microchip that was adapted for NO_2 measurement of a living environment level was designed and fabricated by Takabayashi and coworkers. The results of the basic experiment using a simple gas absorption microchip showed that microchip-based gas absorption was dependent on the gas–liquid contact time related to the absorption area and the flow rate of the gas absorption solution. However, as the form of the gas absorption unit was a huge hexagon-like pool, it turned out that the heterogeneity in a gas absorption rate takes place because of stagnation of the gas absorption solution. In order to prevent the staying of gas absorption solution in the gas absorption unit, it was ideal for all the fluid that flows along several branch lines to gather at a point simultaneously. Therefore, the gas absorption unit was taken as composed of a line and a space, as shown in Figure 19. The gas absorption unit consisted of channels that were 500 μm wide and 100 μm deep; 24 channels were 250 μm wide and 100 μm deep. The internal volume of the gas absorption unit was approximately 6.2 μl. In order to lose the dead volume in the gas–liquid contact, it is desirable that the form of the porous glass plate be the same as the channel pattern (hexagon-like). However, as it was difficult to cut a hexagon precisely, it was decided that a circular porous glass plate be used. Therefore, the actual gas absorption area as the gas–liquid interface was calculated to be approximately 39 mm × 39 mm. The developed micro gas absorption unit was integrated into a microchip for NO_2 determination. The photograph of the structure of the microchip is shown in Figure 20. However, the use of porous glass as the interface of gas and liquid has some disadvantages, for example, there is no suitable bonding technique between the porous glass plate and the quartz substrate. Thus, in order to improve the performance of the proposed method, a poly(tetrafluoroethylene) (PTFE) membrane, which has a thickness of 70 μm, as polymeric porous material, was used. A photograph and schematic illustration of the structure of polymer-based absorption microchip is shown in Figure 21. The microchip was composed of a PTFE membrane and two PMMA

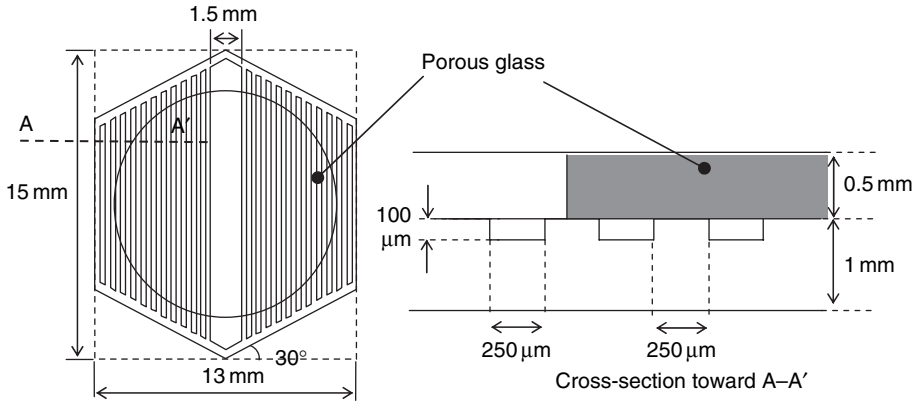


Figure 19 Schematic illustration of channel pattern and cross-sectional view of gas absorption unit.

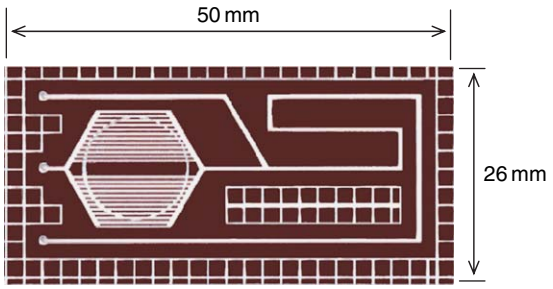


Figure 20 A photograph of NO_2 measurement microchip.

substrates, which are the cover and bottom plates. Each plate was a 12.5 mm × 35 mm rectangle. The thickness, average pore diameter, and porosity of the PTFE membrane (AVDANTEC, FG301) were 70 μm, 0.1 μm, and 64%, respectively. For the bottom plate, a single channel having a width of 1 mm, a depth of 0.1 mm, and a length of 25 mm were formed by mechanical cutting. The penetrated walls were accessible to the PTFE membrane and the inlet and outlet of the absorption solution were mechanically bored on the cover plate. The PTFE membrane was sandwiched between the cover plate and the bottom plate, and these plates were bonded by a thermal adhesive bonding technique. The microchip-based collector using PTFE membrane depends on the effect of wind velocity (ranging from 0 to 1.5 $m\,s^{-1}$), relative humidity (ranging from 30% to 80%), and the form of the absorption unit. The overall mass transfer coefficient was approximately six times higher at 1 $m\,s^{-1}$ than that with no air circulation. For the effect of relative humidity on gas absorption rate, the change of the overall mass transfer

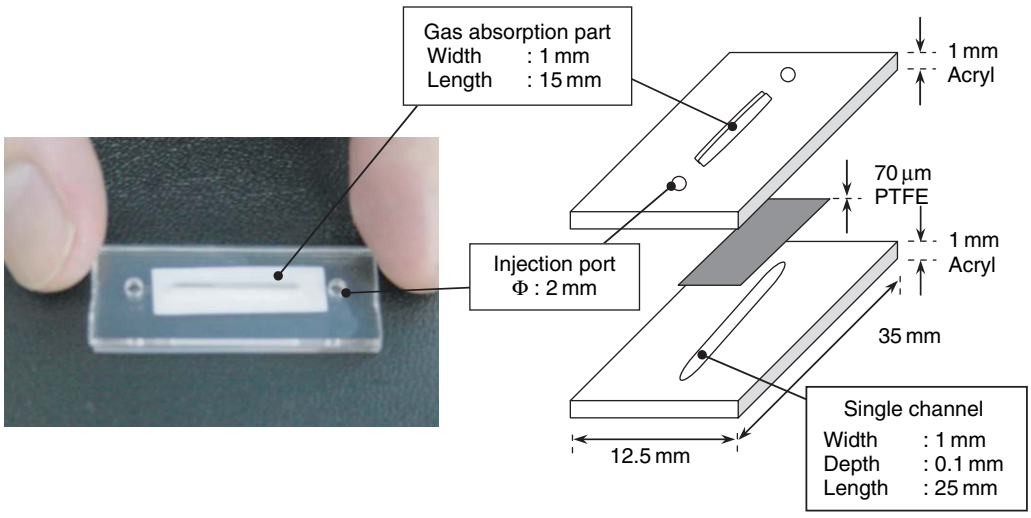


Figure 21 A photograph and schematic illustration of the structure of polymer-based absorption microchip.

coefficient at high (80%) and low (30%) relative humidity was found. However, it was shown that a good linear relationship exists between NO_2 concentration and absorbed NO_2^- concentration under a constant relative humidity. Under optimized condition, the detectable limit was 0.5 ppb. Considering the absorption area of 15 mm × 15 mm and an internal volume of 1.5 ml, the excellent results obtained by this work are promising and will be applied in devices monitoring air pollutants.

For monitoring volatile organic compounds in air, a micro preconcentration unit was developed by fabricating monolithic silica in a fused silica capillary (Chang and coworkers). The micro preconcentration unit is also expected to integrate into a microchip to achieve a microchip-based micro preconcentration system for the monitoring of volatile organic compounds.

3.13.4.3 Microdrop Formation System for Trace Metal Preconcentration and Determination in Liquid Samples

The droplet generation in a microchannel utilizing its geometric pattern was performed and the formation of a droplet during flow in a microchannel was observed. The droplet was formed at the confluent point of the continuous phase and the dispersed phase by flowing two immiscible solutions, water as aqueous phase and tributyl phosphate (TBP) as organic phase. By varying the flow rate of the two solutions, the droplet volume could be controlled. The dynamic formation of TBP was observed in a 600-μm-wide continuous phase and a 300-μm-wide dispersed phase. The shape of the fluid was varied from slug to droplet. On the other hand, in another microchip having a 600-μm-wide continuous phase and a 100-μm-wide dispersed phase, TBP droplets were formed precisely and continuously at 90- to 200-μm diameter in the range of 5–55 mm s^{-1} in the continuous phase and 47–370 mm s^{-1} in the dispersed phase.

A quick extraction system using dispersed droplets in the T-shaped microchannel has been constructed and investigated for determination of aluminum. The droplet was formed in the T-shaped microchannel 600 μm wide and 50 μm deep in the continuous phase and 70 μm wide and 20 μm deep in the dispersed phase by flowing two immiscible solutions, aqueous buffer solution containing aluminum and TBP as organic phase. Using this droplet generation in the microchannel, aluminum in the aqueous phase was extracted with DHAB as a metal chelate into tributyl phosphate for several seconds. The characteristics of extraction in this microchannel, such as extraction time, volume-to-area ratio of the droplet, and mass transfer, were investigated. Furthermore, this extraction system was applied to environmental analysis for evaluation of its validity, and it was found that the result was in good agreement with that from the conventional extraction method.

DHAB has been used as a chelating agent for spectrophotometric detection of metal ions such as Al^{3+}, cobalt cation (α), iron cation (β), vanadium cation (δ), indium cation, and gallium cation. Metal chelates produced by reacting these metals with DHAB have absorbance in the range from 400 to 600 nm. These chelates are highly colored, kinetically inert, and water soluble. Also, DHAB has been used as a chelating agent for spectrofluorometry detecting Al^{3+}, gallium cation, and magnesium cation. These metals can be determined at each excitation (Ex.) and emission (Em.) wavelength: Ex. 485 nm, Em. 560 nm for Al^{3+}; Ex. 470 nm, Em. 580 nm for magnesium. Selective determination of these analytes can be achieved by choosing these specific wavelengths.

The aluminum standard solution containing DHAB was flown as the continuous phase and the TBP was flown as the dispersed phase. The Al^{3+}–DHAB chelate is composed in the continuous phase, and then it is extracted into the TBP droplet. The extraction reaction starts by contact between the TBP and the Al^{3+} standard solution at the confluent point. The variation of fluorescence intensity to pH for a reaction in the microchannel was investigated. Standard solutions and TBP were flown into the T-shaped microchannel and the fluorescence from droplet was detected in each solution. The fluorescence intensity was highest at pH 6.3. It was shown that the fluorescence intensity was lower in acid and alkali media. The variation in the fluorescence intensity obtained from this experiment corresponded with the value from the conventional extraction method that was reported by Watanabe and coworkers. Additionally, the concentration of DHAB was optimized. The concentrations of each solution were 20, 100, 200, and 1000 μg l^{-1}, which corresponded with 0.093, 0.467, 0.935, and 4.672 μmol l^{-1}, respectively. The concentrations of aluminum in aqueous phase were 0 and 500 μg l^{-1}. In a concentration of DHAB at 1000 μg l^{-1}, two peaks were obtained in the spectrum at 524 and 560 nm. It was

Table 2 Equilibrium and kinetic parameters for extraction of Al^{3+}-DHAB chelate from continuous phase into each droplet

Flow rate of continuous phase (mm s^{-1})	*Fluorescence intensity (arb. unit)*	*Droplet volume (nl)*	*Extraction time (s)*	$\frac{dc_o}{dt}(\times 10^{-6})$	$\frac{\beta_o(\times 10^{-4})}{(mm\ s^{-1})}$	*Nernst film (δ_{aq})*
16	122	11	5.4	0.77	0.62	1.1
26	95	5.9	3.0	2.2	1.5	0.41
37	64	2.7	1.5	4.9	2.5	0.25
41	37	1.2	1.1	9.1	3.5	0.18
57	26	0.6	0.91	13	4.5	0.14

Chelating agent – 2,2-dihydroxyazobenzene (DHAB).

assumed that the peak at 524 nm originated from DHAB unreacted with Al^{3+}. The highest peak was obtained from the 200 μg l^{-1} DHAB solution and the intensity of the blank at 560 nm was lower. Therefore, the concentration of DHAB was optimized to 200 μg l^{-1}.

Table 2 shows the data obtained at $x = 40$ mm on the microchannel by PMA. Each peak represented fluorescence from a droplet. The coefficient variation was 2–20%. The flow rate of the continuous phase and the dispersed phase were 37 and 25 mm s^{-1}. The fluorescence intensity of the Al^{3+}–DHAB chelate in the droplet was gradually increased and reached at about 60 mm from the T-junction, which was assumed to be in equilibrium. This variation of fluorescence intensity indicated that the Al^{3+}–DHAB chelate was extracted from the aqueous phase to the droplet by increasing the connection time. The extraction time was about 2 s, which was calculated as the length from the T-junction to the equilibrium point divided by the flow rate of the continuous phase.

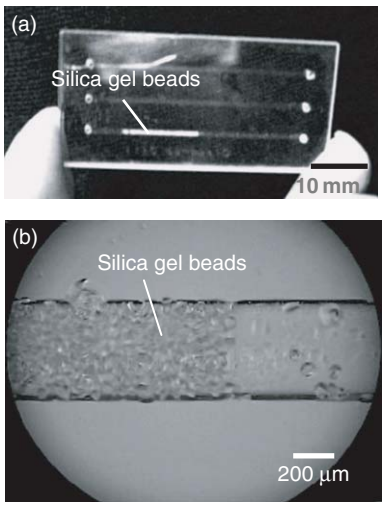


Figure 22 Solid–liquid extraction pretreatment microsystem. (a) A total photograph of the microchip; (b) A microscopic photograph of the microchannel where silica gel beads are packed by the wet etched dam structure.

3.13.4.4 Microsystem for Organic Compounds Pretreatment and Analysis in Solid Samples

As an example of the removal of impurities using a microchip, a monolithic porous column for on-chip solid-phase extraction has been reported. However, the disadvantages of inserting a microchip into a monolithic column are that the duplicability of each chip is bad and the system cannot be used permanently. Moreover, the sample, including fine particles or high-viscosity solution, does not flow properly because of the high pressure in the channel.

Using PAHs containing particles as samples for cleanup operations, the microchip method (Figure 22) and the batch method were compared. The following three samples were used for comparison: diesel exhaust particles that were obtained from the engine whose displacement was controlled by chassis dynamo test, particles sampled with a high-volume sampler at the sides of an expressway, and particles sampled in a tunnel duct. Because a good correlation was found even when the source of particles was changed, the effect of the matrix effect and of the change in the foreign substance on the microchip method would be minimal. In addition, the reproducibility was 5.37% for B[*a*]P and 6.21% for B[*k*]F. These values of reproducibility were almost identical to that obtained with the batch method.

Because the condensation operation is eliminated, the microchip method requires an operation time of not more than one-tenth and a volume of solvent of not more than one-hundredth of that required for the batch method. Furthermore, in the microchip method, silica gels can be activated while being kept in their packed state, and contamination problems at the time of packaging can be resolved.

3.13.5 Development of a Microsystem Detection Device for Environmental Monitoring

In order to detect analytes in a miniaturized space such as a microchannel, a high-sensitivity detection technique is required. For example, for the optical analysis, the size of the space where the analytes exist is important. The small volume in a microchannel is disadvantageous for the analysis of any substance. In order to tackle such a situation, high-sensitivity detection methods are studied and developed. Optical analysis, electrochemical detection, and mass spectrometry (MS) are the detection techniques frequently used for analysis utilizing the microsystem.

Many optical analyses for detection for microfluidic devices have been explored, such as refractive index (Burbbraf *et al.* 1998), absorption (Liang *et al.* 1996), fluorescence (Wang and Morris 2000), and chemiluminescence. Fluorescence detection is mainly used for on-system detection, because it allows selective and sensitive detection. Optical microspectroscopic methods are making rapid progress in their spatial resolution and sensitivity due to the development of microscopic techniques and powerful light sources such as laser. Among them, laser-induced fluorometry (LIF) is one of the most sensitive methods and confocal laser fluorescence microscopy has the highest spatial resolution (Hadd *et al.* 1997, Jin *et al.* 2001). For another optical method, a TLM has been used for ultrasensitive detection of nonfluorescent species in a microchannel (Harada *et al.* 1999).

Electrochemical detection is the technique for measuring the oxidation–reduction current on the surface of an electrode. Metals such as platinum and gold are used as electrodes, and as metal deposition can be easy in a microsystem fabrication process, many examples of applications in microsystems with a built-in electrochemical detection system have been reported. The electrochemical detection methods, such as amperometric detection (Wang *et al.* 1999, Wooley *et al.* 1998), conductivity detection (Lugere *et al.* 2001), cyclic voltammetry (Rossier *et al.* 1999), and electrochemical detection using an electrically isolated potentiostat, are used for microfluidic devices. Because metal layers used as an electrode can be easily deposited in a microfabrication process, electrochemical detection is widely used for microfluidic devices.

3.13.5.1 A Micro Photometric/Fluorometric Detector for Lab-on-a-Chip System

In the field of analytical science, as most of the researches using microsystem have been proposed for analyzing liquid samples, few researches deal with the gas phase. Moreover, there are few reports on the application of microsystem techniques to atmospheric gas analysis because there is no suitable gas sampling method. Because of the large differences between physical parameters of gas and liquid phases, such as viscosity and compressibility, the control of a gas–liquid two-phase flow in a microchannel was relatively difficult. Thus, mostly, gas–liquid two-phase flow on the microfluidic devices is treated as gas–liquid segmented flow. A miniaturized ammonia sensor has been reported (Timmer *et al.* 2004a, b). In this system, gas flow and liquid flow were introduced in the reaction chamber, and an electrolyte conductivity sensor was used for the detector. Tokeshi *et al.* (2002) reported an evaporation utilizing the gas–liquid two-phase laminar flow in the microchannel and applied the technique for the gas analysis system. However, as this system is dependent on special microchannel structures and surface modification, complicated and advanced fabrication techniques are required. In addition, because of the use of TLM for ultrasensitive detection, the system is far from a general-purpose analysis system.

3.13.5.2 A Highly Sensitive Micro Dual-Beam Fluorometric Detector for Lab-on-a-Chip System

A micro dual-beam fluorometric detection device (**Figure 23**) was developed for highly sensitive detection of B[*a*]P and B[*k*]F. The device uses a UV light-emitting diode (UV-LED) that emits light with

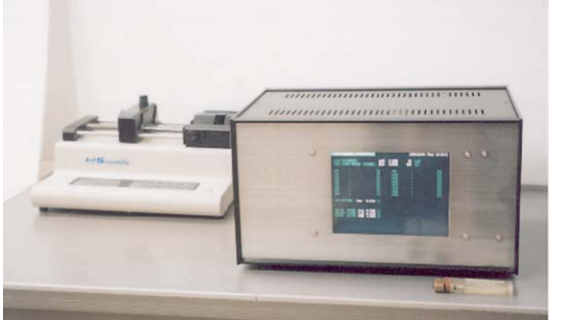


Figure 23 Picture of the micro dual-beam fluorometric detection device.

a wavelength of 370 nm and a violet laser diode that emits light with a wavelength of 395 nm as excitation light sources. The detection method is based on the following observation: characteristic fluorescence is emitted from both B[*a*]P and B[*k*]F, having intensities up to 10 times greater than those from 22 other PAHs. The excitation and emission wavelengths for fluorescence from B[*a*]P and B[*k*]F are 370 and 429 nm, and 395 and 429 nm, respectively.

The microchip was 50 mm in length, 26 mm in width, and 2.0 mm in thickness, and the microchannel was 500 µm in width and 100 µm in depth. All solutions were injected into the microchip using an external syringe pump (IC3200, KD Scientific). The chip used in this study was specifically designed to serve as an optical flow cell considering the simple pattern used in the detection system.

Fluorescence spectra were recorded with a Shimadzu RF-5300PC spectrofluorometer. A modular component liquid chromatographic system was used, which consisted of a Jasco PU-1580 ternary pump, a Jasco DG-2080-53 degasser, a Shimadzu fluorescence detector, and a Jasco RF-10AXL integrator. In the experiments, 20 µl of the sample was injected through a Jasco AS-1555 auto injector. A Supelco 5-mm LC-PAH column (150 mm × 4.6 mm ID) was used for the detection. The experimental results for the determination of B[*a*]P and B[*k*]F in diesel exhaust are given in **Table 3** (Masaki *et al.* 2005).

3.13.5.3 A Highly Sensitive Thermal Lens Microscope for Lab-on-a-Chip Detection

A detection method for the measurement of a microscopic object is required in order to detect specified objects contained in the channel of a chip. A thermal lens microscope was developed by the University of Tokyo as a detection tool for microchip chemistry (Tokeshi *et al.* 2004). The principle of a thermal lens microscope is shown in **Figure 24**. If a material absorbs both an excitation laser and a probe laser through an objective lens, the light energy except the energy absorbed by the sample is all released to the solvent as thermal energy, and the phenomenon produces an increase in temperature. According to the spatial distribution of temperature, a high gradient of the temperature distribution is formed around the optic axis of the laser by the intensity distribution and the thermal diffusion of the laser. In the case of fluids such as water, a lower refractive index at the center of the optic axis and a higher refractive index at the banlieue of the optic axis are formed because the refractive index is lowered by the temperature rise.

The refractive index distribution is equivalent to the concave lens and it is also called thermal lens. The sensitivity of the thermal lens is proportional to the calorific value generated, i.e., it is proportional to the amount and the concentration of a sample; therefore, the quantity of a sample can be determined by measuring the sensitivity of the thermal lens. For an actual measurement, the excitation laser is modulated and the intensity change of the probe laser is simultaneously detected. The subject of ultrasensitive detection of nonfluorescent molecules is extremely important in the field of analytical chemistry and measurement technologies. Although it is presently possible to detect a single fluorescent molecule, it is very difficult to detect nonfluorescent molecules at a single-molecule level. The principle of a thermal lens microscope, which can detect nonfluorescent molecules at the single-molecule level,

Table 3 Comparison of the dual beam fluorometric detector and HPLC/FLD for the determination of B[*a*]P and B[*k*]F in diesel exhaust

	Dual beam fluorometric detector (µg km^{-1})	*HPLC/FLD (µg km^{-1})*
Sample A		
B[*a*]P	3.01	3.25
B[*k*]F	2.98	3.12
Sample B		
B[*a*]P	1.20	1.14
B[*k*]F	1.36	1.46
Sample C		
B[*a*]P	1.65	1.84
B[*k*]F	1.42	1.35

$n = 3$.

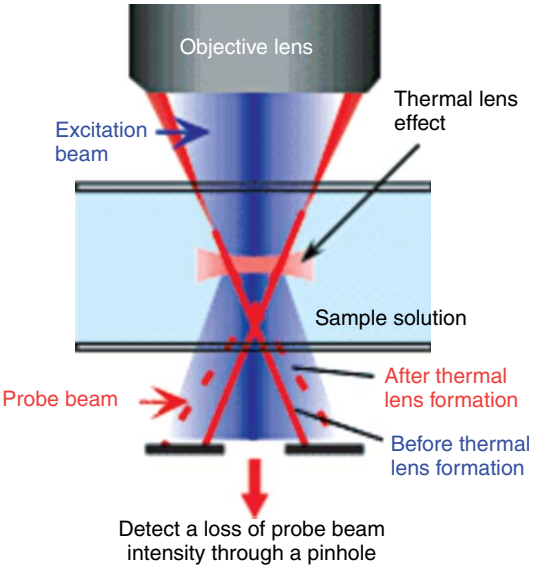


Figure 24 Schematic illustration of a thermal lens microscope.

and its performance and applications are described. A desktop thermal lens microscope and a palmtop thermal lens detection device were also developed.

The thermal lens microscope is a promising measurement method not only for analytical chemistry but also for biological, material, and environmental fields. But the thermal lens microscope setup is relatively large as compared with a chemical microchip. This means that even if a microchip is small, the total analysis setup may become large. A portable size instrument of thermal lens spectroscopic analysis was developed by making use of our fiber optical communication technology.

3.13.5.4 A Downsized Mass Spectrometer for Lab-on-a-Chip Detection

In the field of analytical sciences, a usage at the field of biotechnology and nanotechnology which represents the main themes of the frontier research to elucidate the life sciences, molecular analysis has been heading for not only easily analyzing, but also with higher quality, and understanding the results of analyses for anyone. For its challenging, for example, preparation for their analyses, downsizing and higher quality of instruments have been developed. A downsized mass spectrometer is needed with sample pretreatment microsystems. For the molecular analyses of proteins in the postgenome age, high-performance liquid chromatography–mass spectrometry (LC-MS) or MS/MS are frequently used in the frontier research of proteomics technology.

The downsized MS has been accomplished with a very high-resolution, lightweight, and compact body using a hybrid vacuum system and MEMS technology. We adopted the time-of-flight mass spectrometer (TOF-MS) due to its compact body, high resolution, and very high sensitivity. The property of TOF-MS was tested and the significance of this very small and high-performance instrument was discussed based on specific results (Korenaga 2005).

The adrenocorticotropic hormone (ACTH) was used for the resolution tests for this TOF-MS. The ACTH concentration in the blood plasma reflects the amount of secretion from the pituitary. Thus, ACTH analysis is significant for recognizing the degree and the position of the abnormal part of the pituitary, hypothalamus, and adrenal cortex. The ACTH measurement is indispensable to clarify the diagnosis and the condition. ACTH (0.123 mg) was dissolved in 141 µl of acetonitrile/H_2O (35:65). This solution was diluted to 3.54 pmol μl^{-1}. Four microliters of the diluted solution was used and this was mixed with 4 µl of the matrix solution (CHCA, α-cyanohydroxycinnimic acid). Two microliters of the mixed solution was added on the sample plate. To find the various TOF-MS conditions, the acceleration voltage, the pulse delay of the N_2 laser, and the voltage of the reflectron and of the electronic circuits were changed. Assuming an approximation to the normal distribution-obtained mass spectrum, the gaussian fitting for this mass spectrum was calculated and from the results of fitting, the resolution of ACTH was estimated.

Figure 25 shows a very small TOF-MS. The length of molecular flight is about 50 cm, equipped with a reflectron to reflect molecules and a microchannel plate (MCP). MEMS technology was used in the high-voltage electricity of TOF-MS. The total volume and mass are approximately 14 l and less than 20 kg, respectively.

The mass spectrum of ACTH is shown in Figure 26. The measurement condition was optimized such that the first and the second acceleration voltages and the reflectron voltage were 4.95, 3.35, and 5 kV, respectively. The pulse delay of acceleration after N_2 laser ablation was conditioned to 6 µs. As a result of the gaussian fitting of the ACTH mass spectrum peaks, a surprisingly high resolution of 4966 was obtained.

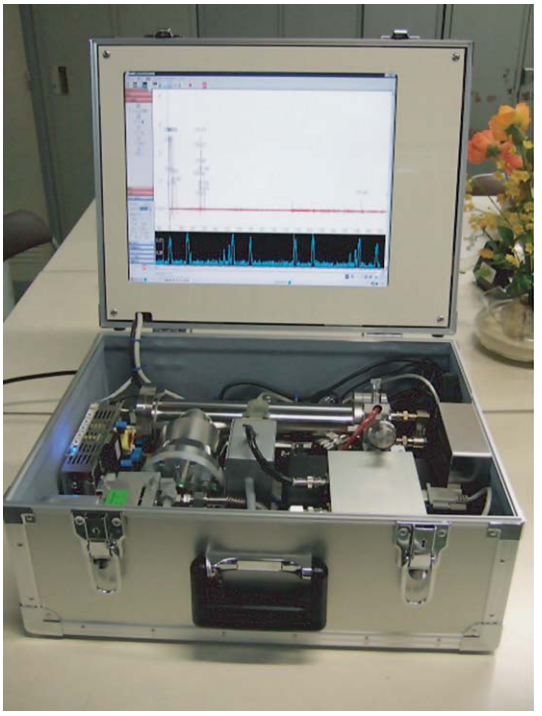


Figure 25 A photograph of the small and high-performance time-of-flight mass spectrometer (TOF-MS) prototype device.

The usage of a high-performance small TOF-MS is described below:

(1) For on-site analysis and on-site monitoring of environmental organic compounds such as dioxins, agricultural chemicals, and halocarbons in water it should be effective.
(2) For explosive organic compounds and drugs, it would also be suitable where many people gather at a public place such as a station or an airport.
(3) For academic purposes such as earth science and archaeology, where it is likely to be accepted.
(4) Probably, for its use in medicine by analyzing proteins as described above with this high-performance TOF-MS. For example, a doctor in a hospital could make a decision immediately by using this.

Futhermore, as the cost performance of this machine could be very high, a quick result without waiting for the tests would be expected in large institutes using very large and expensive machines for a long time. Also, medical researchers at other institutes should be able to use the machine easily. A significant contribution from a high-performance small TOF-MS is expected in the near future.

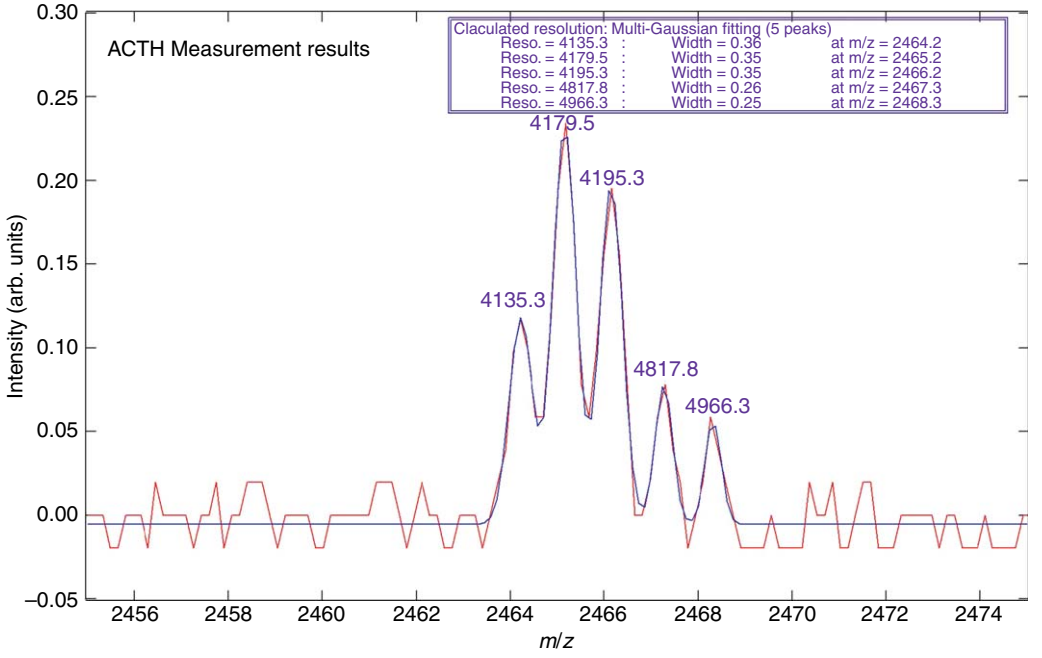


Figure 26 Mass spectrum of adrenocorticotropic hormone (ACTH) obtained by the downsized time-of-flight mass spectrometer (TOF-MS) prototype device: multiple peaks indicate their isotopes from *m/z* 2464 to 2468. High resolution 4966 was observed at *m/z* 2468 as the results of the gaussian fitting.

3.13.6 On-Site and Real-Time Environmental Monitoring Device with Microsystems

The microchip technology provides a miniaturized analytical system, which leads to downsizing of the experimental equipment as well as high-speed, high-throughput analysis, a small sample volume, and a small experimental space. Based on these advantages of the microchip technology, the microsystems have been used for on-site and real-time monitoring of all kinds of environmental pollutants.

3.13.6.1 Industrial Nitrogen Dioxide Gas Monitoring in Atmospheric Environment

A design and a photograph of the NO_2 monitoring microsystem are shown in **Figure 27**. The microsystem integrates three functions:

(1) NO_2 gas absorption
(2) Fluorescent derivatization
(3) Fluorescence detection

The microsystem was constructed by the Technology Research Laboratory, Shimadzu Corp. (Kyoto, Japan). The microsystem substrate consists of a 50 mm × 26 mm quartz glass plate 2.0 mm thick. The microchannels were fabricated using photolithography and wet chemical etching with dimensions ~500 μm wide and 100 μm deep, with a channel width in the absorption region of 250 μm.

The sampling of NO_2 was performed by extracting NO_2 from the atmosphere through the use of a porous glass plate embedded in the microsystem device. It was designed to absorb the gas and then allow it to react with TEA within the microchannel. The NO_2 gas absorption area is a disk-form porous glass plate (Vycor(r) 7930; Corning International K.K.) 11 mm in diameter and 0.5 mm in thickness, having a 28% void volume and a 200 $m^2 g^{-1}$ ratio between its area and weight.

The on-system fluorescence detection system is 320 mm in width, 230 mm in depth, and 177 mm in height. This device is equipped with a UV-LED (NSHU590E; Nichia Corp., Tokushima, Japan) excitation light source of maximum emission wavelength at 370 nm. The observed fluorescence was filtered with a band-pass filter (S76-L42; SURUGA SEIKI Co. Ltd., Shizuoka, Japan) to get the integral fluorescence at wavelengths more than 420 nm, and detected using a photomultiplier tube (PMT). All reagents were injected into the microsystem using syringe pumps (IC3200; KD Scientific) equipped with Hamilton GASTIGHT(r) syringes (25, 100, and 1000 μl). For on-site experiments, it is necessary to keep the wind velocity constant; so the fan to adjust the wind velocity is placed at the outer cover of this device.

An NO_2-absorbing solution, TEA, and a suitable fluorescence reagent were injected into the microsystem via the central and adjacent ports. The flow rate was controlled with syringe pumps. The NO_2 gas is absorbed from the atmosphere into TEA to form the absorbing solution via a porous glass plate. Here, the absorbed NO_2 reacts with water to produce NO_2^-, which is then transferred into the chemical reaction channel, where it reacts with 2,3-diaminonaphthalene (DAN) in acidic solution to produce 1-[*H*]-naphthotriazole (NTA), a highly fluorescing species in alkaline media. At the detection point, NTA is excited at 370 nm (UV-LED) to detect the fluorescence light. For highly sensitive determination, pH adjustment after reaction and before detection are needed, so other systems were used.

In order to determine the optimal conditions for the chemical reaction, the NO_2 absorption area was sealed and the TEA solutions with known NO_2^- concentration were introduced into the microsystem via the central port of system A. Here, the NO_2^- was reacted with the fluorescent reagent (DAN), introduced from a side port on system A, resulting in the formation of NTA. The highly fluorescent NTA was then passed through the PTFE tubing to system B,

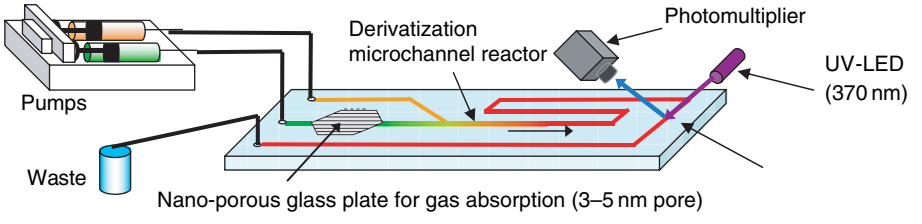


Figure 27 Micro NO_2 monitor prototype device installed with a microchip.

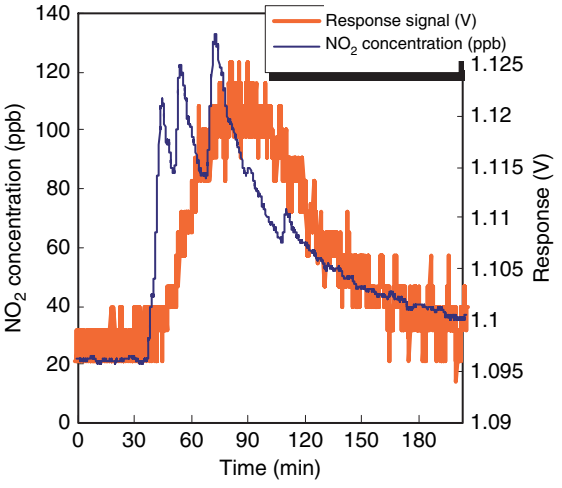


Figure 28 Experimental results for continuous NO_2 monitoring.

where it was combined with the pH buffer solution and delivered to the detection point as shown in Figure 28.

3.13.6.2 Trace Metal Ion Measurement in Aqueous Environment

A schematic illustration of the experimental setup based on laser-induced fluorescence microscopy is shown in Figure 29. Ar^+ laser (488 nm wavelength, 2101-40MLA; JDS Uniphase Corporation, California, USA) was used as the excitation beam source. The excitation beam from Ar^+ laser was introduced into an optical microscope (TE300; Nikon, Tokyo, Japan) via an optical fiber, and, through an objective lens (NA 0.3), was irradiated onto a microchip put on the microscope stage. A dichroic mirror, an absorbing filter and a reflection filter, was placed in the microscope to block off a scattered and reflected excitation beam. A CCD camera, which was mounted on the microscope, displayed a picture image from inside the microchannel. Fluorescence from the extracted solute in the droplet was collected by the same objectives, and led to a polychromator–multichannel photodetector (PMA-11 C7473-36; Hamamatsu Photonics, Shizuoka, Japan) through an optical fiber to record a fluorescence spectrum. The spectrum data were obtained by PMA every 19 ms for 3–10 s, and the variation of fluorescence intensity depending on the contact distance of two solutions was observed by moving the 3D stage of the microscope. The direction from the confluent point to the channel outlet is set to *x*. The performance of this laser-induced fluorescence microscopy system was thoroughly investigated. For the detection of Al^{3+}–DHAB chelate, which is used in this study, a filter block was prepared to block off the scattered and reflected excitation beams. A filter block designed to transmit light at 530–670 nm was used. The transmittance of the filter block was 0% at 488 nm (excitation wavelength, Ar^+ laser) and 88% at 560 nm (emission wavelength), respectively.

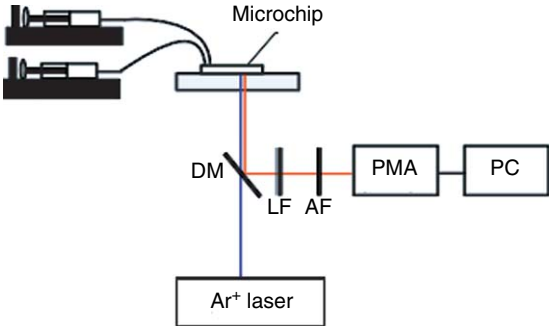


Figure 29 Schematic illustration of experimental setup. AF, absorbing filter; DM, dichroic mirror; RF, reflection filter.

Aluminum exists in our environment; the amount is as much as the oxygen and silicon in the earth. The metal is known to be harmless and easiy to process as an industrial material, and is widely used in industry. However, in recent years, the elution of aluminum as metal ion from soil and rocks, which is especially caused by acid rain, is a problem for plants and microorganisms. Individual chemical formation of aluminum has an effect on this ecological system in varying degrees. Many researchers are engaged in the study in order to clarify the behavior of aluminum in the soil, the chemical formation in the soil and river, or the effect on plants and microorganisms. Tsunoda and coworkers reported the effects of aluminum on the environment. They found the chemical formation of aluminum that existed in rivers. Aluminum exists as ion or compounds with organic matter in lakes in Japan. In the Uzoriko Lake (Aomori, Japan), it was clarified that over 90% of aluminum existed as Al^{3+}, and in the Inawashiroko Lake (Fukushima, Japan), almost all the aluminum formed compounds with organic matter. The chemistry of aluminum in water is complicated, and many chemical parameters, including pH, determine which aluminum species are present in aqueous solutions. In pure water, aluminum has a minimum solubility in the pH range from 5.5 to 6.0; concentrations of total dissolved

aluminum increase at higher and lower pH values. Also, the study of the effects on plants is still continuing. Especially, it is known that growth of the root of a plant is affected by aluminum ions. As described above, the effect of chemical substances on the environment is an important problem that confronts the analyst.

The World Health Organization does not regulate the aluminum in drinking water as a health-based guideline value, owing to limitations in the animal data as a model for humans and the uncertainty surrounding human data. However, they recommend that permissible levels based on optimization of the coagulation process in drinking water and plants using aluminum-based coagulants are 0.1 mg l^{-1} or less in large water treatment facilities and 0.2 mg l^{-1} or less in small facilities.

The maximum value in the extraction of Al^{3+} from aqueous phase to droplet in this microchannel is limited, because the distribution of the amount depends on the ratio of the aqueous phase and the organic phase. If the amount of Al^{3+} in the field sample is large, the extracted amount will saturate in the droplet. For avoiding this problem, some ideas can be proposed. First, the sample solution can be diluted to adjust to a suitable concentration. Second, the extracted amount should be detected before reaching saturation in the droplet. Third, a larger droplet can be used for extraction, which has a larger maximum value in extraction when compared with a smaller droplet. Therefore, a microchip having a larger channel size in the dispersed phase (300 μm × 50 μm) was used for determination. By using this microchip, 18 nl of droplet can be generated. For avoiding saturation, the extraction amount was detected at a fixed position, which is 50 mm from the T-junction.

The calibration curve of Al^{3+}–DHAB chelate was obtained by detection at 50 mm from the T-junction. The correlation coefficient was $r = 0.992$ with a coefficient variation of 5–13%. The detection limit was 2 μg l^{-1} ($\sigma = 3$) and the extraction efficiency was 89%. Generally, the concentration of Al^{3+} in city water was below 80 μg l^{-1}. Therefore, this droplet extraction method has sufficient sensitivity for its application in the determination of Al^{3+} in these samples.

This droplet extraction method was applied for the determination of Al^{3+} in field samples. The sample waters were obtained from the Asakawa River and the Nagaike Pond in Tokyo on August 2004, and city water was also used. These sample waters were filtered with a pore size of 0.2 μm (DISMIC-13cp; ADVANTEC, Tokyo, Japan) and were adjusted to pH 6.3 by adding buffer solution. They were then added to the DHAB solution. As for reference, the conventional extraction method was performed. The result of the determination of Al^{3+} is shown in Table 4. These results show a relationship between the determination by this droplet extraction method and that determinedby the conventional method. The correlation coefficient was 0.92 with the conventional extraction method.

Table 4 Determination of Al^{3+} in water

	This method (μg l^{-1})	*Conventional method (μg l^{-1})*
Asakawa River	55	80
Nagaike Pond	148	155
City water	32	24

Mean ($n = 3$).

3.13.6.3 Polycyclic Aromatic Hydrocarbons Analysis in Solid Samples

An integrated microfluidic system that performs cleanup for PAHs from diesel exhaust particles on silica gel beads in a microchip is described. A column chromatography phase was constructed by filling the silica gel beads into a microchannel that had a dam structure 25 μm in height. The height of the dam structure was determined by the rate of wet etching. This work on the cleanup of PAHs from diesel exhaust particles showed that the microchip-based system has the same performance as the conventional method on the solid-phase extraction column and has advantages, such as less reagent consumption and shorter pretreatment time, over the conventional method.

B[*a*]P and B[*k*]F standards were purchased from Wako Pure Chemicals. Dichloromethane used was from Kanto Chemicals as analytical grade solution. The silica gel beads used for cleanup were C-300HG (40–60 μm, >80%) from Wako Pure Chemicals. The quartz glass plates purchased from Tobita-Rika Glass were 50 mm in length, 26 mm in width, and 1.0 mm in thickness, and they were mechanically polished. Silica gel beads were activated at 120°C immediately before use.

The exhaust gas gathered from a diesel car was processed with a fixed constant volume sampler (CVS) device, and the diesel exhaust particles containing PAHs from the quartz filter were sampled

using a high-volume air sampler. The diesel car was run with a chassis dynamometer to simulate running on a road. The stainless-steel dilution tunnel was 6.5 m in length with a 0.6-m inner diameter; diesel exhaust particles were sampled from the tunnel with a flow rate of 40–120 $m^3\,min^{-1}$. Two 20-mm-diameter disks were cut from the collected filters with a belt punch for PAH analysis, and then 4.0 ml of benzene:methanol (3:1) was added and extracted by ultrasonication. The solution including PAHs was concentrated and the residue taken up in 1 ml dichloromethane solution.

A cleanup operation is one of the pretreatments indispensable for measuring samples collected in an environment. The samples are examined after a pretreatment microchip enables effective cleanup operations when applied to diesel exhaust particles, particularly many foreign substances. Inside the microchip, a dam structure was designed based on calculations so that silica gel can be freely placed and removed, which results in highly reproducible cleanup operations. In **Figure 30**, the cleanup operation is only with solvents extracted from diesel exhaust particles without using an elution solvent. These advantages enable a direct detection of substances immediately after the cleanup operation, and condensation is not required. By employing an appropriate column packing, the range of applications for cleanup operations may be expanded, and complicated cleanup operations can be easily performed in a short time (Masaki *et al.* 2006).

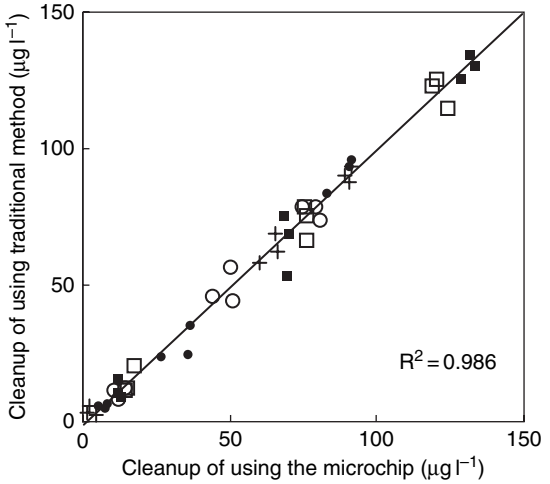


Figure 30 Determination of B[*a*]P (○), B[*k*]F (●), B[*a*]A (■), B[*ghi*]P (□), and DB[*a*,*h*]A (+) by a sample cleanup procedure with the microsystem pretreatment and coventional method in three kinds of particles.

3.13.7 Concluding Remarks

Because microsystem technology has many attractive characteristics, application in a broader field is expected. For environmental monitoring, one of the merits of this technology is miniaturization of analytical instruments that allows on-site and multipoint continuous measurements. In addition, it also saves environmental loads, energy, and valuable resources. A novel gas sampling method was proposed, which is based on gaseous diffusion and gas–liquid contact via a porous material. By the application of this technique to a microfluidic system, a microsystem device for monitoring trace gases will be provided. Additionally, considering that many situations of gas–liquid equilibrium exist in the environment, the technique will have a large ripple effect on the growth of microsystem technology. The basic characteristic of microsystem-based gas absorption was demonstrated. Porous glass was used as the porous material and NO_2 was chosen as the analyte gas. In order to realize the sensitivity of gas absorption, a porous PTFE membrane was also used as a gas–liquid separator in the plastic microsystem. The optimal condition for gas absorption was investigated. Aiming at a practical microsystem for measuring NO_2, a polymer-based NO_2 measurement microsystem was developed. The fluorescence determination of nitrite ion with DAN was demonstrated. The application of the microsystem device to monitoring of NO_2 in the living environment was described.

The monolithic silica synthesized in capillaries and microchannels by the sol–gel process was investigated for all kinds of application in micro analytical systems. With a bimodal pore structure, the monolithic silica has proven to provide better stability and higher performance than conventional columns packed with particles. Monolithic silica can be used without retaining frits or other special structures, especially in micro analytical systems, owing to the stability of the continuous bed. The different kinds of analytical applications of monolithic silica in microchannels showed it was a powerful tool in micro analytical systems.

The possibilities for on-site analysis using the downsized mass spectrometer with chemical instrumentation including microsystems were discussed. A downsized MS system was developed to connect with chemical pretreatment microsystems. The downsized mass spectrometer has been accomplished with very high resolution, lightweight, and compact body using

a hybrid vacuum system and MEMS technology. A chemical microsystem designing was also developed to change easily and quickly from the traditional way such as the method of wet chemistry in the laboratory room. Because the chemical microsystem is currently able to operate separation, extraction, heating, and so on, each design would have a wide variety as required by the operator such as in medical, biological, and environmental studies. Thus, a method using microsystems would be an alternative for wet chemistry in the future. It would have a big potential bringing the data more quickly by the on-site medical, environment, and biological analyses with this system. Measurement and analysis in the nanotechnology and biotechnology field play one of the most important roles for the progress represented as recent proteomics. In the field sample application, MS is often used to determine the configuration of amino acids in proteins and to discover new biomarkers for various diseases.

References

Akiyama M, Tamaki J, Miura N, Yamazoe N 1991 Tungsten oxide-based semiconductor sensor highly sensitive to NO and NO_2. *Chem. Lett.* **237(9)**, 1611–14

Albert G, Cherubini F, Palombari R 1996 Amperometric solid-state sensor for NO and NO_2 based on protonic conduction. *Sens. Actuators B* **37**, 131–4

Axelrod H D, Engel N A 1975 Fluorometric determination of subnanogram levels of nitrite using 5-aminofluorescein. *Anal. Chem.* **47**, 922–4

Burbbraf N, Krattiger B, de Mello A J, de Rooij N F, Manz A 1998 Holographic refractive index detector for application in microsystem-based separation systems. *Analyst* **123**, 1443–7

Chang W, Korenaga T 2006 Application of a monolithic silica capillary adsorbent for the preconcentration of airborne trichloroethylene and tetrachloroethylene. *Anal. Bioanal. Chem.* **385**, 1149–52

Chang W, Ono Y, Kumemura M, Korenaga T 2005 On-line determination of trace sulfur dioxide in air by integrated microchip coupled with fluorescence detection. *Talanta* **67**, 636–50

Chaube A, Baveja A K, Gupta V K 1984 Determination of ultra trace concentrations of nitrite in polluted waters and soil. *Talanta* **31**, 391–3

Cussler E L 1997 *Diffusion, Mass Transfer in Fluid Systems*. Cambridge University Press

Dasgupta P K, Kar S 1995 Measurement of gases by a suppressed conductometric capillary electrophoresis separation system. *Anal. Chem.* **67**, 3853–60

Dombrowski L J, Pratt E J 1972 Fluorometric method for determining nanogram quantities of nitrite ion. *Anal. Chem.* **44**, 2268–72

Effenhauser C S, Paulus A, Manz A, Widmer H M 1994 High-speed separation of antisense oligonucleotides on a micromachined capillary electrophoresis device. *Anal. Chem.* **66**, 2949–53

Fontijin A, Sabadell A J, Ronco R J 1970 Homogeneous chemiluminescent measurement of nitric oxide with ozone. Implications for continuous selective monitoring of gaseous air pollutants. *Anal. Chem.* **42**, 575–9

Fukui S 1963 The filter paper method for measuring sulfur dioxide, nitrogen dioxide, and chloride in atmosphere. *Jpn. Anal.* **12**, 1005–11

George L A, O'Brien R J 1991 Prototype FAGE determination of nitrogen dioxide. *J. Atm. Chem.* **12**, 195–209

Hadd A G, Raymon D E, Halliwell J W, Jacobson S C, Ramsey J M 1997 Microsystem device for performing enzyme assays. *Anal. Chem.* **69**, 3407–12

Hadd A G, Jacobson S C, Ramsey J M 1999 Microfluidic assays of acetylcholinesterase inhibitors. *Anal. Chem.* **71**, 5206–12

Harada M, Shibata M, Kitamori T, Sawada T 1999 Sub-attomole molecule detection in a single biological cell *in-vitro* by thermal lens microscopy. *Anal. Chem.* **15**, 647–50

Harrison D J, Manz A, Fan Z, Lüdi H, Windmer H M 1992 Capillary electrophoresis and sample injection systems integrated on a planar glass system. *Anal. Chem.* **64**, 1926–32

Harrison D J, Fluri K, Seiler K, Fan Z, Effenhauser C S, Manz A 1993 Micromachining a miniaturized capillary electrophoresis-based chemical analysis system on a system. *Science* **261**, 895–7

Heeren F V, Verpoorte E, Manz A, Thormanm W 1996 Micellar electrokinetic chromatography separations and analyses of biological samples on a cyclic planar microstructure. *Anal. Chem.* **68**, 2044–53

Helaleh M, Kumemura M, Fujii S, Korenaga T 2001 A new fluorimetric method for the determination of formaldehyde in air based on the liquid droplet sampling. *Analyst* **126**, 104–8

Helaleh M I H, Ngudiwaluyo S, Korenaga T, Tanaka K 2002 Development of passive sampler technique for ozone monitoring. Estimation of indoor and outdoor ozone concentration. *Talanta* **58**, 649–59

Hibara A, Iwayama S, Matsuoka S, Ueno M, Kikutani Y, Tokeshi M, Kitamori T 2005 *Anal. Chem.* **77**, 943–7

Hisamoto H, Horiuchi T, Uchiyama K, Tokeshi M, Hibara A, Kitamori T 2001a On-System integration of sequential ion-sensing system based on intermittent reagent pumping and formation of two-layer flow. *Anal. Chem.* **73**, 5551–6

Hisamoto H, Saito T, Tokeshi M, Hibara A, Kitamori T 2001b Fast and high conversion phase-transfer synthesis exploiting liquid/liquid interface formed in microchannel chip. *Chem. Commun.* 2662

Jacobson S C, Hergentroder R, Koutny L B, Ramsey J M 1994 Open channel electrochromatography on a microsystem. *Anal. Chem.* **66**, 2369–73

Jin L J, Giordano B C, Landers L P 2001 Dynamic labeling during capillary or microsystem electrophoresis for laser-induced fluorescence detection of protein-SDS complexes without pre- or postcolumn labeling. *Anal. Chem.* **73**, 4994–9

Korenaga T 1992 Aspects of sample dispersion for optimizing flow-injection analysis systems. *Anal. Chim. Acta* **261(1)**, 539–48

Korenaga T 1995 Development of an analysis instrument for convection–diffusion phenomena of the sample solute in capillary flow systems. *Bunseki* (2), 145–51

Korenaga T 1997 Indoor air pollution and its monitoring. *Bunseki* (4), 315–17

Korenaga T 2002 Simple analytical methods: Convenient miniaturized analysis method for atmospheric environment. *Bunseki* (8), 424–9

Korenaga T 2005 *Real-time monitoring device coupled with nanotechnology for environmental pollutants. Bunseki* (10), 550–6

Korenaga T, Stewart K K 1988 Seeking high sensitivity flow injection analysis. *Anal. Chim. Acta* **214(1)**, 87–95

Korenaga T, Yanagisawa Y 1995 Development of a passive sampler for sulfur oxide monitoring. *J. Ecotechnol. Res.* **1**, 98–102

Korenaga T, Shen F, Yoshida H, Takahashi T, Stewart K K 1988 Study of sample dispersion in flow injection analysis. *Anal. Chim. Acta* **214(1)**, 97–105

Korenaga T, Shen F, Yoshida H, Takahashi T 1989a Dispersion behavior of solutes in an ideal laminar flow with small-bore glass tubes. *Bull. Chem. Soc. Jpn.* **62(5)**, 1492–500

Korenaga T, Shen F, Takahashi T 1989b An experimental study of the dispersion in laminar tube flow. *AIChE J.* **35(8)**, 1395–8

Korenaga T, Yoshida H, Shen F, Takahashi T 1989c Peak simulation for non-segmented continuous flow systems. *Trends Anal. Chem.* **8(9)**, 323–6

Korenaga T, Izawa M, Fujiwara T, Takahashi T, Muraki H, Sanuki S 1991a A double-plunger micropump for small flow rate down to several μl/min in flow injection analysis. *Anal. Sci.* **7(3)**, 515–16

Korenaga T, Izawa M, Fujiwara T, Takahashi T 1991b Molecular diffusivity of sample solutes and in a number of eluents and the implications with regard to the design of capillary chromatographic systems. *J. High Resolut. Chromatogr.* **14(6)**, 377–82

Korenaga T, Zhou X, Izawa M, Takahashi T, Moriwake T, Shinoda S 1992 High-sensitivity flow method for the determination of proteins with micro-flow plunger pumps. *Anal. Chim. Acta* **261(1)**, 67–74

Korenaga T, Zhou X, Moriwake T, Muraki H, Naito T, Sanuki S 1994 Computer-controlled micropump suitable for precise microliter delivery and complete in-line mixing. *Anal. Chem.* **66(1)**, 73–8

Korenaga T, Odake T, Ono Y, Rong L 2000 Development of a gas absorption microchip using porous glass and its evaluation for analytical applications. *Bunseki Kagaku* **49(6)**, 423–7

Kumemura M, Korenaga T 2004a Droplets formation from microfluidic device for sampling and measurement of atmospheric nitrogen dioxide. *Analyst* **129(2)**, 105–6

Kumemura M, Korenaga T 2004b Quantitative extraction of Al^{3+} in water using dispersed droplet in T-shaped microchannel. *Anal. Bioanal. Chem.* **8**, 454–6

Kumemura M, Korenaga T 2006 Quantitative extraction using flowing nano-liter droplet in microfluidic system. *Anal. Chim. Acta* **558**, 75–9

Kumemura M, Odake T, Korenaga T 2005 Laser-induced fluorescence microscopic system using an optical parametric oscillator for tunable detection in microchip analysis. *Anal. Bioanal. Chem.* **382**, 992–5

Lee K, Yanagisawa Y, Hishimura M, Spengler J D, Billick I H 1992 A passive sampler for measurement of carbon monoxide using a solid adsorbent. *Environ. Sci. Technol.* **26**, 697–702

Li J, Liu G, Ma H, Korenaga T 1995 Effect of chemical diffusivity on the peak shapes in flow injection analysis with sample or reagent injection. *Anal. Chim. Acta* **310(2)**, 329–36

Liang Z, Chiem N, Ocvirk G, Tang T, Fluri K, Harrison D J 1996 Microfabrication of a planar absorbance and fluorescence cell for integrated capillary electrophoresis devices. *Anal. Chem.* **68**, 1040–6

Liu S, Dasgupta P 1995a Liquid droplet. A renewable gas sampling interface. *Anal. Chem.* **67**, 2042–9

Liu S, Dasgupta P 1995b A renewable liquid droplet as a sampler and a windowless optical cell. Automated sensor for gaseous chlorine. *Anal. Chem.* **67**, 4221–8

Lugere F, Lubking G W, Berthold A, Bastemeijer J, Vellekoop M J 2001 Downscaling aspects of a conductivity detector for application in on-system capillary electrophoresis. *Sens. Actuators A* **92**, 109–14

Maeda Y, Aoki K, Munemori M 1980 Chemiluminescence method for the determination of nitrogen dioxide. *Anal. Chem.* **52**, 307–11

Manz A, Graber N, Widmer H M 1990 Miniaturized total chemical analysis systems: A novel concept for chemical sensing. *Sens. Actuators B* **1**, 244–8

Masaki H, Susaki H, Korenaga T 2005 Development of a micro dual beam fluorometric detector specific for microchip analysis of benzo[*a*]pyrene and benzo[*k*]fluoranthene in diesel exhaust particulate samples. *Analyst* **130**, 1253–7

Masaki H, Hoshi J, Amano S, Sasaki Y, Korenaga T 2006 Integration of the cleanup process: Pretreatment of polycyclic aromatic hydrocarbons from diesel exhaust particulates on silica gel beads in a microchannel. *Anal. Sci.* **22**, 345–8

Matsumi Y, Murakami S, Kono M, Takahashi K 2001 High-sensitivity instrument for measuring atmospheric NO_2. *Anal. Chem.* **73**, 5485–93

Mikuaka P, Vecera Z 1992 Determination of nitrogen dioxide with a chemiluminescent aerosol detector. *Anal. Chem.* **64**, 2187–91

Miura N, Yao S, Shimizu Y, Yamazoe N 1993 Development of high-performance solid-electrolyte sensors for NO and NO_2. *Sens. Actuators B* **13**, 387–90

Motomizu S, Mikasa H, Toei K 1986 Fluorometric determination of nitrite in natural waters with 3-aminonaphthalene-1,5-disulphonic acid by flow-injection analysis. *Talanta* **33**, 729–32

Nelli P, Depero L E, Ferroni M, Groppeli V, Guid F, Ronconi F, Sangaletti L, Sberveglieri G 1996 Sub-ppm NO_2 sensors based on nanosized thin films of titanium–tungsten oxides. *Sens. Actuators B* **31**, 89–92

Ohta T, Arai Y, Takitani S 1986 Fluorometric determination of nitrite with 4-hydroxycoumarin. *Anal. Chem.* **58**, 3132–5

Palmes E D, Gunnison A F 1973 *American Industrial Hygiene Association Journal* **34**, 76

Penza M, Vasanelli L 1997 SAW NO_x gas sensor using WO_3 thin-film sensitive coating. *Sens. Actuators B* **41**, 31–6

Raymond D E, Manz A, Widmer H M 1994 Continuous sample pretreatment using a free-flow electrophoresis device integrated onto a silicon system. *Anal. Chem.* **66**, 2858–65

Rossier J S, Roberts M A, Ferrigno R, Girault H H 1999 Electrochemical detection in polymer microchannels. *Anal. Chem.* **71**, 4294–9

Saltzman B E 1954 Colorimetric microdetermination of nitrogen dioxide in atmosphere. *Anal. Chem.* **12**, 1949–55

Sato T, Susaki H, Iwata T, Ymamoto K, Odake T, Korenaga T 2001 NO2-fluorescence detection system using a quartz microchip. *Trans. IEE J. Sensors Micromach.* **121E(9)**, 507–12

Shoji S, Nakagawa S, Esashi M 1990 Micropump and sample-injector for integrated chemical analyzing systems. *Sens. Actuators A* **21**, 189–92

Sigsby J E, Black F M, Bellar T A, Klostermen D L 1973 Chemiluminescent method for analysis of nitrogen compounds in mobile source emissions nitric oxide, nitrogen dioxide, and ammonia. *Environ. Sci. Technol.* **7**, 51–4

Simon P K, Dasgupta P K 1993 Wet effluent denuder coupled liquid/ion chromatography system: Annular and parallel plate denuders. *Anal. Chem.* **65**, 1134–9

Takabayashi Y, Uemoto M, Aoki K, Odake T, Korenaga T 2006 Development and optimization of a lab-on-a-chip device for the measurement of trace nitrogen dioxide gas in the atmosphere. *Analyst* **131**, 573–8

Takahashi T, Korenaga T, Shen F 1990 A numerical solution for the dispersion in laminar flow through a circular tube. *Can. J. Chem. Eng.* **68(2)**, 191–6

Tanaka T, Guilleux A, Ohyama T, Maruo Y Y, Hayashi T 1999 *Sensors Actuators B* **56**, 247

Timmer B H, van Delft K M, Otjes R P, Olyhuis W, van den Berg A 2004a Miniaturized measurement system for ammonia in air. *Anal. Chim. Acta* **507**, 137–43

Timmer B, Olthuis W, van den Berg A 2004b Sampling small volumes of ambient ammonia using a miniaturized gas sampler. *Lab on a System* **4**, 252–5

Tokeshi M, Minagawa T, Kitamori T 2000 Integration of a microextraction system on a glass system: Ion-pair solvent extraction of Fe with 4,7-diphenyl-1,10-phenanthrolinedisulfonic acid and tri-*n*-octylmethylammonium chloride. *Anal. Chem.* **72**, 1711–14

Tokeshi M, Minagawa T, Uchiyama K, Hibara A, Sato K, Hisamoto H, Kitamor T 2002 *Anal. Chem.* **74**, 1565

Tokeshi M, Mawatari K, Hibara A, Kitamori T 2004 Ultrasensitive detection of non-fluorescent molecules. *Oyo Butsuri* **73(6)**, 741–8

Wang J, Tian B, Sahlin E 1999 Integrated electrophoresis systems/amperometric detection with sputtered gold working electrodes. *Anal. Chem.* **71**, 3901–4

Wang S-C, Morris M D 2000 Plastic microsystem electrophoresis with analyte velocity modulation. Application to fluorescence background rejection. *Anal. Chem.* **72**, 1448–52

Wooley A T, Lao K, Glazer A N, Mathies R A 1998 Capillary electrophoresis systems with integrated electrochemical detection. *Anal. Chezm.* **70**, 684–8

Yanagisawa Y, Nishimura H 1982 *Environmental International* **8**, 235

Yang Y, Zhang X, Korenaga T, Higuchi K 1997 Determination of passive-sampled sulphur dioxide in ambient air as sulphate ion by flow injection analysis with an in-line reaction column. *Talanta* **45(2)**, 445–50

Biography

Takashi Korenaga was born in Okayama in 1949. He attended Okayama University, graduating with a B.S. and M.S. degree in analytical chemistry. He obtained a Ph.D. degree from Kyoto University in 1985 for engineering on automation of water quality monitoring for flow injection analysis. He worked as an industrial process analytical chemistry researcher at Japan Exlan Co., Ltd. from 1974 to 1977. In 1978, he joined the Faculty of Engineering, Okayama University and served as research assistant professor. In 1987, he spent a few months at the Department of Biochemistry and Nutrition, Virginia Polytechnic Institute and State University as a visiting professor. In 1993, he was promoted to associate professor of chemical engineering at Okayama University. Then he was appointed a professor of Toyama National College of Technology from 1993 to 1995. In 1995, he moved to The University of Tokushima as a professor, and then spent a sabbatical leave at the School of Public Health, Harvard University, and also Massachusetts Institute of Technology as a Monbusho Visiting Fellow. He was joined as a professor of the Department of Chemistry, Faculty of Science, Tokyo Metropolitan University from 2001 to the present date.

His specialty and research interest include instrumentation of the downsized analytical chemistry measurement device such as mass spectrometer, fluorimetric detector, and microfluidic chemical and biochemical pretreatment systems.

He was the recipient of The Japan Society for Analytical Chemistry Award for Younger Researchers in 1984, The 26th Japan Environmental Award in 1999, and The Takeda Techno-Entrepreneurship Award 2001. He is President of the Japanese Association of Environmental Safety and Waste Management for Universities and successfully hosted the 8th Asian Academic Network for Environmental Safety and Waste Management in Chennai, India in 2006.

3.14 Medical Applications

Sarah Audet,[1] Jim Carney,[1] Greg Haubrich,[1] David A Ruben,[1] Erik Herrmann,[2] Vincent CMH Larik,[3] Rogier AM Receveur,[3] Erik Scott,[4] and Glen Vaughn[2], [1]Medtronic Inc., Minneapolis, MN, USA, [2]Medtronic Inc., Tempe, AZ, USA, [3]Medtronic Bakken Research Center BV, Maastricht, The Netherlands, [4]Medtronic Inc., Brooklyn Center, MN, USA

Glossary

Acute A medical condition that occurs abruptly and lasts for a relatively short time. It may also mean a condition that requires urgent medical attention

Aneurysm A widening of a blood vessel

Angina pectoris Intermittent pain in the chest due to ischemia

Arrhythmia An abnormal heart rhythm

Arteriosclerosis Hardening and thickening of the arterial walls. A form of arteriosclerosis where plaques are formed within the inner layers of large- and medium-sized arteries

Arthroplasty Surgical restoration of the integrity and functional power of a joint

Astigmatism Defect of an optical lens causing rays from a point to fail to meet at a focal point resulting in a blurred and imperfect image

Auditory Relating to hearing

Biomarker A specific biochemical in the body that has a particular molecular feature that is generated in response to a pathological event, exposure to environmental factors, progress of a disease, or the effects of treatment

BMI Body Mass Index (weight (kg) divided by height squared (m^2))

Bradycardia A heart rhythm that is abnormally slow

CAD Coronary Artery Disease

Cardiac output The average rate of blood flow generated by the heart, usually measured in $l\,min^{-1}$

Cataracts A clouding of the lens of the eye or of its surrounding transparent membrane that obstructs the passage of light

CF_x Carbon Monofluoride

CMOS Complementary Metal Oxide Silicon

Cochlea A division of the bony labyrinth of the inner ear of higher vertebrates that is usually coiled like a snail shell and is the seat of the hearing organ

Cochlear nerve Part of the vestibular nerve associated with hearing

Coronary artery disease The narrowing of the coronary arteries, restricting blood flow to the heart. This condition is usually the result of arteriosclerosis

CRT Cardiac Resynchronization Therapy

CVD Cardiovascular Disease

DELAS Diagnostic Endo Laser Scanner

ECG Electrocardiogram

EMI Electromagnetic Interference

EWL Excess Weight Loss (weight lost divided by the excessive weight)

FCC United States of America Federal Communications Commission

FDA United States of America Food and Drug Administration

Fibrillation A condition of the heart in which there is a complete lack of synchronization of the contraction process, resulting in the loss of cardiac output and in death

GERD Gastroesophageal Reflux Disease

GI Gastrointestinal

Glaucoma A disease of the eye marked by increased pressure within the eyeball that can result in damage to the optic disk and gradual loss of vision

Hypertension Abnormally high arterial blood pressure

IGS Implantable Gastric Stimulation

In vitro In an artificial environment outside the living organism

In vivo Within a living organism

Incidence The number of people per year who are newly diagnosed with a specific chronic condition or have suffered a specific event

Intramedullary Within the bone marrow

IPG Implantable Pulse Generator

Ischemia Deficiency of blood to an organ or a tissue due to functional constriction or actual obstruction of a blood vessel

ISFET Ion-Sensitive Field Effect Transistor

IST Information Society Technologies

IVP Intracorporeal Video Probe

LED Light Emitting Diode

LVAD Left Ventricular Assist Device: a mechanical pump that is surgically connected to a damaged heart to increase the flow of blood to a patient

Macula An anatomical structure (as the macula lutea) having the form of a spot differentiated from the surrounding tissues

MEMS Microelectromechanical System

MI Myocardial Infarction

MIS Minimally Invasive Surgery

Myocardial infarction A localized area in the heart of necrosis due to an interruption of the blood supply to the area

Myoelectrical Utilizing electricity generated by muscle

Necrosis Cell death

OCT Optical Coherence Tomography

Ophthalmic Of, relating to, or situated near the eye
Ophthalmologists A physician who specializes in ophthalmology
Optometrists A specialist licensed to practice optometry
Orthopedics Prevention or correction of injuries and disorders of the skeletal system and associated muscles, joints, and ligaments
Orthotics Use of specialized mechanical devices to support or supplement weakened or abnormal joints or limbs
Otolaryngologists A medical specialty especially concerned with the ear, nose, and throat
PPM Part Per Million
Prevalence The total number of people who are living with a specific chronic condition or have suffered a specific event
Retina The sensory membrane that lines the eye, is composed of several layers including the one that contains the rods and cones, and functions as the immediate instrument of vision by receiving the image formed by the lens and converting it into chemical and nervous signals that reach the brain by way of the optic nerve
RF Radio Frequency
Sinoatrial node Specialized cells in the right atrium of the heart that begin the contraction sequence
ST segment elevation A condition in which the ST segment of the ECG is elevated above the baseline. This is usually associated with an acute myocardial infarction
Tachycardia A heart rhythm that is abnormally fast
USA United States of America
Vestibular nerve Associated with hearing
VF Ventricular Fibrillation
Visual acuity The relative ability of the visual organ to resolve detail that is usually expressed as the reciprocal of the minimum angular separation in minutes of two lines just resolvable as separate and that forms in the average human eye an angle of 1 min
VNS Vagus Nerve Stimulation
Vulnerable plaque A lesion in the wall of the artery that is filled with a soft lipid material and is covered by a thin fibrous cap
Wafer fab The physical facility, including property, plant, and equipment, where integrated circuits are fabricated on silicon, or other semiconductor, wafers. Testing of finished wafers can be performed in the same facility

3.14.1 Clinical Applications of Medical Microsystems

3.14.1.1 Cardiovascular Applications

3.14.1.1.1 Introduction to cardiovascular disease

Cardiovascular disease (CVD) is a large and growing problem in the United States and the industrial world. There are an estimated 71 300 000 people in the United States who have one or more cardiovascular conditions (Thom *et al.* 2006). CVD includes a number of conditions of the heart and the vasculature including coronary artery disease (CAD), arteriosclerosis, heart failure, hypertension, arrhythmia, heart valve disease, congenital heart defects, aortic defects, and peripheral vascular disease. CVD is the cause of common events such as myocardial infarction (MI), cardiac arrest (sudden death), stroke, heart failure, and aneurysms.

CVD is by far the leading cause of death in the United States. **Table 1** shows the extent to which CVD has grown and its devastating effects in lives lost. In 2003, CVD accounted for more than 37% of all deaths, more than the total of the next four causes combined.

This section contains an overview of the current strategies to meet this enormous challenge. In

Table 1 Prevalence, incidence, and number of deaths per year (2003) for many of the manifestations of cardiovascular disease (Thom 2006)

Condition	*Prevalence*	*Incidence*	*Deaths per year*
Coronary heart Disease – total	13 200 000	1 200 000	479 000
Myocardial infarction	7 200 000	565 000 (First MI) 300 000 (Recurrent)	171 000
Heart failure	5 000 000	550 000	57 000
Stroke	5 500 000	700 000	158 000

Section 3.14.1.1.2, the clinical needs and different manifestations of CVD are described. This is followed by a description of the current strategies to diagnose and manage patients each with specific condition. Research to find new ways to diagnose and manage CVD is very active. In Section 3.14.2, some specific microsystems that hold great promise for improving the management of these patients are described.

3.14.1.1.2 Clinical needs for CVD

The clinical needs for managing CVD are associated with two overlapping strategies: increasing life expectancy and improving quality of life. Current medical practice heavily focuses on patients who have already had one or more significant medical events. The treatment is aimed at getting the patient through the acute event, followed by a separate strategy to manage the consequences of the event. In this section, the technologies for diagnosing the diseases as well as the treatment of the diseases are described.

3.14.1.1.3 Current solutions to the management of CVD

3.14.1.1.3.(i) Coronary artery disease CAD is characterized by the narrowing of the coronary arteries sufficiently to prevent adequate blood supply to the heart muscle. If the condition persists, it may lead to cardiac cell death (necrosis) called an MI, heart failure, and/or an arrhythmia. This condition is usually the result of atherosclerosis that causes plaque deposits inside of the coronary artery and may obstruct the flow of blood. The common perception of this condition is of a hard plaque due to cholesterol that may grow over time leading to chest pains (angina pectoris) during exercise or to an MI. Research since the mid-1990s has pointed out that atherosclerosis is actually a much more complex disease (Ross 1999). An inflammatory process resulting from an injury to the arterial wall can cause a lipid-filled region covered by a thin fibrous cap to form. If the cap ruptures, the immediate formation of a clot can cut off all blood flow through that artery. This lipid-filled region is known as a vulnerable plaque (Muller *et al.* 1994) and may be responsible for approximately 70% of fatal acute MIs and the associated sudden coronary deaths (Naghavi *et al.* 2003).

3.14.1.1.3.(i).(a) Diagnosis of CAD In clinical practice today, the typical way for the diagnosis of CAD is in the emergency room or in the physicians' clinic after the symptoms have appeared. Consider a patient who arrives with chest pains or other threatening symptoms. The confirmation can involve a number of tests, each having its own ability to identify people who are having an MI while rejecting those who are not. Today, nearly everyone who shows up at the emergency room complaining of unexplained chest pain will start on a defined protocol (Braunwald *et al.* 2002). One test is the 12-lead electrocardiogram (ECG). Certain characteristics of the ECG, such as ST segment elevation, are highly predictive that the patient is having an MI. Unfortunately, less than 50% of the people coming to the emergency room while suffering an MI will test positive immediately (Slater *et al.* 1987).

Chest pain patients in the emergency room who had a negative ST segment test are given a series of blood tests looking for specific chemical markers indicating cardiac necrosis. When cardiac cells become starved for oxygen or are damaged, they inject different enzymes and other molecules into the blood stream. Elevated levels of such markers of heart damage may help identify an acute MI in a patient with severe chest pain, diagnose ischemia, or triage patients who are in the highest risk due to CAD. In each case, the goal is to point to the correct treatment.

Following an MI, proteins that normally reside within the heart muscle cells are released from the damaged tissue into the blood stream. By measuring the levels of these proteins in the blood, emergency room personnel can quickly determine whether patients have actually suffered an MI. Several key cardiac marker proteins have been identified. Each biomarker has a response time when it can be first measured in the blood. It also has a time at which it reaches its peak and begins to be cleared from the body. **Table 2** shows the response times of various biomarkers after an acute MI.

Myoglobin is one of the first biomarkers to appear in blood after the onset of an MI. Emergency room personnel will often test for myoglobin to get an early indication of an MI but, unfortunately, myoglobin can be elevated in blood due to a number of other conditions. Troponin I takes a few hours longer to appear after the onset of an MI. It is, however, very specific to necrosis of cardiac cells and is becoming the gold standard because of its specificity.

Testing for these markers was historically performed in a central lab using the blood drawn from the patient. Recently, however, a number of companies have introduced small, portable systems that can be used at the point of care. **Figure 1** shows the

Table 2 Cardiac biomarkers used in the diagnosis of acute myocardial infarctions (modified from Braunwald *et al.* 2001)

Cardiac marker	*Range of times to initial elevation (h)*	*Mean time to peak elevations*	*Time to return to normal range*
Heart fatty acid binding protein	1.5	5–10 h	24 h
Myoglobin	1–4	6–7 h	24 h
Troponin I	3–12	24 h	5–10 days
Troponin T	3–12	12 h–2 days	5–14 days
Creatine kinase – MB	3–12	24 h	2–3 days

Figure 1 Triage MeterPlus portable fluorimeter and diagnostic devices from Biosite. (Courtesy: Biosite.)

Biosite® Triage® MeterPlus and four different protein arrays from Biosite Incorporated (San Diego, CA). One of the protein arrays, the cardiac panel, measures three key cardiac markers: myoglobin, CK-MB, and troponin I. Results are delivered approximately 15 min after a blood sample is injected into the panel and the panel is inserted into the meter.

The technology to perform the measurement involves the use of microfluidics, signaling chemistry, and fluorescence. A diagram of one of the protein arrays is shown in **Figure 2**. In order to measure the analytes of interest, a sample of blood is added to the protein array. The blood comes in contact with a fibrous filter inside the array that separates the red blood cells from the plasma. Capillary action draws the plasma through the filter where it enters the reaction chamber. In the reaction chamber, analytes in the plasma bind to specific antibodies that are linked to fluorescent particles. The plasma containing the analytes and antibodies is then pulled by the capillary action into the diagnostic lane that contains discrete zones of immobilized antibodies that bind to any analytes bound to a fluorescent particle. Each discrete zone is scanned by a laser diode inside of the Triage MeterPlus and the fluorescence is measured to quantify the concentration of the analyte that had been present in the sample of blood.

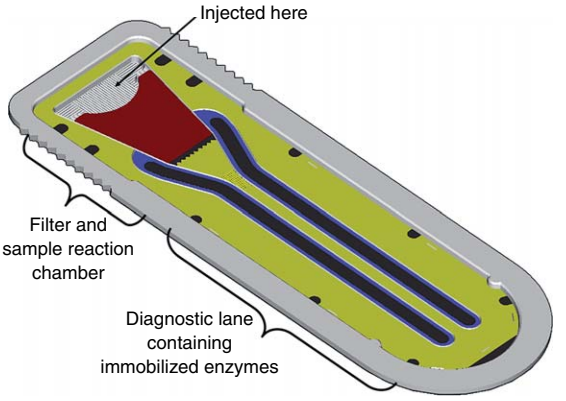


Figure 2 Diagram of protein array used in the Triage MeterPlus. (Courtesy: Biosite.)

3.14.1.1.3.(i).(b) Therapy for CAD The coronary artery bypass graft has been the standard of care to repair a blocked coronary artery. In 2003, an estimated 268 000 (Thom *et al.* 2006) patients had bypass surgery in the United States. This operation involves opening the chest of the patient and cutting through the sternum to expose the heart. The surgeon sews one end of a length of vessel that was removed from elsewhere in the patient above the blockage in the coronary artery, and the other end below the blockage, allowing blood to bypass the occluded region.

In the last few years, the treatment of choice has shifted away from bypass surgery to angioplasty with the placement of coronary stents. Over 660 000 percutaneous coronary interventions were performed in the United States in 2003 (Thom *et al.* 2006). In angioplasty, the physician threads a catheter from an entry point in the femoral artery, through the vasculature to the occluded artery. The catheter contains a deflated balloon at its end. This balloon is inflated once it is inside of the occlusion, effectively opening up the vessel to allow blood flow. In order to keep the vessel open, a stent is usually placed over the balloon before it is inflated. The stent is a tubular wire mesh that, when expanded inside of the occlusion, provides enough structural support to keep the vessel open.

Although far less invasive than the coronary bypass procedure, stenting is plagued by the possibility that tissue or clots will form inside of the stent (restenosis) occluding the vessel and requiring further repair. The prevention of restenosis is an area of aggressive research, involving not only designs of the metal mesh but also the coating of the stent with drugs that are intended to prevent the formation of clots. Restenosis in state-of-the-art devices has been reduced to 5–9% (Lemos *et al.* 2004).

3.14.1.1.3.(ii) Aneurysms One manifestation of CVD is a weakening in the wall of a blood vessel that results in a localized widening of the artery or vein. This widening is called an aneurysm, and it is particularly dangerous if it occurs in the aorta or in a blood vessel in the brain. The risk is that the aneurysm will rupture, causing excessive injury or death. There are an estimated 15 000 deaths per year in the United States (Thom *et al.* 2006) due to the rupture of an abdominal aortic aneurysm (AAA).

Figure 3 is an illustration of an AAA. The diameter of the abdominal aorta below the renal arteries is typically around 2 cm. A diameter of 3.5 cm or greater is classified as an aneurysm. The prevalence of AAA appears to be increasing. The prevalence of AAAs found in screening studies from various countries is about 4% to 9% in men and 1% in women (U.S. Preventive Services Task Force 2005).

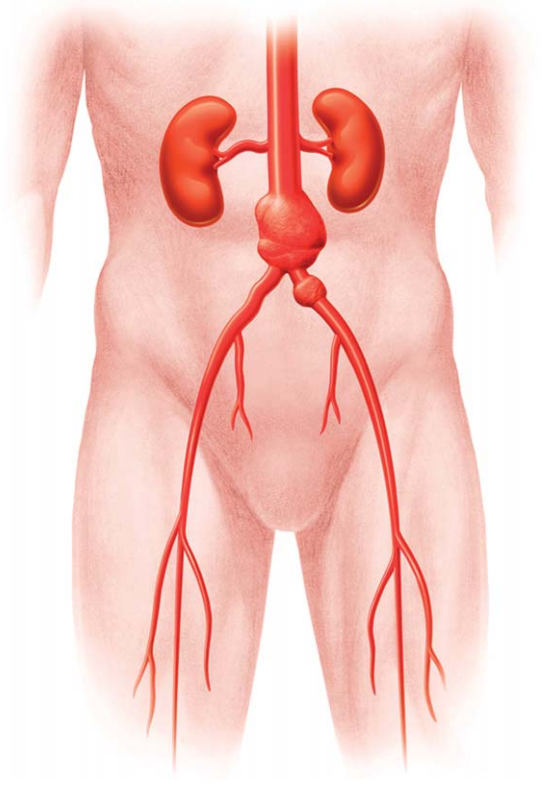


Figure 3 Drawing of an abdominal aortic aneurysn. (Courtesy: Medtronic.)

If the AAA rupture (Harris 1991), the prognosis for the patient is extremely dire. Current statistics indicate that only 10% of the patients are still alive 30 days after AAA ruptures (Kazmers *et al.* 2001).

3.14.1.1.3.(ii).(a) Diagnosis of AAAs The diagnosis of an AAA can be done in a number of ways. Very often, the AAA is discovered on an X-ray or image when diagnosing a completely different condition. This highlights the fact that an AAA usually does not have any symptoms. Due to the high mortality rate associated with an AAA rupture, the United States has passed a law allowing a one-time, free screening test using abdominal ultrasonography for men between the ages of 65 and 75 who have ever smoked. Medicare will also cover the procedure for men and women with a family history of AAA (US Preventive Services).

3.14.1.1.3.(ii).(b) Therapy for AAAs Historically, open abdominal surgery was the standard treatment for AAAs that had grown to the size of 5.5 cm or larger. In this procedure, the damaged section of the aorta was removed and a graft made from Dacron or other fabric was placed in its place. The repair, while usually successful, was also associated with an on-table mortality of about 3.8% (Lee *et al.* 2004). In 1991, the first endovascular stent graft was used to repair an AAA (Parodi *et al.* 1991). An illustration of a stent graft that was placed to repair an AAA is shown in **Figure 4**.

A stent graft is a long tube made of a thin metal framework (stent) covered by a thin material such as Dacron or polytetrafluoroethylene (PTFE). The physician places the stent graft via a catheter that is inserted into the femoral artery in the patient's groin. Once the stent graft is in position, the stent portion is expanded against the sidewall, effectively channeling all the blood through the inside of the stent graft. This effectively excludes the aneurysm and removes the pressure from the aneurysmal wall. Although this approach is far less invasive than the open surgical approach, the current designs of stent grafts may not be 100% successful at excluding the aneurysm. If blood continues to leak around the stent graft or if the graft develops a leak, the patient remains at risk. Hence, monitoring of this repair is done using X-rays once or twice per year.

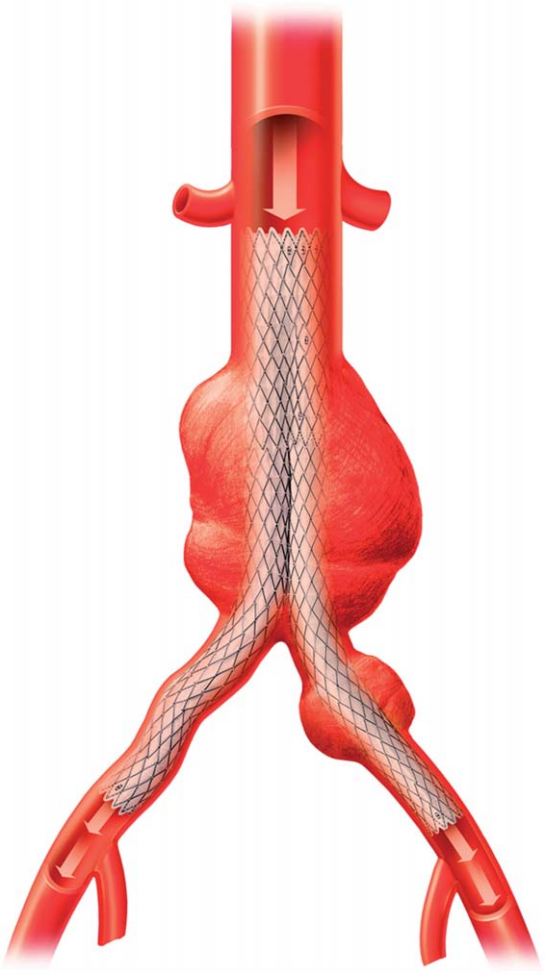


Figure 4 Drawing of stent graft placed in an abdominal aortic aneurysm. (Courtesy: Medtronic.)

3.14.1.1.3.(iii) Arrhythmias The heart's rhythm is controlled by an electrical impulse that is generated from a clump of cells in the right atrium called the sinoatrial (SA) node. These cells trigger the heartbeat based on the changing environment including oxygen in the blood, presence of specific chemicals, and neural inputs. If these cells become damaged or if cells other than the SA node begin to trigger heartbeats, then arrhythmias can result. Bradycardia, or slow heartbeat, can result from damage to the SA node or the cells that conduct the signal to the rest of the heart. Implantable devices called pacemakers have been developed to take over the job of starting the heartbeat at a rate required to meet the needs of the patient.

On the other end of the scale, various disease states or drugs can cause the heart to beat too fast. This is called tachycardia. If the heart rate becomes exceedingly fast, the ability of the heart to pump blood decreases and the patient can become symptomatic. In ventricular fibrillation (VF), the contraction process of the heart is completely uncontrolled and little or no blood is pumped. If this happens, the patient may die suddenly, i.e., in a matter of minutes. Of the 330 000 sudden deaths from coronary disease, most are believed to be from VF (Thom *et al.* 2006).

3.14.1.1.3.(iii).(a) Diagnosis of an arrhythmia Arrhythmias are diagnosed using instruments that can measure the electrical activity of the heart. The standard is to measure an ECG using electrodes that are pasted on the chest of the patient.

In some patients, arrhythmias are not present all the time and can be missed during a physical examination at the medical clinic. These patients are given a Holter monitor to wear for 24 or 48 h. A Holter monitor is similar to an ECG machine but usually has only two or three electrodes on the patient. The signals generated between the electrodes are stored in the memory of the Holter. Once the Holter is returned to the physician, the memory is analyzed to determine whether any arrhythmias are present.

The Reveal® Insertable Loop Recorder (Krahn *et al.* 1999) takes this concept one step further. A photograph of the Reveal® recorder is shown in **Figure 5**. It is about the size of three sticks of gum and is inserted underneath the skin in the chest of a patient who has infrequent but recurring fainting spells. The Reveal® recorder measures the electrical activity of the heart and, when the patient faints, stores a few seconds of the rhythm for review by the managing physician. This was the first implantable device for a monitoring-only application.

Figure 5 Photograph of the Reveal® Insertable Loop Recorder. (Courtesy: Medtronic.)

3.14.1.1.3.(iii).(b) Therapies for arrhythmias The standard of care for treating bradycardia is an implantable pacemaker. There are approximately 200 000 pacemakers implanted each year in the United States (Thom *et al.* 2006). As shown in **Figure 6**, a pacemaker is implanted under the skin in the pectoral region. The pacemaker is a device about the size of a matchbook that is electrically connected to the heart through wires, called leads, that enter the large vein just below the collarbone and advance into the heart. There are two leads in the illustration. One lead is attached to the wall of the right atrium and the other is attached to the right ventricle.

The pacemaker contains a significant amount of processing capability and other electronics to perform multiple functions. It senses the activity of the heart and can also deliver an electrical stimulating pulse to start the heart contraction. Most pacemakers today also have the ability to estimate the rate at which the patient's heart should be beating in order to supply sufficient blood flow to meet the demands of the patient's activity level. This capability is called rate-responsive pacing and requires the input from one or more sensors. The most prevalent method of sensing the patient's demand is through the use of a vibration sensor built into the pacemaker itself. This sensor may be a small piezoelectric crystal attached to the side of the pacemaker or an accelerometer attached to the pacemaker's circuit board. In either case, daily activities like walking or running will generate vibration signals picked up by the sensor. An algorithm inside of the pacemaker converts this level of physical activity into a heart rate for the patient.

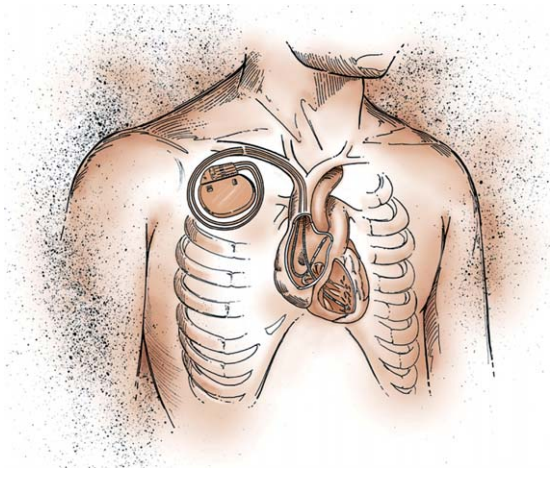


Figure 6 Illustration of implanted pacemaker with two leads into the heart. (Courtesy: Medtronic.)

Figure 7 is a photograph of the inside of a rate-responsive pacemaker. The accelerometer is attached to the circuit board inside the pacemaker. The large silver object that covers the right half of the pacemaker is the battery. These batteries are primary cells, so when they are depleted, the pacemaker must be replaced. Advances in electronic circuitry and battery technology have made it possible to add valuable features to the pacemaker for therapy and diagnostics while simultaneously increasing the lifetime of the device. Currently, pacemakers generally last between 6 and 12 years before needing to be replaced.

Implantable cardioverter defibrillator (ICD) is a more recent invention than the pacemaker (Mirowski *et al.* 1980). The ICD is intended to monitor the electrical activity of the patient's heart, looking for rhythms that can be classified as either VF or life-threatening tachyarrhythmia. If either of these conditions is detected, the ICD will provide appropriate electrical therapy to try to break the arrhythmia.

In 2003, approximately 64 000 ICDs were implanted in the United States (Thom *et al.* 2006). This number is growing rapidly as new classes of high-risk patients are identified and the therapy becomes more accepted.

3.14.1.1.3.(iv) Heart failure Heart failure is a condition in which the heart is compromised to such an extent that it can no longer pump enough blood to meet the body's requirements. The condition may be the result of various causes but is usually associated with damage to the heart muscle

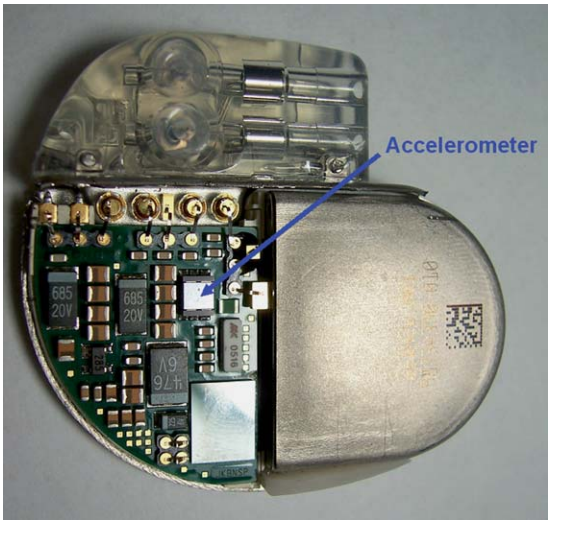


Figure 7 Photograph of the internal circuitry, accelerometer, and battery of a pacemaker. (Courtesy: Medtronic.)

associated with an MI, infection, or congenital defect. When this happens, the patient will have symptoms such as fatigue, shortness of breath, inability to sleep comfortably at night, or lung congestion.

3.14.1.1.3.(iv).(a) Diagnosis for heart failure

Most patients are diagnosed with heart failure after the onset of symptoms that brings them to their physician or the emergency room. Heart failure can be confirmed through the use of history and physical, chest X-ray, and 2D echocardiography. Recently, the B-type natriuretic peptide (BNP) blood test has become a standard method to diagnose and manage heart failure (Morrison *et al.* 2002). BNP is a substance secreted from the ventricles, or lower chambers of the heart, in response to changes in pressure that occur when heart failure develops and worsens. The level of BNP in the blood increases when heart failure symptoms worsen, and decreases when the heart failure condition is stable.

If a patient has severe or unstable heart failure, the managing physician may decide to place a catheter in the patient's heart to more closely monitor the heart's performance. The catheter is a thin, flexible tube that is inserted through one of the large veins (the inferior or superior vena cava) that return blood to the heart. A small balloon at the end of the catheter allows the catheter to follow the flow of blood through the right side of the heart and into the pulmonary artery. Once there, the catheter is purposely positioned in a small branch of the pulmonary artery and a balloon at the end of the catheter is inflated to stop the blood flow through the pulmonary branch. This allows the measurement of a pressure called the pulmonary capillary wedge pressure in front of the temporarily inflated and wedged balloon.

An alternative to the cardiac catheter to measure intracardiac pressure is a small catheter that contains a silicon pressure sensor. **Figure 8** is a diagram of a pressure catheter that is only 0.36 mm in diameter manufactured by Radi Medical (Uppsala, Sweden). This catheter is small enough to fit into small blood vessels to measure pressure. It is also inserted through coronary occlusions in order to assess the effect that the occlusion have on blood flow.

Figure 8 Illustration of an ultrathin PressureWire® sensor. The diameter of the guidewire carrying the sensor is 0.36 mm. (Courtesy: Radi Medical Systems.)

A tiny silicon pressure sensor die is mounted on a small circuit board inside of the catheter, as shown in the close up diagram of **Figure 9**. The silicon sensor is side-mounted a few centimeters back from the tip of the catheter and is protected by a polymer coating.

3.14.1.1.3.(iv).(b) Therapy for heart failure

It is generally accepted that heart failure is a disease that can be controlled for a time, but not cured. One possible exception to this is the heart transplant in which a patient's damaged heart is replaced by a healthy heart from a donor who has died. In 2004, however, only 2016 heart transplants were performed in the United States (Thom et al. 2006).

The left ventricular assist device (LVAD) is a mechanical device that can overcome the issue of shortage of hearts for transplant. An LVAD, such as the HeartMate II from Thoratec (Pleasanton, CA, USA), is an auxiliary pump implanted in the chest of a patient with an extremely damaged heart. The input of the pump is surgically attached to the left ventricle and the output is inserted into the aorta. The effect is to off-load the left ventricle while providing an adequate blood supply to the patient. The LVAD has shown itself to be of value as a bridge to heart transplant (Long 2001) and has recently been approved for a long-term use in those patients who are not eligible for a heart transplant (Rose *et al.* 2001). However, their use is still very limited because of the size, surgical invasiveness and cost of LVADs.

The standard of care for heart failure patients is through various medications that are intended to improve quality of life and increase life expectancy. Quality of life is affected by optimizing the pumping capability of the heart. The heart is a pump and has the property that increasing the input pressure increases the output of the pump. This relationship is called the Frank–Starling curve and is illustrated in **Figure 10**.

Figure 9 Illustration of silicon PressureWire® sensor inside recess in guidewire from Radi Medical Systems. (Courtesy: Radi Medical Systems.)

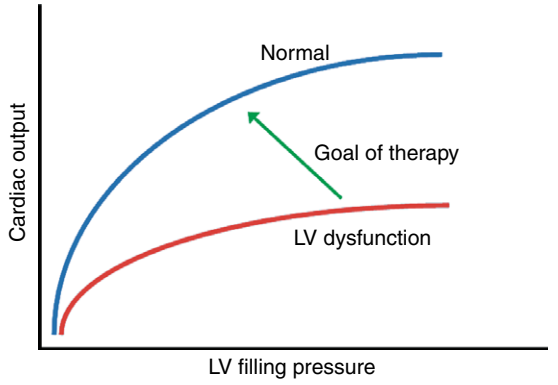


Figure 10 Schematic representation of the relationship of the Cardiac output to left ventricular filling pressure for a normal heart and one with a dysfunctional left ventricle. (Courtesy: Medtronic.)

If the heart is damaged and its pumping capability is diminished, the cardiac output at a specific input pressure is reduced. The body, however, has compensatory mechanisms that attempt to achieve the desired cardiac output by increasing the input pressure. This is done by shutting down the kidneys and increasing the volume of blood in the system, resulting in an increasing input pressure. Unfortunately, the Frank–Starling curve for the damaged heart shows only a minor increase in the cardiac output for the increase in pressure. This high pressure on the input of the heart is also applied to the lungs. If the pressure is high enough, the fluid is forced out of the blood, through the thin tissue membrane that separates the pulmonary blood from the air sacs, and into the lungs. The result is called congestive heart failure and can often lead to a patient coming to the emergency room with extreme breathing difficulty. The goal of the physician is to keep the patient's symptoms under control in order to avoid hospitalization. The first line of treatment for heart failure is drug management. Diuretics activate the kidneys and remove excess fluid, reducing the filling pressure. Vasodilators reduce the vascular resistance and allow blood to flow more easily. The desired outcome is to move the patient's Frank–Starling curve up and to the left.

In the last few years, there has been an emergence of a new type of implantable device called cardiac resynchronization therapy (CRT), which paces both sides of the heart approximately simultaneously to coordinate the contraction and to improve the pumping ability (Leclercq *et al.* 2000). The clinical results for CRT have been very impressive. In the MUSTIC randomized clinical trial (Cazeau *et al.* 2001), not only were there improvements in the hemodynamic measurements, but those patients who received CRT had approximately two-thirds fewer hospitalizations than the control group.

3.14.1.1.4 Cardiac solutions on the horizon

Because of the severity of the problem of CVD, cardiovascular research is extremely active. This section describes a number of areas of active research and new product development.

3.14.1.1.4.(i) Diagnostic tools for CAD The applications for cardiovascular catheters containing sensors continue to increase. The ultraminiature pressure wires described in Section 3.14.1.1.3.(iv).(a) are now being used to monitor the effectiveness of using a stent. In addition to measuring whether an occlusion has stopped blood flow in a coronary artery, the pressure pulse in a coronary artery measured by these catheters can be analyzed, providing an estimate of the level of occlusion by calculating indices such as fractional flow reserve (Bech *et al.* 2001).

A second device that continues to improve is the intravascular ultrasound (IVUS). The device provides images that are used to guide the placement of stents and to assess the procedure afterward (Schiele *et al.* 1998). The original versions of this catheter contained a single ultrasound element that was rotated at high speed to obtain a view of the vessel wall and lesion (Elliott and Thrush 1996). Volcano Therapeutics Inc. (Rancho Cordova, CA, USA) has developed an advanced version of this catheter, as shown in **Figure 11**. It contains 64 elements and has higher resolution, giving it the ability to determine the structure and the composition of the interior of the vessel walls. This is extremely helpful in the identification of vulnerable plaques.

3.14.1.1.4.(ii) Monitoring of AAAs The quality of stent grafts for protecting patients from AAAs continues to improve. In addition, the screening of

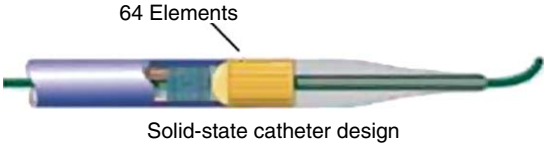


Figure 11 IVUS catheter from Volcano Therapeutics containing an array of 64 ultrasound elements. (Courtesy: Volcano Therapeutics.)

AAA has been recommended as a reimbursable procedure (US Preventive Services Task Force 2005). Therefore, more AAAs will be diagnosed and more patients will get AAA stent grafts. Unfortunately, the stent grafts are not perfect and a percentage of these devices do not seal correctly, maintaining a dangerous level of blood pressure in the aneurysm. A number of companies have begun to investigate the use of small, chronically implanted pressure sensors to monitor the pressure inside of the aneurysm. **Figure 12** is a wireless pressure sensor manufactured by CardioMEMS (Atlanta, GA, USA). This device is about the size of a small paper clip and is placed in the region of the aneurysm that is excluded by the AAA graft at the time the graft is implanted. This sensor consists of a silica-based capacitor that varies as a function of pressure and an antenna that doubles as an inductor forming a resonant circuit. The resonant circuit is interrogated using an external instrument that can precisely determine the resonant frequency. This resonant frequency is translated to pressure and the physician can determine whether the pressure inside of the excluded region of the aneurysm has dropped or is still dangerously high.

3.14.1.1.4.(iii) Advanced therapy for arrhythmias

Currently, there is a significant amount of activity directed at improving the algorithms used by an ICD to determine whether or not the patient is having a lethal arrhythmia and whether a shock should be delivered. Much of this effort is directed at looking for different sensors that can provide a measure of the hemodynamic status of the patient. St. Jude Medical (Sunnyvale, CA, USA) presented an abstract that evaluated the use of an optical device integrated into the side of an ICD (Weiss *et al.* 2006). The optical device is intended to identify the presence or the absence of pulsatility in the arteries using the light emitted into the underlying tissue by light emitting diodes (LEDs). This device is in the investigation stage but points out the need to improve the ability of ICDs to accurately identify fatal arrhythmias.

Figure 12 Implantable, ultrasound-based pressure sensor. (Courtesy: CardioMEMS.)

3.14.1.1.4.(iii).(a) Advances in heart failure therapy

The CRT device has proven to be an extremely effective device in the treatment of heart failure. One issue that it faces, however, is that the time delays between the delivery of the electrical pulses to the heart have a significant effect on the performance of the device (Porciani *et al.* 2005). The optimization procedure is currently a time-consuming process involving the use of echocardiography or other techniques to visually confirm that the selected timing resynchronizes the contraction of the left ventricle. The use of implanted sensors that could perform the same function would be a significant advancement in this procedure. The Sorin Group (Via Crescentino, Italy) announced the NewLiving CRT device that contains a sensor for automatic optimization. This device includes a small accelerometer at the end of the lead that measures peak endocardial acceleration (PEA™). The PEA sensor was originally developed to provide rate response to pacemakers (Langenfeld *et al.* 1998) by measuring a value closely correlated with the strength of the heart's contraction. This sensor was designed into a CRT device called NewLiving chf and is used to automatically optimize the timing of the device by maximizing the strength of contraction.

3.14.1.1.4.(iii).(b) Advances in heart failure diagnostics and monitoring

In addition to the effort to improve heart failure therapy devices, there is a significant effort in the development of devices for monitoring heart failure patients. Because heart failure is a chronic condition, many of these devices are intended to be implanted permanently in the patient. One such device is the Chronicle® implantable hemodynamic monitor (**Figure 13**) from Medtronic Inc. (Minneapolis, MN, USA). The Chronicle system consists of an implanted pacemaker-sized monitor and a pressure-sensing lead placed in the right side of the heart. Pressures inside

Figure 13 Chronicle pressure sensing monitor for the management of heart failure. (Courtesy: Medtronic.)

Figure 15 Photograph of implantable, battery-powered pressure sensor with ultrasound-based communication. (Courtesy: Remon Medical.)

the heart are measured continuously and stored in the device. The data are then intermittently transmitted by the patient to a secure website accessible by treating clinicians. Information from the Chronicle system, together with other patient-specific data as available, such as symptoms, daily weight, and physical signs, are used to guide treatment decisions. (Bourge *et al.* 2005).

A close up of the Chronicle pressure sensor is shown in **Figure 14**. The sensor is a titanium housing that has been machined to have a thin diaphragm in one section. This diaphragm is one plate of a capacitor and deflects due to cardiac pressure and produces a change in the capacitance. This change is measured by a circuitry inside the capsule and is transmitted to the Chronicle device for processing and storage.

Other companies are also developing implantable pressure monitors to manage heart failure patients. The device, shown in **Figure 15**, is a pressure sensor being developed by Remon Medical (Caesarea, Israel) that is intended to be placed in the pulmonary artery of a patient suffering from heart failure. The device is an advanced version of the wireless pressure sensor described in Section 3.14.2.1. This device contains a microelectromechanical system (MEMS) pressure sensor, an ultrasound transducer, an integrated circuit (IC), a miniature battery, and a Remon's proprietary acoustic switch (Penner and Doron 2003). This sensor is dormant until a short low-intensity ultrasonic activation pulse is sent to it from a master device that is either inside or outside the patient's body. The ultrasound pulse flexes the piezoelectric ultrasound transducer, generating sufficient energy to close a switch that connects the battery to the circuit. The device then starts measuring the pressure and transmitting the measurements using acoustic communication. After a predefined time (10 s in the pulmonary artery sensor), or upon a deactivation pulse sent from the master, the implant becomes dormant again. The miniature battery can last for more than 10 years at two to three measurements per day without charging.

Figure 14 Close-up of Chronicle® pressure sensor capsule. (Courtesy: Medtronic.)

The measurement of cardiac pressures using a batteryless implanted sensor that is interrogated using an external device has attracted a number of competitors. ISSYS (Ypsilanti, MI, USA) is developing a small pressure sensor that is powered and interrogated using a radio frequency (RF) external signal. Savacor (now St. Jude Medical) has also developed a pressure sensor that is inserted into the right atrium and poked through the interatrial septum into the left atrium.

The chronically implanted accelerometer and pressure sensors are the first entries into implantable systems for disease management. It should be expected that, in the not too distant future, new sensors such as flow sensors or blood chemical sensors will appear to either allow the managing physician to better prescribe medications or to control a therapy-providing

device. Providing this new level of control will allow patients with CVD both to live longer and to live with a higher quality of life.

3.14.1.2 Neurological Applications

Neurology is the medical study focusing on wellness and disorders of the nervous system. Physicians specialized in the field of neurology are called neurologists, and neurosurgeons are specialized neurologists with neurosurgical training.

Some of the diseases and conditions of the neurological system include autonomic nervous system disease, basal ganglia disease, brain tumors, cerebral palsy, cerebellar disease, cerebrovascular disease, cervical spondylosis, CNS infections, coma, dementia, demyelinating diseases, epilepsy, hydrocephalus, migraine, movement disorders, multiple sclerosis, neurological syndromes, neuromuscular disorders, Parkinsonism, phakomatoses, peripheral neuropathy, speech and language disorders, spinal muscular atrophy, and tension headache (GP Notebook 2006).

Over 50 million people in the United States suffer from chronic pain and another 25 million Americans experience acute pain each year (MedTech Insight, February 2005). The estimated cost associated with pain including health care costs, disability, and loss of productivity exceeds $100 billion annually in the United States (MedTech Insight 2005a, b). In addition to pharmaceuticals, medical devices based on neurostimulation and drug delivery are available to provide treatment for patients suffering from acute or chronic pain. Neurostimulation devices are also market-released to provide treatment for patients suffering from Parkinson's disease and epilepsy. Applications of microsystems in these three areas of neurology are discussed below.

There is a growing awareness of the connection between the heart and the brain, and the field of neurocardiology is emerging. The Earl and Doris Bakken Heart Brain Institute was created at the Cleveland Clinic with the goal of advancing the understanding of the heart–brain connection (The Cleveland Clinic 2006).

It should be noted that in chronically implanted medical devices, the implant materials must adhere to stringent requirements, especially those related to biostability. The interface between the foreign material and its surrounding tissue becomes stable within 1–2 weeks after implant, which then provides continued and unchanged performance for the duration of the implant. This is true for electrical stimulation, and also for (unblocked) delivery of drugs at the targeted sites. This implies that the choice of materials is restricted to only a few known biocompatible metals and polymers. All the materials in transient or in permanent contact with tissue and body fluids must adhere to the biocompatibility standard (ISO 10993).

To achieve uninterrupted performance, reliability standards are applied, and components are being subjected to a very rigid assessment. Very often, this implies the design of specific components, not commercially available components. Most critical are batteries and feedthroughs to achieve hermetic-sealed enclosures in the harsh biological environment. Implanted battery-powered devices require a carefully designed power management to achieve the longevity goals and to reduce the replacement (surgical) frequency.

3.14.1.2.1 Chronic pain

Depending upon the underlying cause, electrical stimulation with a neurological stimulator or local drug delivery with an implantable drug pump provides methods for alleviating pain. While the pump provides pharmacological intervention, an electrical stimulation system blocks sensory signals (**Figure 16**).

Intrathecal morphine delivery has been studied in a prospective clinical trial, managing chronic nonmalignant pain (Anderson and Burchiel 1999). Anderson and coworkers concluded that continuous infusion, with carefully selected patients, resulted in a long-term improvement in several areas of daily function. The treatment of intractable chronic benign pain by intrathecal morphine administration has been found favorable in a retrospective multicenter survey by Njee *et al.* (2004). Vogelzang *et al.* (1987) evaluated prospectively the performance of the implantable Medtronic SynchroMed pump used in systemic chemotherapy and concluded a favorable infusion performance. The treatment of failed back surgery syndrome by electrical stimulation of the spinal cord has been evaluated by Van Buyten (2006) and concluded that spinal cord stimulation provided a long-term 50% reduction in pain in over 60% of the patients. Furthermore, pain medication could be reduced.

The drug delivery system consists of an implantable infusion-rate programmable pump that works like a roller pump but with a precise motor-driven mechanism. The device is refillable (transcutaneously) and is reprogrammed by telemetry means. In order to allow for local delivery of the drug, catheters from

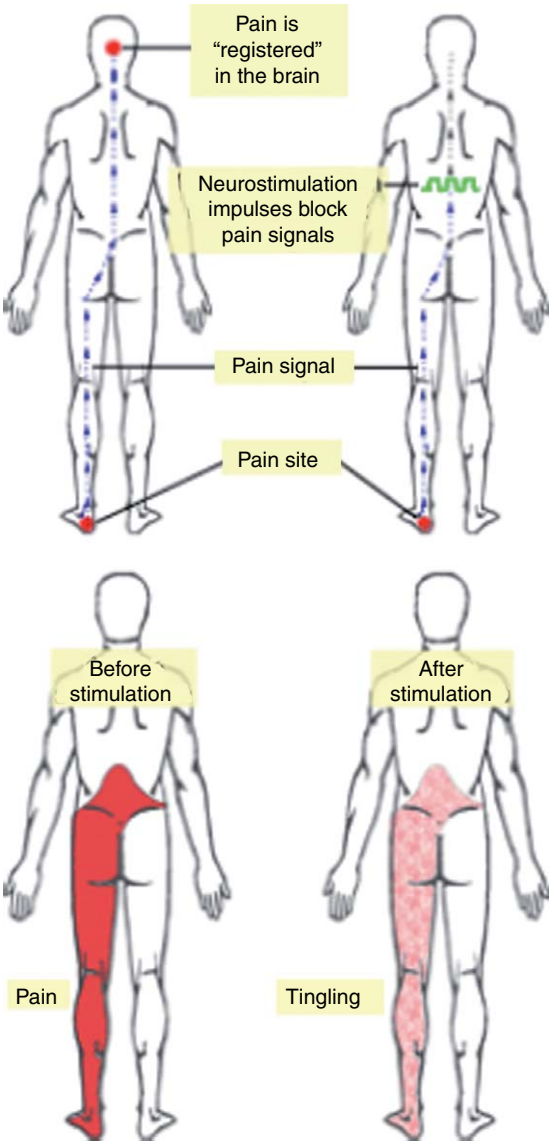


Figure 16 A simplified diagram of an electrical stimulation system blocking sensory nerve signals. (Courtesy: Medtronic.)

polymeric materials connect the target site to the pump. Figure 17 shows the system components.

A system providing electrical stimulation consists of an implanted stimulator, one or more electrodes, a patient activator, and a programming unit. The implantable pulse generator (IPG) is a versatile stimulator generating customizable pulse trains as required in the therapy. The patient activator allows for some intervention by the patient, like switching the stimulation on and off. The treatment of multifocal pain requires an extreme stimulation scheme urging the need for rechargeable batteries in

Figure 17 (*Left*) The implantable pump with catheter and (*right*) the device to reprogram the pump via telemetry. (Courtesy: Medtronic.)

neurostimulators. Figure 18 shows the several system components including a modern rechargeable implant.

3.14.1.2.2 Parkinson's syndrome

Parkinson's syndrome is characterized by movement rigidity, resting tremor, and bradykinesia, causing a serious impairment in the functioning of severely disabled patients. The incidence is up to 346 cases per 10 000. Pioneer Professor Benabid *et al.* (1998) studied the long-term effect of electrical inhibition of deep brain targets in movement disorders. The stimulation frequency (in the brain) used in suppressing these movement disorders is higher than practiced for pain suppression (in the spinal cord). Devices used successfully in the past years are Medtronic's Kinetra devices, now indicated for essential tremor and Parkinson's disease. Before implanting devices, a rigid clinical procedure aimed at precise implantation of special deep brain electrodes is executed. This procedure includes the use of a stereotactic frame, magnetic resonance imaging (MRI) recordings, and screening/signal recording with temporary microelectrodes in order to achieve the correct implant target size. After localization of this target lesion, electrodes are implanted for the definite therapeutic procedure.

3.14.1.2.3 Epilepsy

Epilepsy, from the Greek word epilepsia meaning 'a taking hold of or seizing', is a neurologic disease commonly affecting those younger than 18 years and those older than 65. Clinically diagnosed epilepsy affects approximately 1% of the population and is derived from congenital defects in the brain, tumors, trauma, illness, or strokes. But these known causes only account for 35% of the cases; there is no known cause for the remaining 65% of those afflicted. The susceptible age span of under 18 and

Figure 18 (*Top*) Physician programmer (*left*) and external and implantable rechargeable neurological stimulator. (*Bottom*) Traditional implant with extension cable and stimulating electrode (*left*) and patient activator (*right*). (Courtesy: Medtronic.)

over 65 years comes from the fact that most children born with epilepsy tend to outgrow it, and those people over 65 have increased risk due to age-related maladies such as stroke and anomalies. In any case, those with epilepsy live full lives with the proper management of the disease.

3.14.1.2.3.(i) Epilepsy overview Epilepsy manifests itself with seizures. Seizure is a state when normal, orderly electrical activity in the brain breaks down and a storm of activity occurs. There are two types of seizures: partial and generalized. Partial seizure is further subdivided into simple and complex seizures. A partial seizure involves a localized portion of the brain and may not even be noticed by observers (simple seizure) or create a loss of normal awareness (complex seizure) such as a blank stare or a repetitious movement or sound. A general seizure encompasses the entire brain in the storm of electrical activity manifesting itself in the more banal conception of an epileptic seizure. These seizures carry the terms petit mal and grand mal (among other types) and have symptoms ranging from unresponsiveness for the former to muscle rigidity and convulsions for the latter. These seizures last for as little as 30 s to as long as several minutes. Any person having two seizures is clinically diagnosed with epilepsy.

3.14.1.2.3.(ii) Disease treatments As stated, a person is clinically diagnosed with epilepsy when two seizures have occurred. To further understand when a seizure occurs, doctors employ several tests and pieces of equipment including electroencephalogram (EEG) to measure the electrical activity in the brain, magnetic resonance imaging (MRI) to image the activity of the brain, single-photon emission-computed tomography (SPECT) to measure blood flow in the brain, and magnetoencephalography (MEG) to measure electrical current loops in the brain. Each of these measurement methods is a tool to identify when the brain loses its orderly electrical process. Once the diagnosis has been made, there are four current practices in which to manage it: medication, diet, surgery, and vagus nerve stimulation (VNS). Of these four methods, only surgery has the opportunity to cure epilepsy. The remaining therapies can only manage it.

The vast majority of people with epilepsy manage the seizures through antiepileptic drugs (AEDs). This regime often requires taking a daily pill and living with their inevitable side effects. In some cases though, people are resistant to drugs and thereby require alternative treatments.

For those who have tremendous willpower, or are solely dependent on a caregiver, a ketogenic diet may control the seizures. This diet consists of forcing the body to produce ketones, a by-product of energy production from the burning of fat, not carbohydrates. Hence, the diet must virtually consist of all high-fat foods and none containing carbohydrates. Apart from being extremely unpalatable, it has its own medical consequences because of kidney and heart problems.

The third option that is often of last resort is surgery. This procedure involves either the removal of the problematic portion of the brain (curative) or simply severing pathways between areas of the brain (reduction of symptoms). Surgery is an option only when partial seizures occur since generalized seizures involve the entire brain and localized therapies therefore do not exist.

3.14.1.2.3.(iii) Vagus nerve stimulation The final solution is also the most recent: VNS. This therapy was first conceived in the 1930s but could not be realized due to the lack of contemporary technology. Finally, with the advent of microelectronics, specifically in the field of cardiac pacing, the United States of America Food and Drug Administration (FDA) approved the use of a pacemaker-like device for VNS in 1997. As with other treatments for epilepsy, researchers do not know how VNS relieves seizures, but efficacy has shown that 45% of people had fewer seizures, with 20% of those people reducing their seizure frequency by 75% (Schachter 2002).

This option is often much more acceptable to the patient than is brain surgery or even complete dependence on medication. Though the FDA has not allowed exclusive VNS therapy (medications are still required), it has been able to reduce the amount of pharmaceuticals required thereby lessening the side effects from them.

VNS does require surgery, but the surgery is minor in comparison to brain surgery, often with the patient going home on the same day of the procedure. Like the implantation of a cardiac pacemaker, a small pocket is made just below the shoulder on the left side of the body. A lead is then run subcutaneously up to the base of the neck where an electrode is then wrapped around the vagus nerve. Once activated, the stimulating device will pulse the nerve with electrical signals using parameters such as an amplitude of 1–12 mA, a pulse width of 120–1000 μs at frequencies of 1–145 Hz for intervals up to 270 s on and 180 min off (Cyberonics 2006 http://www.cyberonics.com). For emergency situations, a hand-held magnet that recognizes the onset of a seizure can be waved over the device to provide instant therapy for the patient.

The microsystems for these devices are extremely similar to those used in cardiac pacemakers. The key technology involved is microelectronics, which continues to provide capacity for an ever-growing set of features, miniaturization, and increased longevity. Because these devices are programmed in the doctor's office postoperatively, RF communication is a necessary component. Finally, advances in energy storage continue to increase the implant life of VNS devices and also allow the reduction of their volume for increased patient comfort (and acceptance).

VNS does offer many benefits with few side effects. Because the vagus nerve is being stimulated, patients have complained of annoyances such as tickle in the throat, difficulty swallowing, and coughing or hoarseness of the voice, but all diminish or disappear over time.

3.14.1.2.3.(iv) Epilepsy solutions on the horizon Today, researchers are investigating the use of electrical stimulation to disrupt the storm of activity associated with epileptic seizures. Currently, deep brain stimulation is being used to address movement disorders (e.g., Parkinson's disease) by placing electrodes deep within the brain and pulsing electrical currents into specific areas of the brain to disrupt the disorderly electrical activity. It appears that the same approach can be applied to epilepsy using devices similar to the cardiac pacemaker.

3.14.1.3 Spinal and Orthopedic Applications

3.14.1.3.1 Introduction

The orthopedics industry encompasses all aspects of human musculoskeletal system, including major and minor joints, bones, muscles, spine and the surgical tools, and the equipment necessary to correct and repair traumatic, degenerative, and congenital defects. Orthopedics has naturally evolved into categories with each utilizing different approaches to care. These

categories commonly consist of major joints, including the knees, hips, and shoulders, and are frequently the subject of replacement and revision. Minor joints include wrists, elbows, fingers, and toe joints, which, because of their size and ability to convalesce without severe impact to daily routine, do not receive the level of attention that major joints currently receive. Bones, aside from the obvious need to repair those broken, can require extensive surgery and extended therapies when it is necessary to correct them when affected by severe trauma, congenital defects, or degenerative conditions including sports-related injuries (**Figure 19**).

The field of spinal repair is extensive. Special focus is afforded to the spine because of the criticality of the spinal column and the spine's ability to be extraordinarily flexible while also providing a key structural role to the construction and the locomotion of the human body. The spine consists of 33 vertebrae, 24 having different degrees of mobility (cervical, thoracic, and lumbar) with the remaining 9 (sacrum and coccyx) being fused. Each of the three mobile portions consists of a vertebra responsible for providing protection to the spinal cord, points of exit for nerve bundles, anchor points for the muscular system, and shock-absorbing intervertebral discs (Cleveland Clinic Spine Institute 2006).

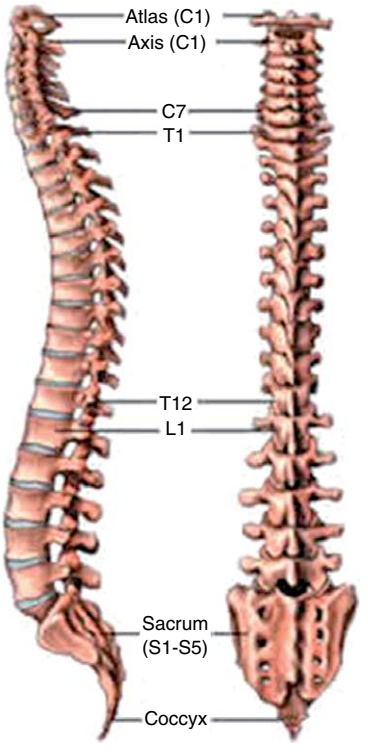


Figure 19 Anatomy of the spine. (Countesy: Medtronic.)

Orthopedics also includes the surgical tools and equipments necessary for repairing the skeletal system. Though traditionally this field has included the creation of specialized surgical tools, plates and screws, there is a growing use of microsystem- and biology-based solutions. In addition, the surgical field includes the navigational tools required to minimize the invasiveness of surgery, enabling new, previously impossible procedures and allowing the patient to return to normalcy sooner than otherwise.

In 2005, the worldwide orthopedic industry reached $19 billion in sales and was continuing its steady, double-digit growth. Hip and knee arthroplasty, or joint replacement, are the dominant segments followed by spine, sports medicine, and trauma repair. The industry's continued growth is expected to continue, supported by studies showing a strong correlation between the major joint replacement (primarily knee) and the increase in age and weight of the US and Western Europe populations. In addition, China, India, Latin America, and Mexico are beginning to emerge as new markets for orthopedics, which will further improve the continued growth of the industry (Quirk and Clair 2005).

3.14.1.3.2 Orthopedic and spinal diseases

There are two primary drivers for corrective procedures: degenerative conditions and trauma. Regardless of the reason, the need to return the patient to normal activity of daily living drives all advancements in technology (Szivek *et al.* 2002). These technological advancements themselves have two fields of focus: first is to provide higher quality information as to the state of injury and the second is to reduce the trauma induced by the corrective procedure. The first focus often implies placing sensors *in vivo,* while not exacerbating the condition or inconveniencing the patient, followed by the transmission of the collected data out of the body so that the doctor has the required information to make actionable decisions (Morris *et al.* 2001). The second focus is to develop methods to place corrective devices (e.g., artificial joints, rods, screws, intramedullary nails, or sensors) and to minimize the trauma required for adequate placement.

Today, there is limited use of microsystems in orthopedics. There are three reasons speculated as to why this is. The first reason is that traditionally the surgical solutions have been founded in the mechanical engineering discipline. Because of this, the knowledge to apply microsystems was not as readily on hand as were mechanical solutions. The second reason is that

there is little space in the patient's joint to contain the added functionality provided by microsystems. Regardless of the additional benefits that microsystems can provide, there is still a minimal amount of physical material needed to repair the damaged joints (e.g., prosthesis). Historically, the addition of sensors, electronics, and telemetry systems could not fit into the space left without debilitating or discomforting the patient. Finally, until recently microsystems could not provide the benefit for justifying the high costs of inventing, producing, and implanting the technology. Today, the advancement of these three areas is such that microsystems are now beginning to appear in commercialized use.

Of the orthopedic conditions warranting monitoring and correction, osteoarthritis (the deterioration of the cartilage meniscus between the ends of the bones) is the leading cause for both knee and hip replacement and continues to increase in prevalence due to an aging worldwide population. The repair of knees for osteoarthritis involves the use of a fixed bearing implant where a metal tray is fused with the tibia (lower leg) and the cartilage meniscus is replaced with a polyethylene bearing. For hip prostheses, the deteriorated joint is replaced with four parts consisting of a stem, ball, cup, and liner (**Figure 20**).

There are three conditions where the surgeon would like *in vivo* data with regard to the described knee and hip replacement. First, when a knee is replaced, the surgeon must ensure that the pressure across the joint is evenly distributed to ensure patient's comfort and minimize wear on the bearing (D'Lima *et al.* 2005). Today, the materials selection and loading models used to design the prosthesis are based on empirical data, mathematical models collected from cadaver studies, and nearly 40 years of clinical experience (Morris *et al.* 2001). After the joint has been implanted and is in use by the patient, this approach often uncovers discrepancies in the loading, wear, and the use of the joint. A second need for data comes after the surgery with the desire to know how well the cement-to-bone fusion is, or during the life of the prosthesis, if loosening occurs within that interface from the stresses of use (Cristofolini *et al.* 2000, Jeffers *et al.* 2005). The difficulty arises because radiographic techniques are unable to detect these microcracks and bone erosion until the interface has weakened to the point where major revision may be

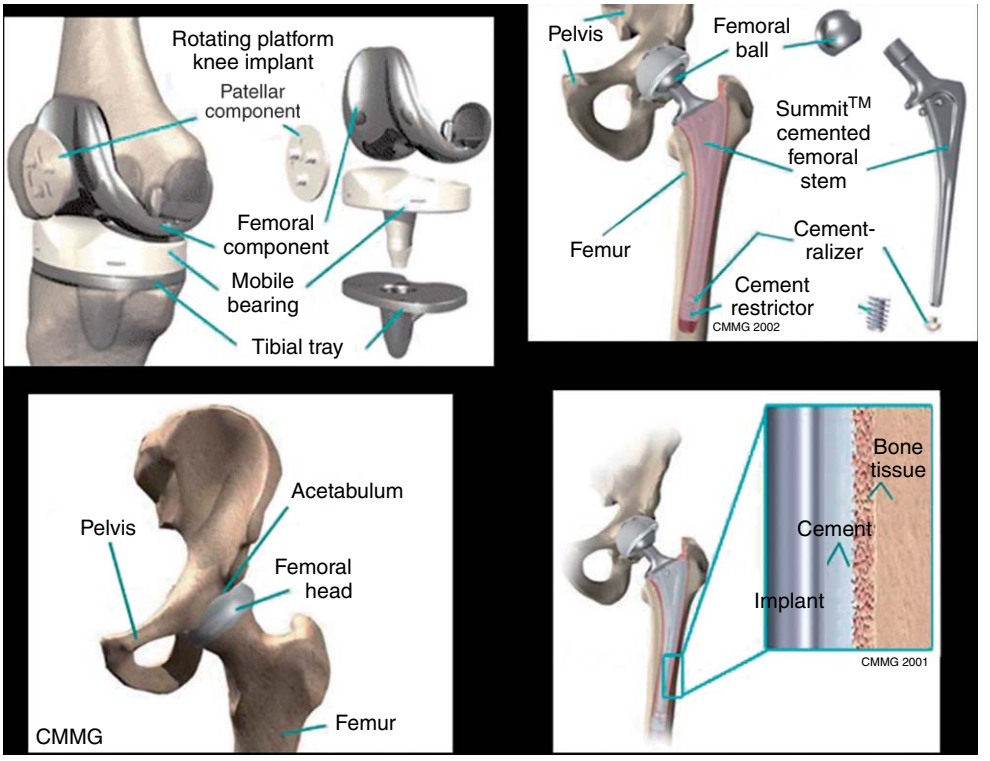


Figure 20 Construction of artificial knee and hip. (Courtesy: DePuy Orthopaedics Inc. and Medical Multimedia Group, LLC.)

required (Robinson 1997). The third situation that patients of knee and hip replacement face is the formation of osteolytic lesions. Fine debris is naturally generated from the friction between the prosthesis' metal tray and the polyethylene bearing. When this debris comes into contact with the living bone around the artificial joint, the body attacks the foreign material and destroys the living tissue in the process. Large voids evolve around the joint leading to loosening and eventual revision (Gallo *et al.* 2002). Knowing that this is a natural evolution to a joint replacement, which ultimately leads to loosening and revision, the treating physician has to monitor the progression of the lesion and intervene prior to significant damage.

Spinal procedures and subsequent corrections are the fastest growing segments of the orthopedic and spine market with 895 000 corrective procedures conducted in 2003, yielding $2.9 billion in sales. Like the rest of orthopedics, spinal disorders fall into degenerative and trauma-based causes. Disc injuries, such as herniated discs and degenerative disc disease, are the most common malady such that "approximately 80% of Americans experience at least a single episode of significant back pain in their lifetime, and for many individuals, spinal disorders become a lifelong malady" (Traynelis and Haid 2001). But, of these episodes, only 200 000 culminate with a surgical correction. In addition to discs, other spinal disorders include osteoarthritis (inflammation of the facet joints), spondylolithesis (slippage of vertebra), scoliosis (curvature of the spine), and traumatic events resulting in vertebral compression fractures (Quirk and Clair 2005) (**Figure 21**).

There are several progressions to the treatment of disc-related ailments. During the initial onset, bed rest, limited activity, and the use of nonsteroidal, anti-inflammatory drug (NSAID) often goes far toward reversing the pain. But, if the disc herniates or does not respond to the previous therapy, surgery may be required, including decompression of neural elements, discectomy, fusion of the adjacent vertebrae, or the implant of an artificial disc (WebMD 2001). For spondylolithesis and scoliosis cases, rods and fusion are treatments for the more advanced cases. Similar to the situations found in knee and hip replacement, physicians dealing with the spine need *in vivo* information. This information includes the pressure distribution across an artificial disc or the state of a vertebral fusion's progress. Because the doctors lack this information, they are forced to err on the conservative side potentially resulting in delayed return to normal or in limited range of motion. Potentially, the ability to monitor intervertebral pressures or prosthesis position during surgery can avoid postsurgical complications.

Figure 21 BRYAN® Replacement cervical disc. (Courtesy: Medtronic. BRYAN® (Cervical Disc System incorporates technology developed by Gary L Michelson, MD.)

Surgery, along with the need to keep the trauma induced by the surgical procedure itself to a minimum, is a critical part of orthopedic and spinal corrections. Tremendous strides have occurred in recent times toward this end with the development of tools such as endoscopes, Medtronic Sofamor Danek's (Memphis, TN, USA) Sextant Rod Insertion System, and surgical navigation equipment. Each of these solutions serves to reduce the size of the incisions and the resulting trauma by providing greater placement control of the surgical tools and corrective devices.

3.14.1.3.3 Current solutions

The orthopedics and spine industry has historically been dominated by mechanical engineering solutions. Today, microsystems are beginning to appear more prevalently in research and are showing early signs of being incorporated into products.

The most significant use of microsystems in orthopedics and spine has been with strain gauges. These devices are thin film (on the order of 0.076 mm in thickness and of planar dimensions on the order of 10 mm × 10 mm) sheets of polymer with metal, serpentine traces embedded in them that change the impedance when deformed (Szivek *et al.* 2002). Placing a gauge (or multiple gauges) in a Wheatstone bridge configuration allows the detection of minute strains while nulling out the effects due to temperature, aging, etc. Within this scope of use, companies such as Microstrain Inc. (Williston, Vermont, USA) have made significant headway with the use of strain

gauges to solve the previously described challenges. The direct application of these gauges have provided the *in vivo* data needed for measuring hip joint movement long before radiographic techniques could (MicroStrain 2001), for monitoring bone elongation (Noonan *et al.* 2004) and for the study of load-bearing effects on spinal ligaments (Roberts *et al.* 1998).

A notable use of strain gauges was realized during the collaboration between The Center for Orthopaedic Research and Education at the Scripps Clinic (La Jolla, CA, USA), DePuy Johnson and Johnson (Warsaw, IN, USA), NK Biotechnical Corporation (Minneapolis, MN, USA), and Microstrain Inc. Collaborating in the e-Knee project, they addressed the need for *in vivo*, real-time, joint pressure distribution data by creating a tibial tray specifically designed to transduce the pressures exhibited across the joint into electrical signals and telemeter the data to an external recorder. Four cavities were located on each of the four corners of the tibial tray with resistive strain gauges mounted on the roofs of each cavity. Pressures from a metal upper plate attached to the femur were collected by posts separating this upper plate from a similar lower plate that contained the strain gauges. The pressures transferred by the posts, each located over a corresponding cavity in the lower plate, deformed the cavity roof, which was then converted to a proportional electrical signal via changes in the resistance in the strain gauges (Kovacevic 1993). These data were then collected by a microprocessor, also embedded in the prosthesis, and were telemetered out of the patient to a data logger via an embedded RF circuitry and an antenna threaded into the end of the stem (Morris *et al.* 2001). The current version does not have a self-contained power source but is powered through an external inductive coil. Several companies, including Microstrain Inc., are continuing to research advanced energy harvesting techniques creating the possibility that the next generation of e-Knee is powered through the normal motion of the joint (**Figure 22**).

The e-Knee is a major advancement toward the understanding of the long-term dynamics of an artificial joint and how particular surgical methods and material effect the ultimate success of the replacement (D'Lima *et al.* 2005). Similar types of measurements have been made for hip implants (Bergmann *et al.* 1988, Davy *et al.* 1988) with creative solutions to address the limited space available for the electronics achieved only through the cooperation of the mechanical and the electronic disciplines.

Figure 22 e-Knee and inductive power coil (Source: Morris BA, D'Lima DD, Slamin J, Kovacevic N, Arms SW, Townsend CP, Colwell CW Jr 2001 Evolution of the electronic knee prosthesis: Technology Development. *J. Bone Joint Surg Am 83* (Supp 2), S62–S66. © The Journal of Bone and Joint Surgery, Amc.www.ejbjs.org.

The study of human joints in motion is an important area of study in order to understand how changing loads generating axial and lateral pressures on the bones, cartilage, and ligaments affect joint wear, longevity, and the patient's quality of life. One of the factors that goes into the study is how fatigue, footwear construction, and the characteristics of locomotion change a person's walking or running gait, which in turns changes the kinematics of their joints (Arndt 2003). Novel GmbH (Munchen, Germany) offers a complete system to fully analyze the pressures across the soles of the foot under varying conditions. This system is based on an array of capacitive sensors and their supporting electronics and power. These sensors are constructed of two parallel conductive plates separated by an insulator. As pressure from the foot is applied to the sensor, the two plates are compressed toward each other thereby changing the capacitance between the plates. The supporting electronics can detect this change in capacitance and can transduce it to a pressure measurement, then transmit it to a PC-based data collection system, or a handheld PDA, via a wire harness (suited for controlled lab conditions) or wirelessly using Bluetooth. These sensors can also be applied for similar studies on the hand and body surfaces subjected to long durations of pressure (e.g., the upper legs for those confined to wheelchairs or those wearing artificial limbs). Novel GmbH also offers their line of capacitance-based sensors with

different coatings and in arrays for *in vivo* joint studies.

As described previously, knee and hip joint replacements constitute the largest part of the orthopedic industry. A major reason for joint revisions is due to the interfacial failures between the prosthesis and the cement. The materials and designs selected for the artificial joint replacement are based on finite element analysis models and studies done *in vitro*, but these methods do not always provide a complete picture of the stresses experienced by the joint. As stated earlier, strain gauge is a solution to obtain *in vivo* data, but piezoelectronic sensors do offer an alternative means to obtain these data (Cristofolini *et al.* 2000). Thin-film strain gauges are ideal for measuring tensile stresses along its X–Y-axes but are unable to accurately measure compressive stresses in the same plane and completely unable in a plane perpendicular to that containing the serpentine traces. Piezoelectronic sensors are ideally suited for this particular case, i.e., for measuring forces perpendicular to the surface of the prosthesis, which are those that impact the integrity of the cement bond.

Though the piezoelectronic sensor has the added advantage of requiring a lesser use of surface area than would a thin-film strain gauge, there are two significant detractions with its use. First, they consume more volume than the thin-film strain gauges, and second, they are unable to hold the charge resulting from the physical deformation for long periods of time. This means the charge accumulated on its surfaces will leak off back to a neutral voltage over a period of tens of minutes. Consequently, the sensor has to be measured on a relatively frequent basis to measure the changes in stresses across the sensor.

The use of accelerometers is still another method to detect loosening in the cement–bone interface. A micromachined accelerometer, a device used to measure accelerations induced by movement, can be placed into the head of the artificial joint. At periodic checkups, the doctor applies a vibration at the distal end of the bone with the artificial joint and collects the vibratory motion at the proximal end (where the accelerometer is located within the prosthesis). The resulting signal from the accelerometer is collected and converted to meaningful data by microelectronics and is telemetered out of the body to be analyzed by a computer. If the joint has loosened, the proximal end of the bone, as well as the accelerometer, vibrates in a nonlinear manner generating harmonics of the sine wave. It is the detection and the measurement of these harmonic amplitudes that allow quantification of the joint failure (Puers *et al.* 1999). The primary advantage of this approach is that a qualitative measurement of the entire cement–bone interface can be acquired versus the status of a point location. Another use of this technique also includes the quantification of progress with the fusion of a bone fracture.

A final sensor method used in orthopedics involves the use of acoustic emissions. Though this technique does not have *in vivo* applications yet, it has been used to research the materials used for prosthesis in studies (Jeffers *et al.* 2005). Based on piezoelectronic sensors, "the acoustic emission technique can be used to obtain information about the microstructural changes that are occurring in any loaded material" (Surgeon 2004). The principle of the technique is the measurement of the stress waves that are propagated as microscopic bonds begin to break. These stress waves are recorded and processed, resulting in a picture of what is occurring within the joint under study as different loading conditions are applied. The limitation at the current time is the amount of processing required. But as microelectronics continues to increase in processing power while shrinking their physical size and power consumption, it will be a short time before the technology improves to the point of allowing the microelectronics to be implanted in the artificial joint to listen to the joint kinematics resulting in extremely precise, real-time analysis.

With orthopedic implants comes the necessity for surgery. The direction of technology development is to further minimize invasive surgery. Some of the more extreme surgeries in orthopedics involve the need to place rods along bones and the spines to stabilize them until they have healed, or fused. In time past, this would have required an incision of the full length of the rod through the skin and the muscle. Today, companies such as Medtronic Sofamor Danek are making tools such as the Sextant that minimize this surgical trauma. Through a very few, small incisions, cannulated multiaxial screws are attached to the bone followed by a rod slid and secured into place. This avoids the disruption of blood supply and innervation seen with previous procedures.

In conjunction with the tools for minimally invasive surgery (MIS), which by definition severely restricts or eliminates direct viewing of the procedure, comes a need to navigate through the body. Endoscopic surgery allows the surgeon to view inside the body through a few, 6.4 mm incisions. Larger

pieces of equipments using state-of-the-art computed tomography (CT) and 3D tracking continue to advance the state of remote surgery. Previous generations used optical cameras arranged in an array to identify markers on surgical tools with respect to references on the patient. This information is then displayed on a monitor superimposing the surgical tool tips onto the CT image, which provides the view for the surgeon. A major drawback of this system is a loss of tracking if something or someone is placed between the cameras and the tools. This issue is being resolved as the next generation of navigation tools, which utilize an array of electromagnetic sensors tracking comparable references on the tools, is rolled out to the market. This advancement eliminates the temporary loss of information during surgery and allows the surgical staff to move in a normal manner within the operating room.

Microsystems are proving to make the operating room safer as well as reducing the opportunity for errors. Collaboration between orthopedic surgeon Alfred Durham, Virginia Tech Applied Biosciences Center (Blacksburg, VA, USA) and Carilion Biomedical Institute (Roanoke, VA, USA) has created a handheld device that allows the surgeon to align a drill with pre-drilled holes in an intramedullary nail placed down the center of a broken bone. In the past, the surgeon attempted to align the drill bit with the hole using real-time X-ray resulting in exposure of the surgical staff to the radiation. By utilizing the principles of magnetic fields, the team created custom electronics to determine the placement of a drill sleeve with respect to a fixed embedded magnetic sensor in the nail providing feedback to the surgeon as to the direction needed to achieve the alignment. Location and alignment are obtained in much shorter time and eliminate the danger presented by the X-rays (DeGaspari 2005).

Finally, to speed the patient to full recovery, companies such as EBI, L.P. have developed implantable spinal fusion and bone growth stimulators. These stimulators can generate up to 60 μA of constant direct current for spinal fusions and up to 40 μA for bone healing, over a period of approximately 24 weeks, and have been shown to increase the probability of a successful fusion by 27% for spinal fusion (Kane 1988) and 38% for bone healing (EBI 2006). EBI, L.P. (Parsippany, NJ, USA) also offers external stimulators thereby eliminating the need for surgery to implant the device (**Figure 23**).

Figure 23 Spinal electrical stimulator. (Courtesy EBI Medical. Inc.)

3.14.1.3.4 Spinal and orthopedic solution on the horizon

Looking toward the future, it becomes apparent that the transition from a mechanical only based area of medicine to one of the integrated microsystems and advanced surgical navigation will continue to accelerate. The need to improve the design and materials of implants for longer prosthesis life and better therapeutic use comes clearly from the information gathered under real-world conditions, meaning real-time, *in vivo* data. This need is seen not only by commercial firms looking for the new untapped market, but also by governments. A project created under the Brite–Euram Program for the Commission of the European Community is chartered to develop smart orthopedic implants. This project will serve as a basis to advance the state of the materials, modeling, and data acquisition (Burny *et al.* 2000).

The use of microsystems in orthopedics continues to gain more attention as evidenced by the recent application of MEMS. Dr. Huikai Xie of the University of Florida (Gainsville, FL, USA) has proposed

using the AccuMicroMotion, a 1 mm × 1-mm sensor integrating a three-axes accelerometer and a single axis gyroscope MEMS device that "· · · could be implanted in bones during surgeries, giving orthopedic surgeons a unique way to monitor the progress of repairs" (Hoover 2006, Sadat *et al.* 2005) (**Figure 24**).

Theken Disc (Akron, OH, USA) is also expanding the use of microelectronics with the incorporation of a 2 cm^3 electronic circuitry into an artificial spinal disc. As of April 2005, the eDisc had been implanted into a primate for evaluation prior going to human clinicals. The eDisc "· · · redefines postoperative patient management through wireless data collection. This provides surgeons the unprecedented ability to assess and mange their patients by tracking the motions and loads on the lumbar spine" (Theken 2005) (**Figure 25**).

Amazing advances are also being made in orthotics, the development of electromechanical artificial limbs. One of the great limitations in this field has been the inability to control more than one joint movement at a time. This is now changing with the research being conducted at the Neural Engineering Center for Artificial Limbs, a program within the Rehabilitation Institute of Chicago (Chicago, IL, USA). This single degree of motion is now eliminated with the incorporation of new physiological advances using muscles as electrical amplifiers for the nerves that had operated the muscles of a severed limb, as myoelectrical sensors, and as controllers for the artificial limb. The new electronics improve the detection, amplification, and decoding of the electrical signals generated by the nerves, which in turn are translated into electromechanical instructions that operate the limb. The result is a much more natural control and simultaneous movement of multiple joints. In fact, two patients who have been outfitted with experimental devices are now able to throw a ball (Kuiken *et al.* 2005).

The future of surgical navigation is bright as well, with the use of robotics in the operating room being explored in limited applications (Sherman 2006). As advances in sensors, motors, imaging, and computing power continue forward, these assistants will do more and more of the delicate repair while minimizing the trauma further.

3.14.1.4 Obesity, Gastroenterology, and Urology Applications

3.14.1.4.1 Obesity

Obesity is the condition in which the energy stores of the body are too large (Canham 2003). Since the mid-1970s, the prevalence of those people who are classified as overweight or obese has increased dramatically in the United States, as well as in other parts of the world, with children and adolescents tripling and nearly two-thirds of adults now overweight or obese. This dramatic increase in weight is associated with an increase in the prevalence of associated comorbidities such as type-2 diabetes, hyperlipidemia, coronary heart disease, hypertension, stroke, asthma, and obstructive sleep apnea (Buchwald *et al.* 2004).

Metabolic syndrome describes a group of physical and metabolic conditions that are known health risks

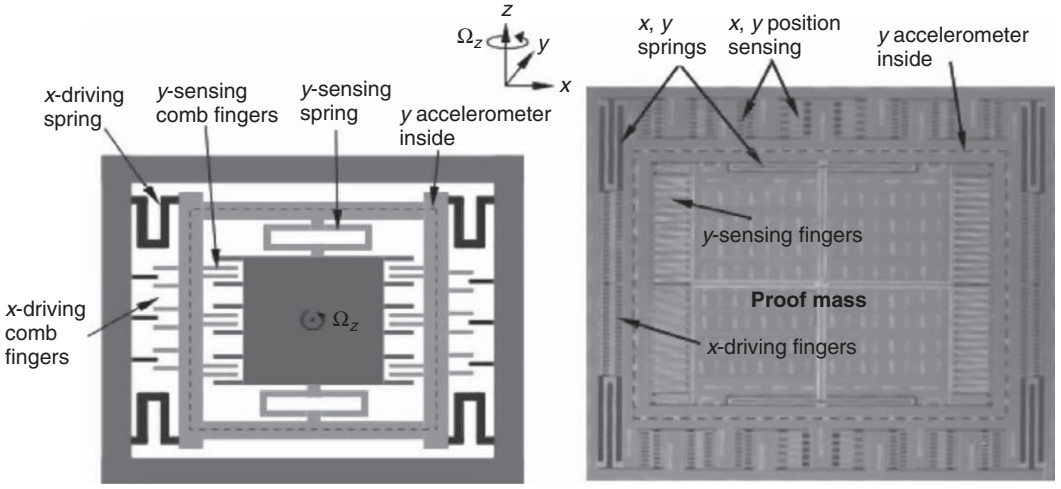


Figure 24 Diagram of AccuMicroMotion Sensor. (Source: Sadat A, Qu H, Yu C, Yuan J, Xie H 2005 Low-Power CMOS Wireless MEMS Motion Sensor for Physiological Activity Monitoring. *IEEE Transactions on Circuits and Systems.* v52, n12, pp. 2539–2551.)

Figure 25 Theken e-Disc (© 2004 Theken Disc.)

for patients, including conditions such as large waist circumference, insulin-resistant glucose metabolism, dyslipidemia, and hypertension. For example, patients with metabolic syndrome have 1.5–3.0 times increase in risk of coronary heart disease and stroke. But this condition is reversible. It has been shown that weight loss can improve all morbidities of metabolic syndrome (Klein *et al.* 2004), and therefore a means of reversing this obesity trend has become a major focus of research.

Following the first law of thermodynamics, the underlying cause of weight gain can be directly attributed to a long-term imbalance of energy consumption being greater than energy expenditure resulting in the excess energy stored as adipose (fatty) tissue. Other complications are caused by hunger and are also due to psychological factors such as taste, stress, and social reasons. On the consumptive side of the energy equation, expenditure is composed of the basal metabolic rate, thermic effect of food, and activity thermogenesis. While both energy consumption and expenditure appear to be controllable by an individual, uncontrollable factors such as hormones and genetics also play an influential role in determining the weight (Bouchard *et al.* 1990, Levine *et al.* 1999).

Obesity is often characterized in terms of body mass index (BMI) and obesity class. BMI is calculated by dividing the weight (in kilograms) by the square of the height (in meters), with an obesity class of I, II, and III having a BMI between 30 and 35, 35 and 40, and >40, respectively.

Therapy efficacy is quantified in terms of excess weight loss (EWL), defined as the amount lost divided by the excessive weight (amount above the ideal weight). Pharmacological treatments have shown a low EWL. Therefore the primary means of EWL is derived from surgical techniques. The surgical techniques employed for treating obesity today focus on reducing the body's ability to absorb calories. The more common techniques include gastric bypass, which relies on physical changes in the size and the intestinal pathway, and (more widely used) placing a band around the stomach. There are several gastric banding procedures such as laparoscopic gastric bypass, adjustable gastric band 'Lap Band' (Kuzmak), vertical banded gastroplasty (Mason), and silicone ring vertical gastroplasty. While these surgical procedures have demonstrated high EWL (40–60%), these procedures also present a mortality risk (Flum *et al.* 2005), as well as being difficult to reverse.

An alternative way to encourage weight loss is to create a sense of satiety, or reduction in appetite. Electrical stimulation of the stomach is an emerging therapy seeking to achieve this satiation (**Figure 26**). The stimulator consists of a chronically IPG and is known in the industry as an implantable gastric stimulation (IGS). The IGS contains a hermetically sealed microcircuitry and a power source with the electrical stimuli delivered by means of leads from the internal circuitry (using hermetic feedthroughs) to specific points on the stomach. Much of this technology is based on implantable cardiac pacemaker technology. Though the electrical pulses from the IGS influence the mechanical activity of the stomach as well as the gastric hormone levels, it is thought to primarily bias the satiety level in the brain through vagal afferent conduction.

3.14.1.4.2 Gastroenterology

The gastrointestinal (GI) system is composed of the alimentary canal (or tract), salivary glands, liver, pancreas, and gall bladder (Young and Austrin 1979), and the function of the GI is to ingest and absorb nutrients while also excreting unabsorbed nutrients and waste products. The branch of medicine concerned with functions and disorders of this system is called gastroenterology (Canham 2003).

There are many GI diseases (gastroenterology including ulcers, abnormal tissue growth, GI bleeding, and inflammation). Crohn's disease, in particular, is an inflammation of the lower part of the small intestine that can lead to severe intestinal blockage. Though it is critical to identify and diagnose these GI diseases as early as possible, small and hard to reach anatomy can make visual image identification

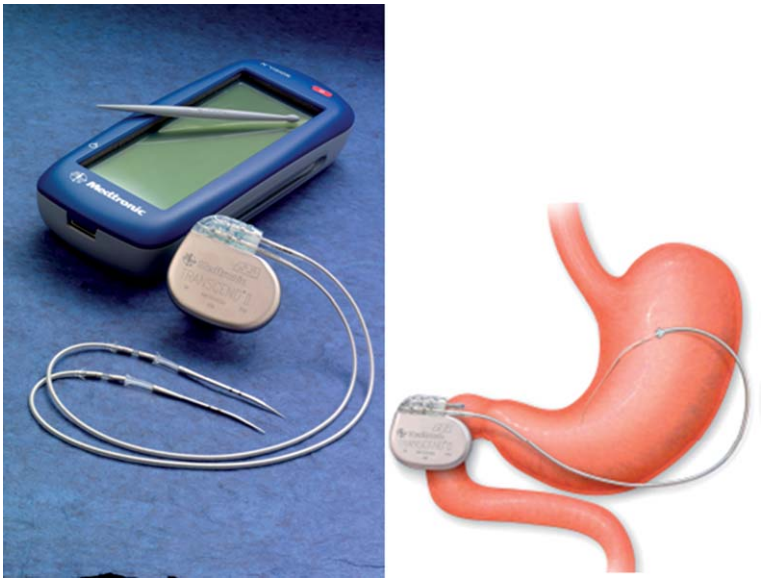


Figure 26 Physician programmer, gastric stimulator, and stimulation leads. (Courtesy: Medtronic.)

of anatomical and structural anomalies, as well as surgical procedures, very difficult.

Endoscopy is the key instrument for visualizing the interior of hollow organs (Merriam 1993) such as the upper GI tract (gastroscopy), the colon (colonoscopy), or the rectum (Yamada and Alpers 2006). GI scopes take up 37% of the total endoscopy market, which was estimated to be $324 million in the United States in 1999 (Technology for Industry 2005). Endoscopy has evolved from rigid tubes with fixed lenses in the 1970s to multicore fiber optics enabling flexible endoscopes in the 1980s (Wilkinson and Hitchings 2003). Further use of microsystems has led to the development of endoscopes with video cameras at the distal end (Panescu 2005). Other instances of microsystem technology enhancing the functionality of next generation endoscopes include Diagnostic Endo Laser Scanner's (DELAS, Laser and Medizin Technology GmbH, Germany) proposed use of a scanning imaging technique. The optical components consist of single fibers and microscanning mirrors to avoid the spectral limitations of conventional image guides (Schutz *et al.* 1997). More recently a team of researchers has developed a miniature confocal optical scanning microscope as an endoscopic application and has successfully obtained a real-time image. The confocal microscope can observe the longitudinal direction of a scanning head, including the electrostatic 2D MEMS scanner, and an aspherical objective lens. A scan angle of 12°, a mirror flatness of 80 nm, and a mirror reflectance of over 85% have been achieved. The diameter and the length of the scanning head are 3.3 and 8.0 mm, respectively, allowing it to be inserted into a channel of the endoscope. The lateral resolution is 0.5 μm, the depth resolution 2.9 μm, the field of view of 100 μm × 75 μm, and the frame rate is 20 frames per second (Murakami 2005).

Disadvantages of endoscopy are the pain it causes; the need for it to be done in the hospital, and it's inability to allow optimum view in difficult areas of the anatomy. More recent developments aimed at overcoming these limitations have resulted in the emergence of capsule endoscopy:

> The PillCam™ SB capsule endoscope is a naturally ingested device for use in the GI tract. Natural peristalsis moves the capsule smoothly and painlessly throughout the GI tract (**Figure 27**). The capsule contains an optical lens, complementary metal oxide semiconductor (CMOS) imager, white LED illumination, battery and radio transmitter to send color video images as it passes. Information is transmitted via an array of sensors secured to the patient to a DataRecorder worn around the patient's waist. The procedure is ambulatory, allowing patients to continue daily activities throughout the endoscopic examination. The effectiveness of this pill has been investigated for GI bleeding, Crohn's disease, small bowel tumors, pathology, esophageal mucosal pathology, and endoscopy (Given Imaging, Yoqneam, Israel) (Panescu 2005).

Figure 27 Inside the Pillcam™ (© Given Imaging).

An intracorporeal video probe (IVP) has been developed within the framework of a European Information Society Technologies (IST) project in both a catheter version (IVP1) and a stand-alone version (IVP2) for applications in video endoscopy and gastroenterology. It uses a CMOS camera and micromechanical tilting mechanism (Arena *et al.* 2005, Neidlinger *et al.* 2000, Sell 2005).

A failure in the normal physiological mechanism to prevent abnormal amounts of acid from entering the esophagus is known as gastroesophageal reflux disease (GERD) causing symptoms of heartburn and/or mucosal injury (esophagitis, peptic strictures, and Barrett's esophagus). The prevalence in the United States is 20%, while that based on injury is 7% and drives an annual cost for antacid medication in the United States estimated at $4 billion (gastroenterology).

The underlying mechanisms, therapeutic approaches, and potential complications are different for three unique groups (non-erosive reflux disease, erosive esophagitis, and Barrett's esophagus) of GERD (Fass 2002b). Current methodology for diagnosing entails passing a catheter, with a pH-sensitive ion-sensitive field effect transistor (ISFET) mounted on the tip, transnasally into the esophagus and leaving the catheter in place. This catheter is uncomfortable, cumbersome and conspicuous, leading most patients to modify their activity (Schepel *et al.* 1987, Smit *et al.* 1997).

Another diagnostic device, the Bravo pH capsule, is an oblong device (6.0 mm × 5.5 mm × 25.0 mm) (Wong *et al.* 2005) with an antimony pH and a reference electrode located at the tip, an internal battery, and a transmitter all encapsulated in epoxy (**Figure 28**). The Bravo sends data to an external receiver every 12 s via a 433-MHz RF signal while achieving data security via digital data transmission and a unique identification code. Prior to placement, the capsule is activated by a magnetic switch and is calibrated by immersing in pH buffer solutions of pH 7.01 and 1.07 (Medtronic, Shoreview, MN, USA).

The GI system is also used for drug delivery using oral tablets, capsules, solutions, rectal suppositories, and transmucosal (Technology for Industry 2005). Targeted delivery of chemotherapy for Crohn's disease and colorectal cancer has been achieved using nondigestible capsules containing a power source, a radio transmitter, and a medicament storage that can be opened remotely for delivery (GastroTarget Inc., Tonawanda, NY, USA). "The capsule signals a remote receiver as it progresses through the alimentary tract in a previously mapped route and upon reaching a specified site is remotely triggered to release a dosage of medicament" (Gastrotarget Corp. 1994).

A rise in CO_2 level of the gastric wall in critically ill patients can be an early warning for a reduction

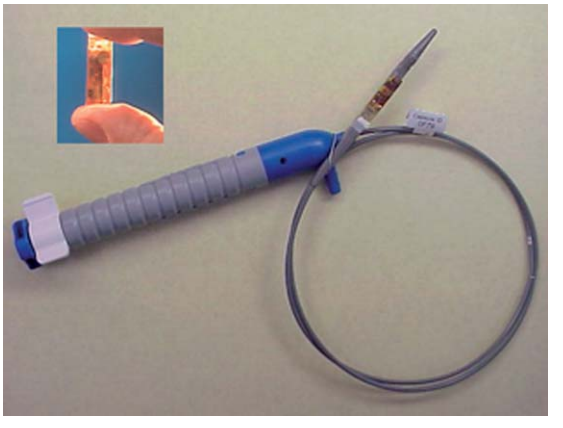


Figure 28 Bravo pH Monitor. (Courtesy: Medtronic.)

in peripheral blood flow, causing multiple organ failure. Companies such as Nihon Kohden (Tokyo, Japan) are investigating the measurement of this parameter using CO_2 ISFETs (Canham 2003). At the University of Twente (The Netherlands), a pH-sensitive hydrogel disk, which swells and deswells in response to pH changes clamped between a pressure sensor membrane and a porous metal screen together with a bicarbonate solution, is also being used as a CO_2 sensor. CO_2 reacts with the bicarbonate solution resulting in a pH change that is converted into pressure change by the enclosed hydrogel. The main advantage of this sensor is that a reference electrode required for potentiometric sensors is no longer needed (Herber *et al.* 2004b). The data collected showed that under isochoric conditions, the static and the dynamic stimulus-dependent behavior of hydrogels show fast reactions (Herber *et al.* 2004a).

3.14.1.4.3 Urology

Urology is the branch of medical science dealing with the study of the urinary tract in both sexes as well as the genital tract in males (Young and Austrin 1979). Urologists study and treat disorders of the urinary outlet of the kidney, urinary bladder, urethra, prostate, testicles, and penis diseases including obstructions from stones, tumors, inborn diseases, infections, impotence, sterilizations, and incontinence (Zeijlemaker *et al.* 2006). The renal system, pertaining to the kidneys, regulates water and electrolyte content; maintains the blood pH; excretes waste metabolic products, toxins, and drugs; retains vital nutrients; and regulates specific hormonal and metabolic functions (Canham 2003). The urological examination and diagnostic techniques include physical examination and urinalysis, instrumentation, and urinary tract imaging (Walsh *et al.* 2002).

Urinary incontinence is defined as involuntary loss of urine more than once a month at inappropriate places or times regardless of the amount that is lost (van Halteren 2003), which affects an individual's social interactions, self-esteem, and daily activities (O'Donnell 1997). Prevalence of urinary incontinence in the United States for the age group of 15–64 years is estimated at 10–20% for women and at 1.5–5.0% for men with an economic cost estimated to be $10.6 billion in 1987.

The urinary bladder is spherical in shape and is located at the end of both ureters in the abdomen behind the pubic bone with several blood vessels and a network of nerve bundles connected to it (Zeijlemaker *et al.* 2006). Incontinence, as a standalone disorder, is divided into stress incontinence caused by a weak sphincter, urge incontinence caused by a hyperactive bladder, or a combination of the two (mixed urinary incontinence). It can also be a part of more complex medical situations such as a neurological disorder, an obstruction, and mobility or cognitive limitations (van Halteren 2003).

Microsystems play a role in urodynamic investigations to diagnose urinary problems. This investigation, which takes about 30–60 min, involves artificial interventions in the bladder and simultaneous measurements. Cystometry is the study of the neuromuscular mechanism of the bladder by means of a cystometer that can measure pressure and capacity. A replacement device was proposed in 1984 for the catheter tip transducer used in cystometry. This telemetry pill enables pressure measurements under close to normal life circumstances for the patients and is introduced into the bladder using a cystoscope. A thick-film substrate serves as a carrier and as an interconnection for the components with the entire circuit protected in a micromachined ceramic package containing a removable container for the disposable battery. Using a dedicated receiver, developed with emphasis on noise reduction, the system is able to transmit pressure data over 4 m for up to 4 days per battery (Sansen *et al.* 1984, Puers *et al.* 1984). Micromachined piezo and capacitive pressure sensors for bladder pressure measurements are described in Puers (2005).

The wide range of therapies for various forms of urinary incontinence is classified as conservative treatments (physiotherapy, behavioral training, and physical stimulation), medicines, surgical, and external or internal assistance devices (van Halteren 2003) with the prescribed treatment depending on the cause of the disease (Zeijlemaker *et al.* 2006). For patients who are diagnosed with urinary urge incontinence (uncontrolled release of urine) or urinary urgency frequency (unstable bladder), electrical stimulation using an implantable stimulator may be a solution. This therapy is for those patients who have failed, or could not tolerate, more conservative treatments. Patients are contraindicated for implant if they have not demonstrated an appropriate response to test stimulation or are unable to operate the neurostimulators (Weil 2000). The implantable InterStim Therapy (Medtronic Inc., Minneapolis, MN, USA) system uses mild electrical stimulation of the sacral nerves that influence the behavior of the bladder, sphincter, and pelvic floor muscles (see

Figure 29 Implantable Interstim Therapy (Courtesy: Medtronic.)

Figure 29). In properly selected patients, InterStim therapy can dramatically reduce or eliminate symptoms.

In 1987, a closed-loop control of the unstable bladder was proposed. It relies on a biofeedback system that makes use of the bladder pressure as a parameter for functional stimulation. The key element is an implantable bladder pressure telemetry device, which relies on an RF power link that converts the pressure signal into a pulse position-modulated signal. To close the biofeedback system, an independent programmable stimulator has been developed, which is triggered by the pressure data (Sansen *et al.* 1987).

A urinary tract stimulator that is intended to restore normal bladder function in patients who have spinal cord injuries was proposed in 1993. The system is composed of two principal parts. The first part, which is external, consists of a handheld device based on an inductive-coupling technique. The second part, a fully programmable implantable device, includes a CMOS array IC that controls eight monopolar (or four bipolar) stimulation channels. In the experimental phase, the researchers investigated the effect of early electrical stimulation of the bladder during the spinal shock phase in paraplegic dogs. In addition, using the stimulator, they localized the parameters of stimulation that give best results in terms of effective bladder pressure and void a high volume of urine (Sawan *et al.* 1993). The combination of microfabricated sensors and actuators to form a closed-loop control system is currently still under investigation. An artificial sphincter and a sphincter sensor are actively being developed in the European project Healthy Aims (Tooley 2004).

Very small flexible endoscopes enabled by microsystem technology are used in urology to observe stones in the bladder (Technology for Industry 2005). Endoscopic optical coherence tomography (OCT) based on a microelectromechanical mirror has been used in thickness measurement and bladder tissue imaging. An endoscopic OCT system based on a microelectromechanical mirror to facilitate lateral light scanning is described. The front-view OCT scope, adapted to the instrument channel of a commercial endoscopic sheath, allows real-time cross-sectional imaging of living biological tissue via direct endoscopic visual guidance. The transverse and axial resolutions of the OCT scope are roughly 20.0 and 10.2 μm. Cross-sectional images of 500 × 1000 pixels covering an area of 2.9 mm × 2.8 mm can be acquired at five frames per second and with nearly 100 dB dynamic range (Pan *et al.* 2001).

The prostate is a relatively small chestnut-sized organ, which exists only in the male and is found right beneath the bladder at the beginning of the urethra. Current treatment of enlarged prostate involves surgical procedures via the urethra to physically remove or ablate the tissue using a heat source. Heat sources utilized today include laser treatment, microwaves, high-energy ultrasound waves, or heat. On the other hand, devices such as little plastic or metal tubes exist to be implanted in the prostate to keep the urethra open (Zeijlemaker *et al.* 2006).

3.14.1.5 Ophthalmic and Auditory Applications

3.14.1.5.1 Ophthalmic applications

3.14.1.5.1.(i) Introduction to ophthalmic applications The ophthalmic industry serves the medical needs of people with aberrations and diseases of the eye. Approximately 1 in 28 Americans 40 years of age and older is affected by loss of vision or blindness

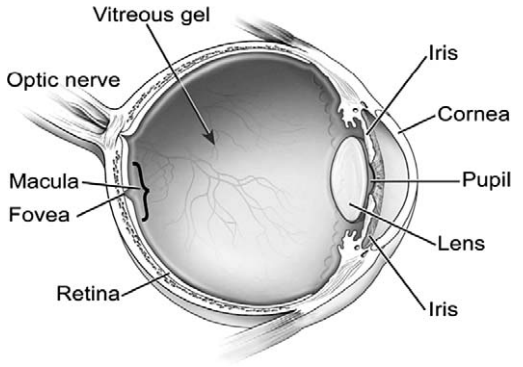


Figure 30 Schematic of the eye. (Source: National Eye Institute 2006 http://www.nei.nih.gov/)

(Gohdes *et al.* 2005). External medical devices to improve or correct vision range from prescription eyeglasses and contact lenses to nonprescription reading and sunglasses. Visual acuity, refraction, and intraocular pressure are some of the tests performed during eye examination. Ophthalmologists perform ophthalmic exams to check for diseases of the eye (American Academy of Ophthalmology 2006). Optometrists also perform many of these tests (American Optometrists Organization 2006). Laser-based corneal refractive procedures such as Lasik have made significant progress in providing long-term medical corrections for vision aberrations.

A schematic of the eye is shown in **Figure 30** (National Eye Institute 2006, National Institutes of Health 2006). Light enters the eye in a controlled manner through the pupil. Much like a camera, the light that enters this shutter-like structure is then focused by a lens. In the case of the eye, the lens focuses the light onto the retina. The macula is composed of photoreceptor cells. The macula is the part of the retina that receives and processes light and sends signals to the brain via the optic nerve. Degeneration of the macula photoreceptor cells, and optic nerve fiber damage, result in impaired or loss of vision. Further details on the anatomy and operation of the eye can be found in Eye Diseases and Conditions (2006), National Eye Institute (2006), and National Institutes of Health (2006).

An individual is defined as being legally blind when his or her best visual acuity is 20/200 or worse in their better eye or whose visual field, regardless of vision acuity, is restricted to 20° diameter. According to the American Foundation for the Blind (2006), approximately 1.3 million Americans are legally blind.

3.14.1.5.1.(ii) Ophthalmic diseases Some of the aberrations and diseases of the eye include myopia, hyperopia, presbyopia, astigmatism, cataracts, glaucoma, macular degeneration, and retina pigmentosa. Myopia or near-sightedness occurs when the eye's lens focuses incoming light in front of the retina. Hyperopia, or far-sightedness, occurs when the eye's lens focuses incoming light behind the retina. Eyeglasses or contact lenses typically correct for myopia and hyperopia, but do not slow their progression. The number of people in the United States estimated to suffer from myopia and hyperopia is currently 30.4 million and 11.7 million, respectively. Presbyopia results from the loss of the eye's ability to accommodate vision from near to far and is age related. People lose their ability to focus near vision and require reading glasses. The 119 million Americans 40 years and older are currently, or will soon, suffer from presbyopia (MedTech Insight 2005a, b). Astigmatism may result from an irregularly shaped cornea or lens and typically can be corrected with eyeglasses, contact lens, or refractive surgery. A comprehensive list can be found in Eye Conditions (2006).

A cataract is a clouding of the lens or the surrounding membrane that results in the obstruction of light (**Figure 31**) (National Eye Institute 2006, National Institutes of Health 2006). Cataracts are typically associated with older age and can be removed by surgery. Over the past 20 years, there has been widespread adoption of intraocular lens

Figure 31 A white congenital cataract. (Source: National Eye Institute 2006 http://www.nei.nih.gov/)

implants in conjunction with cataract surgery (MedTech Insight 2005a, b). Over 20 million people in the United States aged 40 and over have cataracts. This number is expected to increase to 30 million by 2020. Over half of the Americans aged 65 and over have cataracts. In 2005, cataract surgery-related products generated $3 billion in worldwide sales (MedTech Insight 2005a, b; National Eye Institute 2006, National Institutes of Health 2006).

Glaucoma is a condition where the optic nerve fibers die from lack of sufficient blood supply. Intraocular blood pressure is believed to play a role in this, as well as in lifestyle choices and comorbidities such as diabetes. External medical devices that measure intraocular pressure are used by ophthalmologists to diagnose and monitor glaucoma.

Retina pigmentosa and macular degeneration are retinal degenerative diseases. Macular degeneration is the most common cause of loss of vision in people over 55 years of age. Studies have found that both genetic and environmental factors influence this eye disease, but in cases of early onset of macular degeneration, the factors are primarily genetic. Retinitis pigmentosa is the name for a group of related eye diseases that affect the retina. All associated diseases are genetic in nature and cause a gradual degeneration of the rods and cones within the retina and gradual loss of vision, typically with the loss of peripheral vision preceding loss of central vision. Some retina pigmentosa syndromes include Usher syndrome, Bardet–Biedl (Laurence–Moon) syndrome, Best disease, choroideremia, cone–rod dystrophy, gyrate atrophy, and Leber's congenital amaurosis. There are at least 30 million people worldwide who suffer from retina pigmentosa and age-related macular degeneration. These are the two most common causes of as-yet untreatable blindness in developed countries (World Health Organization 2006). There are microsystems in development and in clinical trials studying the potential to provide therapy to patients suffering from age-related macular degeneration and retinitis pigmentosa. These are discussed below.

3.14.1.5.1.(iii) Ophthalmic current solutions Microsystems that include pressure and strain sensors, stimulating electrodes, microelectronics, and additional microsystems, such as Reichert Tonopen (Reichert Tonopen 2006), Keeler Pulsair Easy Eye (Pulsair 2006), Bausch and Lomb Proview Eyelid (Bausch and Lomb 2006), and Canon TX-F (Canon 2006), have been integrated into medical instruments to measure intraocular pressure. Medical

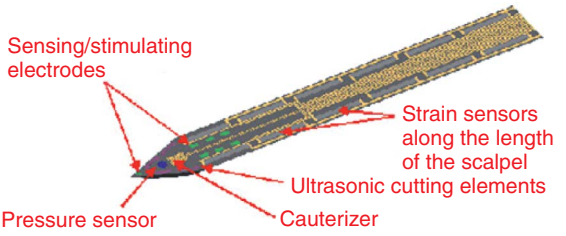


Figure 32 Verimetra's Data Knife, in development.

microsystems are also being integrated into medical instruments and tools to aid in ophthalmic surgery, such as Verimetra's Data Knife (Karoub 2003) (**Figure 32**). Significant progress has been made in the field of multivision intraocular lenses (IOLs) to replace lenses diseased with cataracts (**Figure 33**) (MedTech Insight 2005a, b).

3.14.1.5.1.(iv) Ophthalmic solutions on the horizon Shunts are currently implanted for patients with glaucoma to enable built-up fluid to exit. A feasibility study evaluating micromachined fluidics, sensors, and actuator valves integrated into drainage devices for glaucoma details potential applications of microsystems in this field (Howorth 2002) (**Figure 34**).

The major ophthalmic applications for medical microsystems are in prostheses that aid the near-blind or the blind to see. IOLs may become more widely used than just for the replacement for lenses diseased with cataracts. IOLs may be used as a pretreatment for precataracts. In addition, some companies are integrating microsystems with lenses for use in patients who require cataract surgery and who have glaucoma (Mesotec 2006).

Mesotec (Hanover, Germany) has developed a continuous intraocular pressure monitoring system that is powered externally by RF energy, reads the intraocular pressure, and delivers the data externally by RF telemetry (Mesotec 2006). An external antenna is required to deliver the RF signal and to receive the data from the implanted sensor, and can be integrated into special eyeglasses worn by the patient. The implanted device includes a foldable coil for coupling for the telemetry, a micromachined pressure sensor, and microelectronics. The sensor, microelectronics, and coil are integrated into a modified soft intraocular lens (**Figure 35**). This implantable medical device is intended to be used in patients that have both glaucoma and cataracts (Mesotec 2006).

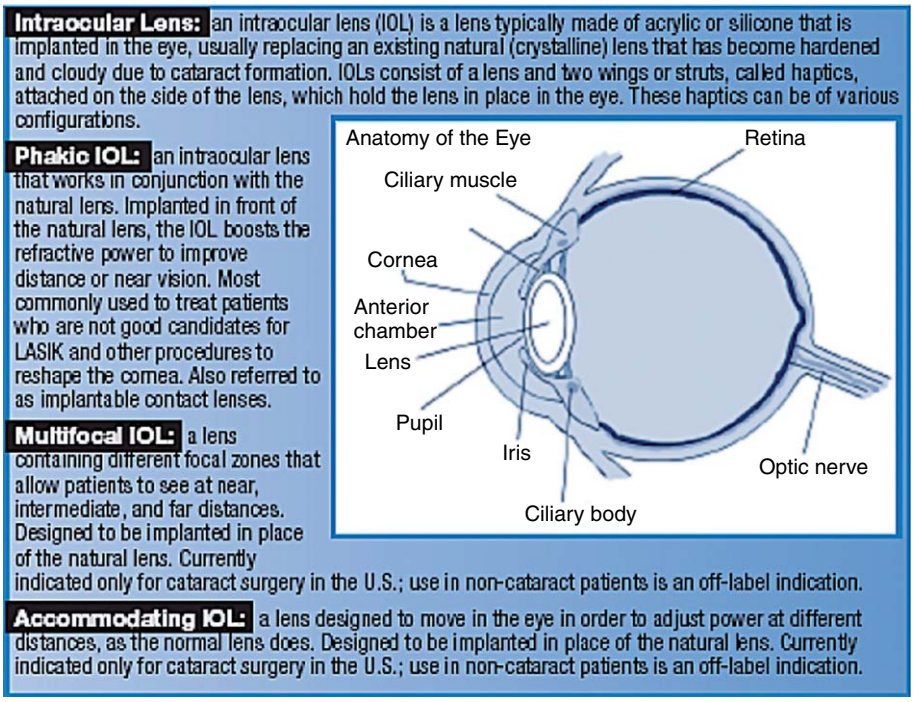


Figure 33 Intraocular lenses. (Source: MedTech Insight 2005 a *Novel Pain Products Address Growing Need* **7(2)**; MedTech Insight 2005b *Eyeing the Future of Ophthalmology* **7(11/12)**.)

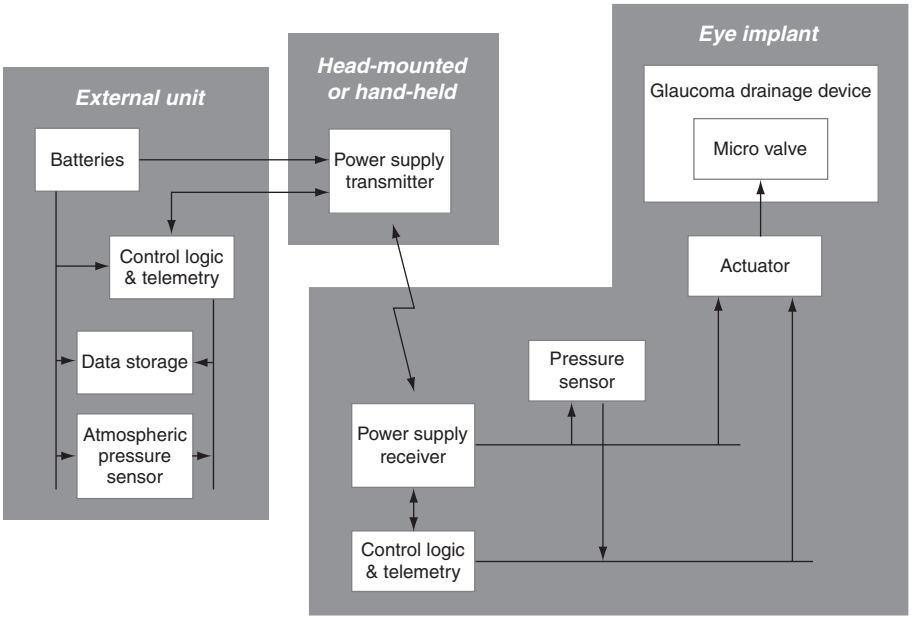


Figure 34 System design for active pressure-controlled glaucoma drainage device. (Source: Howorth D J 2002 Feasibility study for a micromachined glaucoma drainage device. M.Sc. thesis, Cranfield University.)

The most recent application of microsystems in ophthalmic applications is the design and manufacture of implantable medical devices to aid in vision improvement. Academic researchers active in this field include The Doheny Eye Institute and the Intraocular Retinal Prosthesis Group at the University of Southern California (The Doheny Eye Institute 2006), the Harvard–M.I.T. Retinal Implant Project

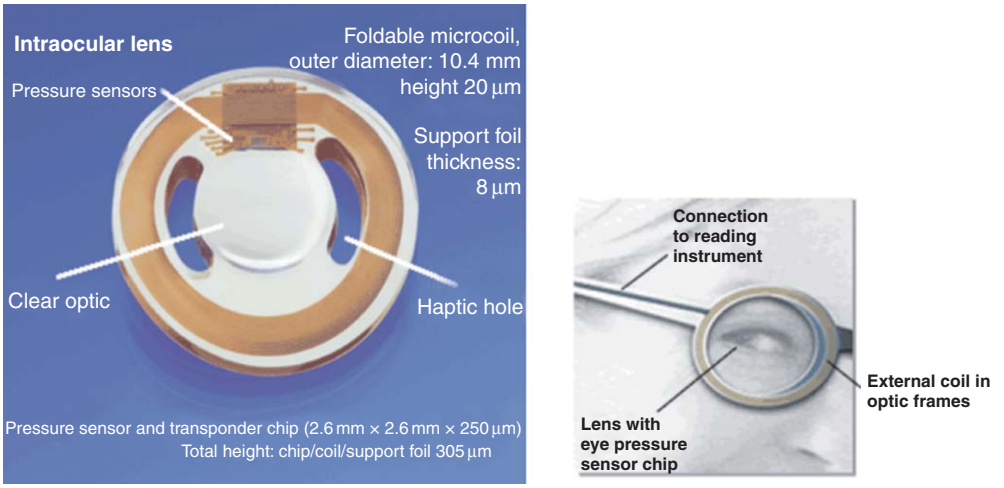


Figure 35 Mesotec-implanted device and external antenna. (Source: Mesotec 2006 http://www.mesotec.com/english/start.htm)

Group (The Retinal Implant Project 2006), Stanford University School of Medicine and the Kresge Eye Institute (Roberts 2006), University of Aachen and the Fraunhofer Institute in Germany (Fraunhofer 2006), the University of Uppsala in Sweden (Uppsala University 2006), and the Katholieke Universiteit in Leuven (Katholieke Universiteit Leuven 2006). Companies active in this field include Optobionics (Optobionics 2006), Second Sight (Second Sight 2006), and Mesotec (Mesotec 2006).

Medical microsystems are improving the lives of those with ophthalmic diseases and are bringing the hope of vision to the blind (About Vision 2006, All About Vision 2006, The Foundation Fighting Blindness 2006).

3.14.1.5.2 Auditory applications

3.14.1.5.2.(i) Introduction to auditory applications

The auditory industry serves the medical needs of people with aberrations and diseases of the ear. Otolaryngologists are head and neck surgeons who diagnose and treat disorders of the ears, the nose, the throat, and related structures of the head and the neck. Otolaryngologists and hearing health professionals perform auditory exams to check for diseases of the ear. External medical devices to improve or correct hearing range from behind-the-ear, in-the-ear, in-the-ear-canal, and on-the-body hearing aids to compact, portable listening devices (FDA 2006). Companies marketing hearing aids can be found (The Hearing Planet 2006). Implantable hearing devices include cochlear implants and partially implantable hearing systems.

A schematic of the ear is shown in **Figure 36** (Wikipedia 2006). Sound passes from the outer ear to the eardrum through the ear canal. The eardrum vibrates from the sound waves that are transmitted to the inner ear by the three bones termed the malleus, the incus, and the stapes. The inner ear contains the cochlea, the semicircular canals, and the vestibule. The vestibule is an oval cavity that contains the saccule and the utricle that communicate with the cochlea and the semicircular canals. The cochlear nerve, which is a part of the vestibular nerve, is associated with hearing. The vestibular nerve passes impulses from the inner ear to the brain and is associated with balance (Wikipedia 2006).

The market for medical devices to provide therapy for hearing loss is large. Twenty-eight million Americans suffer from some type of hearing loss. Age-related hearing loss affects 30% of adults 65 years of age and older and 1.7% of children under 18 years (National Institute on Deafness and Other Communication Disorders 2006). Hearing loss can result from exposure to damaging noise, certain medications, or from heredity, and can vary in severity from mild to total hearing loss.

3.14.1.5.2.(ii) Auditory diseases

The different types of hearing loss are termed conductive, sensorineural, and mixed hearing loss. Conductive hearing loss occurs when sound waves are not conducted to the inner ear and can be caused by the

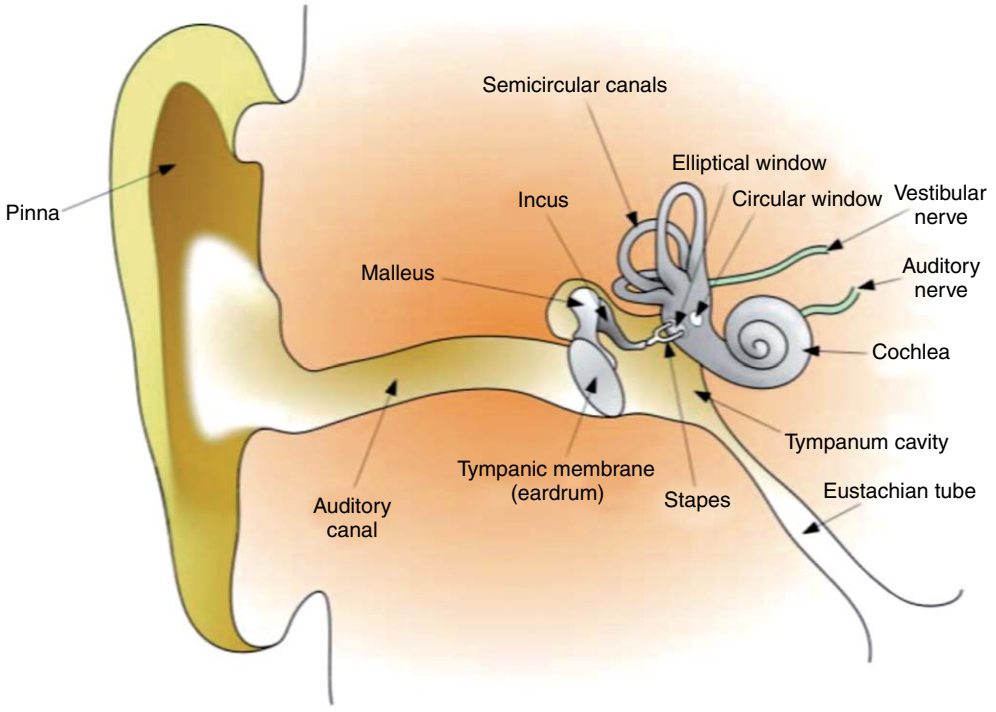


Figure 36 Schematic of the ear. (Source: Wikipedia 2006 http://en.wikipedia.org/wiki/Image:HumanEar.jpg).

buildup of earwax, an infection, fluid in the middle ear, or a punctured eardrum. Conductive hearing losses are typically corrected by medical or surgical therapies. Sensorineural hearing loss develops when the auditory nerve or hair cells in the inner ear are damaged by aging, noise, illness, injury, infection, head trauma, toxic medications, or an inherited condition. Sensorineural hearing loss usually cannot be reversed. Mixed hearing loss is a combination of both conductive and sensorineural hearing loss (Auditory and Vestibular Pathways 2006, Facts on Hearing Loss 2006, Hoversland 2006, National Institute on Deafness and Other Communication Disorders 2006).

3.14.1.5.2.(iii) Auditory current solutions

Microsystems integrated into external and implantable medical devices to aid in hearing aberrations include sensors and actuators such as microphones and implantable electrodes. Microphones pick up auditory signals from the environment and convert these to an electrical signal. Microphone technology and implantable electrode technology are discussed further in Section 3.14.2. Speech processes and microelectronics, such as amplifiers, digital signal processors (DSP), filters, transceivers, and receiv-ers, also play very critical roles in hearing aid and cochlear implant devices. The packaging of implantable systems must be biocompatible and biostable (Berkeley Sensors and Actuators Center 2006, Wireless Integrated Microsystems 2006).

There are a significant number of companies with hearing aid solutions (The Hearing Planet 2006). Three companies have FDA-approved medical implant devices such as cochlear implants that provide therapy for a chronic impairment or a chronic disease (University of Massachusetts Memorial Medical Center 2006): Med-El (Med-El 2006), Advanced Bionics (Advanced Bionics 2006), recently acquired by Boston Scientific (Boston Business Journal 2004), and Cochlear Corporation (Cochlear 2006). A diagram of an ear with a cochlear implant is shown in **Figure 37** (National Institute on Deafness and Other Communication Disorders 2006). The cochlear implant market growth is estimated to be between 15% and 20% per year. It is also expected that growth may accelerate due to the expanded implant eligibility criteria by the Medicare and Medicaid Services in April 2005 and due to the aging of the baby boomer

Figure 37 Diagram of ear with cochlear implant. (Source: National Institute on Deafness and Communication Disorders 2006 http://www.nidcd.nih.gov/health/hearing/hearingaid.asp).

generation (About Deafness 2006, American Academy of Otolaryngology Head and Neck Surgery 2006, Mishori 2006).

Research is continuing in the field of hearing improvement prosthesis. The University of Michigan at Ann Arbor, MI, has a focused research program on neural prostheses (Wireless Integrated Microsystems 2006). There is also significant research in the microsystem subcomponents, such as microphones, speech processors, and wireless telemetry (Berkeley Sensors and Actuators Center 2006, IEDW 2005, Loizou 1998, Wireless Integrated Microsystems 2006).

3.14.2 Emerging Technologies for Medical Applications

3.14.2.1 Novel Sensor, MEMS, and Microoptic Technologies for Medical Applications

3.14.2.1.1 Sensor technology for medicine and health care

Sensors are critical for the management of the patient's health, providing information to the managing physician as well as to therapy-providing devices. The accurate measurement of pressure is a key to the management of many conditions including hypertension, heart failure, hydrocephalous, glaucoma, and incontinence. Everyone is familiar with the sphygmomanometer as a method to obtain arterial blood pressure. Automated blood pressure systems from companies such as SpaceLabs Medical Inc. (Issaquah, Washington, DC, USA) and Sun Tech Medical Instruments (Raleigh, NC, USA) are now being used on ambulatory patients to manage hypertension.

Silicon pressure sensors are a key component in the monitoring of patients in a critical care unit. A cardiac catheter is placed in an artery or is wedged in a pulmonary capillary. The catheter has a lumen that is filled with a fluid. The pressure generated by the blood or the heart pushes on the fluid, which, in turn, pushes on the pressure sensor that is in a housing outside the body.

Silicon pressure sensors are also integrated directly into some catheters. Millar Instruments Inc. (Houston, TX, USA) has a broad product line of single and multiple sensors on catheters. The smallest of these catheters is approximately 0.67 mm in diameter. The PressureWire5 from Radi Medical System AB (Uppsala, Sweden) is even smaller, having a diameter of only 0.36 mm. In order to achieve these small sizes, ultraminiature silicon pressure sensors are required. **Figure 38** shows a number of pressure sensor dies manufactured by Silex Microsystems (Järfälla, Sweden). Each sensor is only 0.16 mm × 0.1 mm × 1.3 mm.

The silicon sensors on these acute catheters are coated with a polymer such as silicone rubber in order to protect the die from the effects of the blood as well as to insulate the electrical connection of the die. If the sensors are intended to be inserted into the body for years, however, a more robust approach is required. ISSYS (Ypsilanti, MI, USA) manufactures silicon sensors for chronic implantation using the dissolved wafer process (Gianchandani and Najafi 1992). In the dissolved wafer process, both a silicon wafer and

Figure 38 Ultraminiature (0.16 × 0.1 × 1.3 mm) silicon pressure sensor die manufactured by Silex Microsystems.

a glass or Pyrex wafer are processed simultaneously. In the process, a thin heavily boron (p^{++})-doped silicon layer forms one side of a capacitive pressure sensor and a metal pad on the glass forms the other side of the capacitive pressure sensor. Bonding of the silicon to the glass also forms a hermetically sealed pressure reference for the sensor. The p^{++} layer is far less affected by the corrosive nature of blood and, therefore, provides a reliable chronically implantable structure.

Mesotec has designed and built a fully integrated pressure/IC for the monitoring of glaucoma (Stangel *et al.* 2001). A photograph of the chip is shown in **Figure 39**. It contains a surface-micromachined pressure sensor, digital and analog circuitry, a circuit to harvest power from the RF, and communication circuitry. The chip is approximately 2.6 mm × 2.6 mm and is connected to an RF antenna that is incorporated into an artificial lens.

Figure 40 is a photograph of a chronically implanted pressure sensor manufactured by Remon Medical (Caesarea, Israel). The unique aspect of this design is that the power is delivered and the data are transmitted using ultrasound. The ultrasound signals from an external activator cause the acoustic transducer to flex, generating electrical energy that is stored in the capacitor. This stored energy is sufficient to activate the sensor for approximately 10 s, during which time it transmits the pressure values that are measured by the silicon pressure sensor.

Today, silicon accelerometers are designed into a number of implantable products. Because of the need

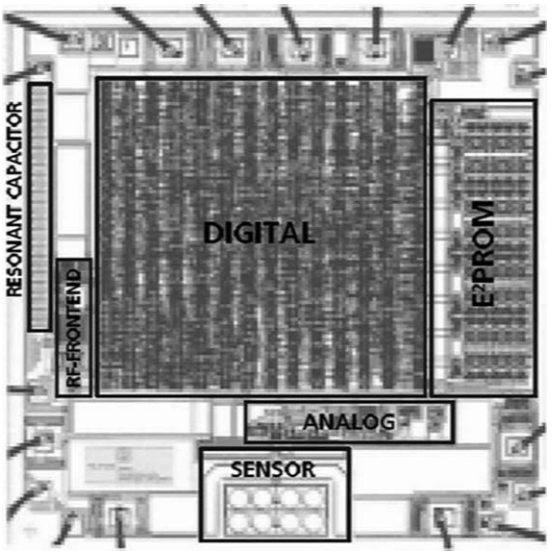


Figure 39 Integrated pressure sensor and IC from Mesotec for intraocular applications.

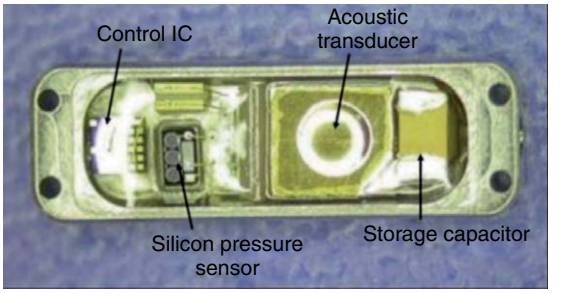


Figure 40 Photograph of the internal components of the initial design of an ultrasound-based chronically implanted pressure sensor. (Courtesy: Remon Medical.)

for extremely low power requirements for the sensor, the accelerometers are usually capacitive devices. **Figure 41** shows a diagram of the symmetric accelerometer structure from VTI Technologies.

This accelerometer is made by bonding three separate silicon wafers together to form a triple stack. The accelerometer has the advantage of being very rugged and immune to temperature effects. Other silicon accelerometers such as comb structures and three-axes accelerometers will be incorporated into products to lower the cost and increase the functionality.

The opticath catheter from Hospira (Lake Forest, IL, USA) is a device used in acute critical care to measure the oxygen saturation of venous blood. The system uses three wavelengths of light transmitted down an optical fiber into the blood to calculate oxygen saturation as well as hematocrit based on the reflection of the three lights.

Sensors for Medicine and Science Inc. (Germantown, MD, USA) is designing and testing an implantable sensor that relies on fluorescence to measure glucose concentration under the skin. A

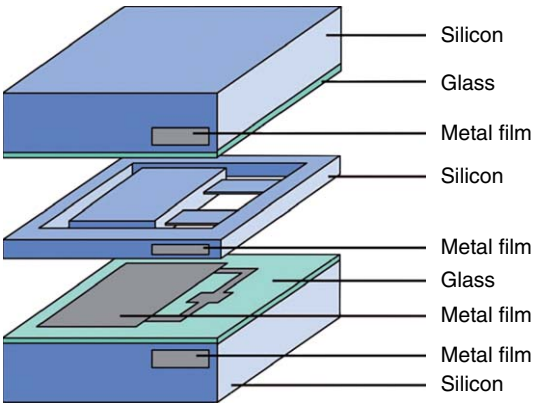


Figure 41 Drawing of the capacitive triple-stack silicon accelerometer from VTI Technologies.

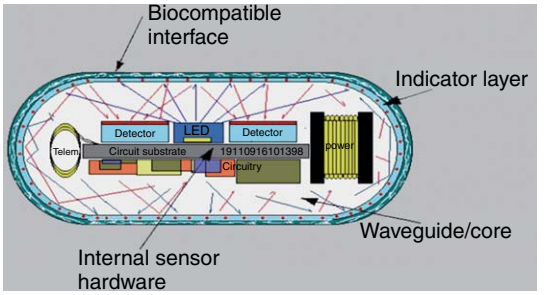


Figure 42 Diagram of a fluorescence-based implantable glucose sensor capsule. (Courtesy: Sensors for Medicine and Science.)

diagram of the sensor is shown in **Figure 42**. The sensor capsule contains an LED to excite the fluorophore in the indicator layer, optical detectors to measure fluorescence for both the signal and reference channels, and a circuitry to drive and control the optoelectronic components. The implanted capsule also has a circuitry to harvest energy from an external inductive power source.

3.14.2.2 Novel Electrode Technologies for Medical Applications

3.14.2.2.1 Introduction

A large portion of clinical applications of medical microsystem technologies deals with electrical stimulation and/or sensing. Section 3.14.1 mentions examples of IPGs such as cardiac pacemakers, cochlear stimulators, neurological stimulators for spinal and deep brain, sacral nerve stimulation, and gastric stimulation, and there are options to expand this into other clinical areas (Schneider and Stieglitz 2004).

Generally speaking, an electrical stimulus is generated in an IPG and electrically conducted to an electrode positioned at the optimum therapy delivery site. Although most of the electrodes currently appearing in medical products are made using conventional fabrication technology, numerous examples of microfabricated electrodes are currently being investigated for clinical applications. Microelectrodes offer the advantage of exact electrode positions, small size, multiple electrodes, and active circuitry (Kewley *et al.* 1997). Microsystems offer possibility for a higher degree of miniaturization and increased numbers of electrodes (Stieglitz 2002). This can lead to concepts featuring adaptable therapy and easy implant procedures (Receveur et al. 2005). We therefore feel that it is good to devote a part of this chapter to this subject.

This section has an emphasis on the technological aspects of microelectrodes. It can be seen as a cross-analysis of the information provided in the previous sections per clinical application area. The various electrodes are ordered in terms of shape, intelligence, materials, and biostability. At the end the information is summarized.

3.14.2.2.2 Electrode shapes, manufacturing process, and intelligence

We distinguish 1D electrodes, 2D electrodes, 3D electrodes, Smart Electrodes, and electrodes for external medical use. By 1D electrodes we mean that the general organization of electrodes is along a line, the shape of the electrode itself does not have to be 1D. One group describes 1D microelectrode arrays that are entirely flexible, made using a boron/EDP approach. It is 40 mm long and 4 μm thick (Bell *et al.* 1997).

2D electrodes are flat electrodes arranged in a 2D way. There are examples of planar fabrication processes that raise the electrodes somewhat in 3D or 2.5D. Electrodes of various materials can be made using photolithographic fabrication processes on a variety of substrates, such as the ones used for stimulation of the retina (Chow *et al.* 2002, Grumet *et al.* 2000, Hammerle *et al.* 2002, Weiland *et al.* 2002, Zrenner *et al.* 1997) and for brain research (Gosselin *et al.* 2004). The 2D arrays of electrodes with micrometer dimensions are also made using wires (Smit *et al.* 1999).

There are various ways reported to make 3D type of electrodes using microfabrication technologies. By making a 3D shape, there might be a gain in electrode surface area compared with 2D (Sachs *et al.* 2005). Generally speaking, 3D microelectrodes are being used to penetrate something. There are many examples of 3D microelectrodes (Birch 2004, Griss and Stemme 2003, Griss *et al.* 2001, Roxhed *et al.* 2005a, b, Schmidt *et al.* 1996, Thiebaud *et al.* 1997) and in the following only a couple of them are highlighted. One group uses probes that are entirely realized in a silicon technology and incorporates 10 recording sites each, with contact areas ranging from 100 to 2500 μm^2, spaced at 100 μm intervals. The probe substrate structure is obtained from a thinned silicon wafer by the application of a two-electrode electrochemical etch stop in potassium hydroxide (KOH) and a subsequent reactive ion etch (RIE) of silicon in sulfur hexafluoride (SF_6). (Peeters *et al.* 1991). Another 3D microfabricated structure is etched into a biocompatible polyimide insulation layer. The polymer layer is micropatterned

by electron beam lithography and etched in an oxygen plasma (Bauerdick *et al.* 2003, Sachs *et al.* 2005). A third group uses a combination of silicon and polymer material to achieve the desired properties. A 5–10-μm-thick silicon backbone layer is attached to the tip of the electrode to enhance the structural stiffness. This stiff segment is then followed by a 1-mm flexible segment without a silicon backbone layer. The fabricated implants have three shanks with five recording sites (20 μm × 20 μm) and two via holes (40 μm × 40 μm) to promote tissue attachment on each shank (Lee *et al.* 2004). Still another group uses a combination of dry and wet etching to form probes with 16 electrodes on each tip. Phase one: thin-film deposition to form the electrodes, bondpads, and insulation. Phase two: etches to define probe shape. Phase three: removes masking layers and package (Kewley *et al.* 1997). The following process yields probes of 1 mm long. A monolithic piece of silicon (6–20 ohm cm^{-1}) is micromachined and acid-etched to form the basic substrate of the electrode array – 100 needles of length 1 mm, each separated by a glass dielectric barrier and arranged in a 10 × 10 configuration with an interelectrode spacing of 400 μm. The silicon needles have electrodes on the first 50–100 μm of each tip. The unmetallized silicon electrode shafts and the substrate are insulated with a 2–3-μm layer of polyimide (Nordhausen *et al.* 1993, Rousche and Normann 1999). As a final example, we mention a 3D multielectrode array that is basically created using a sawing procedure. In 1″ (111) silicon wafers, approximately 4 mm thick, 600-μm-deep grooves are sawn, at distances of 120 μm apart. The walls of the silicon that result from this procedure are shortened in such a way that walls with three different heights result: 600, 400, and 200 μm. Perpendicular to the first sawing direction, new grooves of 600 μm depth are sawn, resulting in rows of silicon needles, with differing heights. The needles in a row have equal heights. The needles are 120 μm apart, and have a size of approximately 55 μm × 55 μm. They get their final size and shape from an RIE process, using an SF_6/O_2 plasma. A flip chip technique to connect the brush to (de)multiplexing circuitry, in which current sources and buffer amplifiers are also integrated, is being studied (Rutten *et al.* 1995).

Others assemble active 3D arrays out of 2D arrays of microfabricated electrodes. Gold beams on the probes are orthogonally bonded to pads on the platform using ultrasound. Active 3D probe arrays containing on-chip CMOS signal processing circuitry have also been made (Bai Qing *et al.* 2000). These arrays can be made in different sizes using programmable platform-based fused interconnects (Gingerich *et al.* 2001)

Microsystem technology provides options to add functionality directly on the electrodes. An 18 polar hybrid cuff electrode with miniaturized multiplexer reduces the number of cables to 4 by integrating this intelligence (Rodriguez *et al.* 2000, Stieglitz 2001). Another example is the actor component of a retina implant stimulator, which is a flexible active silicon multielectrode used for electrostimulation of the retinal ganglion cells (Schwarz *et al.* 2000) or electrodes integrated with CMOS circuits created using a dissolved wafer process (Najafi *et al.* 1985).

In most cases microelectrodes are also being used for external medical applications for biomedical research, not directly for diagnosis or therapy. Some examples are the use of neuronal networks on multielectrode arrays as biosensors (Gross *et al.* 1995); neurons coupled to transistors (Fromherz and Stett 1995, Fromherz *et al.* 1991); electrical or chemical neurotransducers (Bartic *et al.* 2003); 2D array of gold pads on CMOS for *in vitro* cell measurements (Berdondini *et al.* 2004); electrodes on polyimide to measure high density surface EMG (Lapatki *et al.* 2003) and hybrid biosensors: micromachined electrode arrays with living cells above then used for sensors (Kovacs 1998).

3.14.2.2.3 Electrode materials, performance, and biostability

Materials used for electrodes include platinum (Griss *et al.* 2001, Schneider and Stieglitz 2004), platinum black (Rizzo *et al.* 2003), iridium (Kewley *et al.* 1997; Najafi *et al.* 1985, Schneider and Stieglitz 2004), iridium oxide (Rizzo *et al.* 2003), gold (Chow *et al.* 2002, Najafi *et al.* 1985, Rutten *et al.* 1995), nickel chromium (Smit *et al.* 1999), TiW/Au (Peeters *et al.* 1991), copper, silver-coated, roughened, and chlorized (Lapatki *et al.* 2003), nanoporous titanium nitride (Bauerdick *et al.* 2003, Sachs *et al.* 2005), and four-part multilayer of Pt, Ti, W, and Pt (Rousche and Normann 1999). Gold, platinum, and silver are used for recording, and they can be chlorodized at the surface (Prohaska *et al.* 1986).

Important properties of microelectrodes are of course the electrical properties such as impedance, charge transfer capability, and stimulus thresholds. Threshold currents for phosphene generation in the cat's visual cortex with trains of biphasic pulses can be as low as 1.9 μA (Schmidt *et al.* 1996). Required electrode areas range from 1 μm^2 to normal of

$100\,\mu m^2$ to up to $5000\,\mu m^2$ (Peeters *et al.* 1991, Prohaska *et al.* 1986). Electrode–electrolyte impedance depends on the charge transfer processes at the interface (Prohaska *et al.* 1986). Researchers are investigating the use of microfabrication technologies to improve these properties. Examples are electrodes treated to lower the stimulation threshold. Sputtered iridium oxide films yielded the best performance, with a big influence of the oxygen concentration during sputtering (Mokwa 2004).

Arrays of microelectrodes with improved charge transfer capacity are realized by using a 3D micropattern. This yields a surface gain compared to non-3D electrodes resulting in a significant reduction of impedance by a factor of 0.3–0.2 (Bauerdick *et al.* 2003, Sachs *et al.* 2005). The structures in photoresist, pyrolize, become carbon structures mimicking fractals for application as electrode materials (Madou 2005).

The body presents one of the harshest environments since the immune system actively tries to deal with foreign materials. Biostability is therefore difficult to achieve, especially over longer periods of time. The literature contains numerous examples of issues with implantable microelectrodes. Gold is reported to dissolve after subretinal implantation (Chow *et al.* 2002). Photodiode arrays that were implanted for 18 months in the subretina of the rat showed pit corrosion in silicon and silicon dioxide. Titanium nitride is stable (Hammerle *et al.* 2002, Zrenner *et al.* 1997). Silicon dioxide can be used as an insulating layer. It has good adherence to conducting layers and no pinholes but Na^+ and K^+ diffuse through it, creating impurity conduction bands and lowering the insulating properties. Silicon nitride layer can be added to block ion diffusion (Prohaska *et al.* 1986). The other side of biostability is biocompatibility, i.e., the implanted material not causing harm to the body. Extensive work is being performed but is not reviewed here.

3.14.2.2.4 Summary

Electrodes are made on a flat substrate with thin-film techniques (2D array), plating (electroplating or electroless) creating 2.5D, or with combinations of various processing steps to create poles with the possibility to have multiple electrodes along the length of each pole (3D array). The electrode materials are Pt/Au/Ir/Ir oxide/Ag/TiNi/C. There are several concepts in which the electrodes are made on top of electronic circuitry providing some form of intelligence (multiplexing/current sources). The substrates can be thinned or processed to the extent that they become flexible or can be made of a flexible material from the start. Corrosion of silicon and silicon dioxide after implantation in the retina has been reported.

3.14.2.3 Microelectronics Technologies for Medical Applications

3.14.2.3.1 Introduction

We are all familiar with microelectronics and microsystems. We use products that include them everyday. We have become so accustomed to the functionality they provide that we take for granted their capabilities and the speed with which new products expand these capabilities. In addition, we have all seen many articles that describe various technical and nontechnical aspects of microelectronics. This brief overview of microelectronics for medical applications does not duplicate that wealth of information. Instead we focus on the role of microelectronics in medical equipment, instruments, and implantable medical devices, as well as on the requirements of medical applications. We also briefly discuss some of the key differentiators between microelectronics for consumer and medical applications.

Medical applications require the highest levels of quality and reliability. In addition, minimizing size and power consumption is essential for implantable medical devices. Clearly these are the attributes that we desire in almost all of our electronic products. Advances in reliability, size, and power drain have helped drive widespread acceptance of electronic products. This has been a major factor in the global spread of television, personal computers, cell phones, laptops, and personal music players. The key difference is that, as in all high-reliability applications, requirements for medical applications are more stringent, and much more thoroughly tested, than for many commercial products. Likewise, expectations on the quality, reliability, and performance of medical equipments, instruments, and devices are also much higher than for consumer products.

3.14.2.3.2 Microelectronic components and technologies

One place to begin this overview of medical applications of microelectronics and microsystems is a brief discussion of what is included in the broad term microelectronics. Although we all are aware that microelectronics is a key element in many products we use everyday, it is still a very complex subject

encompassing many disciplines, components, and processes. It may help to set the stage for this overview by listing some basic components and processes used to develop and manufacture microelectronics for medical equipment and devices. Within medical products, ICs include many elementary circuits and functional blocks including the following:

- Digital logic circuits
- Analog circuits
- Memories, including flash memory
- Microprocessors (μP)
- Analog-to-digital converters (ADC)
- DSP
- Digital-to-analog converters (DAC)
- RF circuits

and many other types of circuits. In addition, electronic devices include discrete semiconductor and passive components such as the following:

- Capacitors
- Resistors
- Inductors
- Transformers
- Substrates
- Connectors

and many variations of these components, as well as many other components.

ICs for medical products can be developed in several ways. These include the following:

- Commercial off-the-shelf ICs (COTS)
- Programmable logic arrays (PLA), field programmable logic arrays (FPLA), complex programmable logic devices (CPLD)
- Application-specific off-the-shelf ICs (ASSP)
- Full-custom application-specific ICs (ASIC)

Of these four approaches, COTS are viewed as the lowest risk and the lowest cost approach. Using them is also the quickest way to develop electronics for a product. Programmable arrays and logic devices share many of the same characteristics, but can be modified electrically if the original programming of the logic array needs to be updated. They can also be modified at the end of the production line if required to support changes in the product mix. Application-specific standard products are intended to provide some of the optimization of full-custom ICs and some of the advantages of off-the-shelf products. These ICs may also include some sections that are electrically alterable to incorporate some of the advantages of PLA as well. Development of full-custom ICs is the most expensive, involves highest risk, and often takes the longest to achieve the required functionality. But they have the advantage of enabling the most optimized design. For a given level of complexity this minimizes the size of the IC and its power drain. It also enables the development of custom circuits to improve the performance and the reliability of the final product. Clearly all of these approaches offer specific advantages and many medical products incorporate a mix of them, although full-custom ICs are more likely to be used in implantable medical devices than in other medical products.

Development and manufacture of integrated and discrete components require a large number of more fundamental technologies. Examples of these fundamental technologies include the following:

- Semiconductor and wafer fabrication technologies, technologies for packaging silicon die such as die bonding, wire bonding, chip-and-wire, flip-chip, ball grid array (BGA) and micro ball grid array, chip scale packaging (CSP), stacked die packaging, and system-in-package (SIP) packaging
- Printed wiring board (PWB), flex tape, and hybrid technologies
- Electronic materials

This is only a sample of some fundamental circuit blocks, components, and technologies required to develop and manufacture electronic products. But it does suggest the breadth of disciplines, components, and processes required.

3.14.2.3.3 Microelectronics in medical applications

Given the complexity of microelectronic circuits, components, and technologies, a high-level overview of the requirements of medical applications might be in order. Medical instruments and devices must achieve several things. First and foremost, they must be safe and reliable. They must also benefit patients either by providing valuable diagnostic and monitoring data or by delivering effective therapies. Also, in the case of implantable medical devices, they must be small and safe to implant. Clearly they must be useable by physicians, healthcare workers, and caregivers. User interfaces must be adequate to assure that medical instruments and devices are controllable enough to consistently perform the required functions. Requirements like these help to define criteria that drive the selection of microelectronic designs, components, technologies, and test methods.

3.14.2.3.3.(i) Complexity and quality Advances in the capability and sophistication of new medical products can drive the increases in the complexity of the microelectronics. In terms of medical applications, advances in product capabilities include the following:

- Collecting more data
- Analyzing it more accurately
- Responding to it more rapidly
- Performing more sophisticated actions
- Accomplishing all this while being easier to use

In addition to collecting data and responding appropriately, medical devices must process, store, and communicate the data to physicians. This can provide useful information as they perform their analyses of patients' conditions and medical needs. As worldwide information infrastructures have continued to become more tightly interconnected, communicating this information may also include interfacing with computers and other equipments, as well as connecting to physicians and healthcare workers through public communication infrastructures such as telephone systems, cell phones, and the Internet.

3.14.2.3.3.(ii) Patient safety Although no product will gain long-term acceptance unless it is of value to users, without question the first priority for any medical product is patients' safety. It is absolutely imperative that medical products must safely deliver the highest quality performance, and do so reliably. Microelectronics has proven, throughout many decades, that with proper design and thorough testing, very high levels of quality can be achieved. Second, medical products must continue to function correctly throughout the entire life of the product. Thus the reliability of medical products is as important as quality.

3.14.2.3.3.(iii) Cost vs. quality In some technologies, there may appear to be trade-offs between cost and quality. Microelectronics is one area where this is definitely not true. Increasing the level of integration within microelectronic circuitry reduces the number of individual components as well as the number of interconnections required to connect these circuits to inputs and outputs of the finished product. This decreases the number of potential points of failure and increases the intrinsic level of reliability the product is capable of achieving. Therefore, higher levels of integration tend to produce correspondingly higher levels of reliability for a given technology. We are all familiar with the growth in the complexity of electronic products. One example of this is the growth in the complexity levels of leading-edge ICs over the past 40 years. If we take the number of transistors on a silicon die as an indicator of complexity, the complexity level of leading-edge ICs has grown by a factor of approximately 100 million over that time. Also, over these four decades, the cost per transistor has dropped to approximately a ten-millionth of its initial cost. This has been achieved while continuously improving quality and reliability levels attainable with microelectronic circuits.

Since the early 1990s the picture has become a bit more complicated. Some data suggest that failure-in-time (FIT) rates and the rate of failures in the first couple of years have increased for individual chips. In the worst case, they may have increased at approximately the same rate as the growth in the complexity of the die. The usual recommendation for counteracting this trend is to derate operating voltages and temperature. This, and the fact that higher levels of integration will reduce the number of separate components, can improve the attainable level of quality and reliability to historical levels, or better, on a function-by-function basis. Other measures that can augment these efforts to assure the highest levels of quality and reliability are mentioned in some of the following paragraphs.

3.14.2.3.3.(iv) Complexity vs. quality To successfully increase the level of complexity of ICs by 100 000 000 times, the levels of predictability and consistency of wafer fabrication and microelectronics design, processing, and testing have to be raised to extremely high levels. This requires correspondingly high levels of process control throughout these operations. This progress would not have been possible without major advances in process control technologies and methodologies. This includes advances in quality metrics and methodologies, such as six sigma. It also includes advances in hardware and software design tools, as well as in wafer processing, manufacturing, and test equipments. Obviously, much of the progress in computers used for circuit design and in equipments used for processing, manufacturing, and testing electronic products was possible only because of advances in microelectronics. This combination of advances in microelectronic technologies, computers, and equipments has continued to raise the predictability and consistency of processes and technologies used to develop, fabricate, and test microelectronics. This in turn has raised the quality level of microelectronics while reducing both the size and the cost of electronic

products on a function-by-function basis. These advances are part of the mainstream of wafer fabrication processing, microelectronic manufacturing, and state-of-the-art testing. Developers and manufacturers of medical products take full advantage of these advances in quality and reliability, but in themselves they may not be enough. In the next few paragraphs we deal with additional requirements to assure the highest quality and reliability in medical applications.

In some cases, advances in microelectronic technologies have been used to grow the complexity of products as rapidly as possible. Electronic games and Flash memories are examples of how quickly microelectronics can enable the expansion of the complexity and the capability of electronic products. In other cases, advances in microelectronic technologies and processes have not been pushed to the complexity limit, but rather have been used to produce products that are simultaneously both more reliable and more sophisticated. Implantable medical device is an example of this type of product evolution.

Products that become more reliable as their complexity grows may seem inconsistent with some personal experiences with computers and cell phones. This highlights a key difference between consumer electronics and microelectronics for medical applications. As in high-reliability products, medical applications require far more stringent quality, reliability criteria, and testing.

3.14.2.3.3.(v) Design for quality Quality must be designed-in. Good designs can be compromised by inadequate manufacturing and test processes, but these processes cannot increase the levels of quality and reliability above the constraints imposed by the design of a product. This includes design of both the circuits themselves and processes used for fabrication and testing.

3.14.2.3.3.(vi) Wafer yields and quality Design and integration of wafer fabrication processes is extremely important to wafer fab management teams. The financial consequences of poor yields are very real and visible. Whenever wafer fabs run at full capacity, higher yields translate into shipping more wafers. Whenever wafer fabs run at less than full capacity, higher yields reduce the cost of operations and work in process. This may help to reduce adverse impacts of the idle capacity. Yields have significant financial impact on both of these situations. This is especially critical as the cost of constructing wafer fabs is now in the multi-billion-dollar range. Therefore wafer fab management teams focus on achieving the best wafer yields possible.

One consequence of higher wafer yields is that quality and reliability tend to be directly proportional to yield. Therefore medical product manufacturers, even when they rely on full-custom ICs, leverage higher-yielding processes for their products. They match their designs to available wafer fabrication processes that are proven to yield the highest-quality and -reliability ICs. For implantable medical devices, they may also use design methodologies optimized for high reliability as well as for extremely low power drain. Proprietary circuitry may also be added to assure reliable long-term performance of microelectronic circuits. In this way, complex and sophisticated functions are incorporated into medical products without compromising quality or reliability.

3.14.2.3.3.(vii) Design and test for reliability There are also well-established methods for identifying potential reliability and performance factors. These can be analyzed and effective countermeasures are designed into the electronics and manufacturing processes before a product is released to production. Adding additional functionality to assure reliable performance of the end product also increases its complexity. These increases in complexity are almost always transparent to users of high-reliability products. Although invisible, this one-way growth in product complexity can be used to simultaneously increase the functionality of a product and its reliability. This is one more aspect of microelectronics that contributes to making reliable, high-quality medical products available to physicians and patients.

3.14.2.3.3.(viii) Reliability In addition to high-quality components, processes, and fundamental technologies, there are many other factors that must be considered to assure the highest levels of quality and reliability in medical instruments and devices. It is essential that new technologies not only produce products of highest quality, but also wearing out, and other failure mechanisms must be analyzed in depth. This is not trivial. Semiconductor technology has reached the level of extreme miniaturization where critical layers on a die can be less than 10 molecules thick. Analysis of wear-out and other failure mechanisms at these dimensions requires extensive knowledge of metrology and advanced laboratory instruments and skills. At these dimensions even defects that are much smaller than the wavelength of light can be catastrophic.

3.14.2.3.3.(ix) Environmental factors and reliability Besides physical defects within microelectronic circuits, many external factors can also affect the performance and reliability of advanced products. These external factors include robustness to:

- Electrical charge and interference
- Magnetic and electromagnetic fields
- Other sources of radiation
- Subatomic particles
- Temperature extremes
- Mechanical shock

All of these must be analyzed and appropriate steps must be taken to assure the quality and reliability of products. These measures may include redundancy and other electrical and mechanical design approaches. Also, they may include the addition of various forms of stress testing as well as self-test capabilities built into the product itself during manufacturing process. All of these measures are well known within the disciplines that create microelectronic circuits, processes, and testing. While well understood by experienced design, manufacturing, test, quality, and reliability engineers, these measures are applied extensively and meticulously in the development and production of implantable medical devices. In some cases they may add to manufacturing cost and cycle time. But this may be offset, on a total life cycle cost basis, by improvement in the performance and longer lifetimes of medical products.

3.14.2.3.3.(x) Technologies vs. quality In addition to assessing the performance of new technologies against all the factors described in the preceding paragraphs, microelectronic technologies for medical applications must be thoroughly proven with millions of hours of product experience in other applications. Therefore medical products are seldom based on technologies just at the beginning of their use in commercial products. But this does not mean that technology in medical applications lags behind the state of the art. As the global electronics industry approaches the two-trillion-dollar level, it routinely drives very rapid learning curves for new technologies. By intelligently monitoring the learning curves of these technologies, developers of medical equipment and devices can determine which technologies have proven their quality and reliability enough to be incorporated into applications that require the highest levels of quality and reliability.

3.14.2.3.3.(xi) Size Size is also an important consideration in implantable medical devices. In these applications, smaller devices tend to be associated with greater patient safety and comfort. Microelectronics is essential for miniaturizing implantable medical devices. Ongoing advances in microelectronic components, processes, and technologies enable development of sophisticated products that perform many complicated functions in small devices without compromising quality or reliability.

Another key factor in the size of implantable medical devices is power consumption. Clearly this is closely related to advances in the energy density of power sources. Even with the most advanced power sources, it is still important to use power as efficiently as possible. Although some implantable devices can be recharged, most depend on a single primary battery for power. Therefore lifetimes of implanted devices are heavily dependent on the power drain of microelectronic circuits. Many design and processing techniques have been developed to maximize the power efficiency of microelectronics in these devices. These power management techniques, while well understood by experienced circuit designers, are applied extensively and meticulously in the development of implantable medical devices – so much so that even though power management is an increasingly important consideration in consumer electronics, the power drain of many medical devices is still a small fraction of the power drain of typical cell phones and other handheld devices.

3.14.2.3.4 Future

There is little reason to conclude that the progress of microelectronic technologies will end any time soon. In fact, it appears that progress may well continue at roughly the same rate as it has for decades. There are major challenges ahead in some fundamental technologies, but this more likely indicates that the form of some future progress may be different, not that the pace of progress will slow. At the same time, with the present interest in nanotechnology, and other alternatives to silicon-based technologies, there are many interesting potential substitutes, or extensions, of present microelectronic technologies. Ongoing advances in MEMS and MEMS packaging may also enable additional functionality for microelectronic circuits, especially in the areas of sensors and communications. This could reduce the size of future products while enabling them to be more connected with their environment as well as other electronic devices in their vicinity.

3.14.2.3.4.(i) Disruptive breakthroughs vs. rapid evolution It is important to remember that disruptive breakthroughs have been few and far between in IC technology. One of the reasons is that the rate of evolutionary change in silicon-based technologies has been so fast that it often progresses at a faster rate than do new technologies. Furthermore, this rate is likely to continue for the foreseeable future. This is unlike examples of new technologies that are inferior to established technologies when they are first implemented, but improve faster than mature technologies. In those cases, there are many examples of new technologies overtaking and ultimately displacing older, more established technologies. Unlike this scenario, whenever a new technology is both inferior to a mature technology and improving at the same rate or slower, the existing technology is unlikely to be displaced.

The historical rate of progress in silicon-based microelectronic circuits is likely to continue for at least another decade. Therefore, it is likely that the next 10 years will be dominated by a continuation of rapid evolutionary advances, not disruptive breakthroughs by totally new technologies or circuit architectures. Evolutionary advances are likely to occur in many areas of microelectronics, including component sizes and integration, die and circuit packaging, wafer processing and photolithography, semiconductor device design, and power management. The combination of these advances will continue the historical rate of progress in microelectronics, or may even accelerate it (Tummala 2006).

Even if the likelihood of disruptive breakthroughs in fundamental microelectronic technologies is low, this is not an indication that the rate of advances in medical applications will slow. Quite the opposite, the rate of progression may even accelerate in the future. Many factors may indicate that the next few decades will be more exciting than ever. Safety, quality, reliability, and sophistication of medical instruments and devices will continue to advance. At the same time, medical devices will continue to incorporate more functions within the devices of same size, or smaller, as energy sources advance and circuits become both more complex and more efficient. These medical products will communicate with each other, as well as with physicians and caregivers to help them improve effectiveness of diagnoses and treatment.

Can this truly be achieved without disruptive breakthroughs in technology? Yes, continuation of trends currently dominating advances in electronic products is fully capable of achieving progress in medical applications just as in many other products. Two fundamental trends will continue to drive these advances: convergence and the rapid evolution of fundamental technologies. These trends are powerful drivers of progress in both consumer and high-reliability applications, including medical applications. With expanding research and support of developments in biotechnology, nanotechnology, bioinformatics, and the convergence of these technologies, many new and exciting opportunities for advances in medical diagnostics, prevention, and therapies are already emerging. The continued evolution of microelectronics will provide a platform to integrate these and enable reduction to practice of great numbers of opportunities that will benefit patients and their families. Advances in global communication and computing infrastructures, as well as continuing advances in the complexity, quality, and reliability of medical products, may enable systems and capabilities that have the promise of providing patients with access to these benefits anywhere, anytime.

3.14.2.3.4.(ii) Convergence The size of many products will continue to shrink as well. This combination of networked medical applications, integrated with expanding applications of biotechnology and nanotechnology, will make possible creation of medical equipment, instruments, and devices that bring new and more sophisticated diagnostic and therapeutic capabilities to physicians and healthcare workers. These advances, coupled with advances in the informatics capabilities of networked systems that integrate personal health, history, and environmental data, will provide physicians and caregivers with access to information and resources they have never had before. This can enable simultaneous advances in both personal medicine and global health. Clearly, we stand at the threshold of many life-changing advances in healthcare.

3.14.2.4 Communication Technologies for Medical Applications

Wireless biomedical telemetry is used in hospitals, clinics, homes, ambulances, and other healthcare institutions. This telemetry is typically part of an external biomedical monitoring and diagnostic system, a body-worn device, or an implantable medical device designed to gather data, administer therapy,

and possibly provide therapy adjustment for the patient. These devices exchange data with a system designed for medical professional use, or with the patient. Typically, biomedical telemetry is administered, or prescribed for use by a healthcare professional.

Originally the term wireless telemetry was defined as the transmission and measurement of data from remote sources by radio. This definition has expanded to include telecommand that involves the control/reprogramming of devices, and additionally, has expanded beyond radio to include infrared and ultrasonic communications.

Current wireless biomedical telemetry systems can be very sophisticated, are primarily radio-based, and can be divided into the following three general classes of devices:

- Healthcare facility-based wireless (i.e., EKG and pulse oximetry transmitters that allow a patient's mobility about most of one hospital floor while being continuously monitored).
- Implanted and body-worn telemetry systems that are part of portable therapeutic and monitoring systems can be controlled and read by both healthcare professionals with physician-programming devices and by patients with handheld devices. These can be used both in health care facilities and in home use environments.
- Personal health assistant devices. These can be blood glucose monitors, pulse rate, blood pressure monitors, or weight scales. These can be nonprescribed consumer electronics. Use of these types of devices in healthcare management systems is becoming more prevalent.

Wired biomedical telemetry systems, typically hospital-based and connected to the hospital computer network, are quite common. While not covered here, it is worth noting that these systems can automatically measure blood pressure, heart rate, arterial oxygen saturation levels, respiration rate, and allow manual data input for managing hospitalized patients. Wireless systems allow patient mobility to the bathroom, exercise, and allow for general patient case of mobility and comfort. They can also provide increased patient safety due to the continuous monitoring aspect of wireless telemetry.

Wireless biomedical telemetry systems are regulated by both the United States of America Federal Communications Commission (FCC) and the FDA. The FCC defines the RFs allowed for use in transmitting biomedical data, and the power levels allowed. The FDA defines electromagnetic interference (EMI) immunity and safety requirements of therapeutic devices and the associated telemetry interaction at the system level. The frequencies and power levels are important parameters when designing a biomedical telemetry system. The number of users and their power levels, along with man-made noise, can cause significant interference to certain frequency bands. This is one reason why the FCC has created protected frequency bands specifically for biomedical telemetry. These include the wireless medical telemetry system (WMTS) for in-hospital use and the medical implant communications service (MICS) for implantable use. Nonlife critical telemetry often uses the nonprotected industrial, scientific and medical (ISM) bands. The medical portion of the ISM band is related to radio diathermy and other noncommunications use of RF. There are many industrial, commercial, scientific, and home uses of the ISM band, which can lead to significant noise and interference levels on the bands (i.e., cordless phones, WiFi, Bluetooth).

Hospital-based wireless EKG telemetry systems are ubiquitous. Currently manufactured systems utilize the WMTS frequency band. The following paragraph is an excerpt from the FCC WMTS website that describes the creation of this frequency band quite well:

> Prior to the establishment of the WMTS, medical telemetry devices generally could be operated on an unlicensed basis on vacant television channels 7–13 (174–216 MHz) and 14–46 (470–668 MHz) or on a licensed but secondary basis to private land mobile radio operations in the 450- to 470-MHz frequency band. This meant that wireless medical telemetry operations had to accept interference from the primary users of these frequency bands, i.e., the television broadcasters and private land mobile radio licensees. Further, if a wireless medical telemetry operation caused interference to television or private land mobile radio transmissions, the user of the wireless medical telemetry equipment would be responsible for rectifying the problem, even if that meant shutting down the medical telemetry operation.

The FCC was concerned that certain regulatory developments, including the advent of digital television (DTV) service, would result in more intensive use of these frequencies by the primary services, subjecting wireless medical telemetry operations to greater interference than before and perhaps precluding such operations entirely in many instances. To ensure that wireless medical telemetry devices can operate free of harmful interference, the FCC

decided to establish the WMTS, in a Report and Order released on June 12, 2000 (FCC 2006).

The impetus for change occurred in mid-2000, following an incident at Baylor University Medical Center where a HDTV broadcast interrupted medical telemetry in part of the hospital for a short time (Baker 2002). This incident led to WMTS being created by the FCC, and older systems being phased out via attrition. Frequency assignment to channels in the WMTS band are still handled by a frequency coordinator to ensure that multiple uses of the frequency band within the hospital do not interfere with each other.

On the implantable side of wireless biomedical telemetry óne of the earliest uses of implantable telemetry involved the control of the pacing rate of implantable cardiac pacemakers. Originally, the control of the first pacemakers involved piercing the skin of the patient with a custom-sharpened screwdriver to adjust a small potentiometer encased in a waterproof silicone rubber seal. Much to the patients' delight, a simple inductively coupled 160–190-kHz two-way telemetry system replaced the screwdriver and allowed painless adjustment of the pacemaker's pacing rate. This telemetry system, which operates over a short range of just a few inches, was originally used by Medtronic Inc., beginning in the 1960s. Many medical implant manufacturers are still using these short-range inductively coupled systems.

Implantable telemetry systems currently in use today are evolving into longer-range telemetry, which includes the use of the MICS band from 402 to 405 MHz (nearly worldwide allocation), the 902–928-MHz ISM bands in the United States, and the 433 and 868 MHz bands in Europe.

The MICS band is an ultra-low-power band designed to be shared between weather balloons (Meteorological Aids) and medical implantable devices. This band provides licensing by rule, and protection against interference from nonlicensed interferers. The ISM bands allow the use of higher power levels, but must share the use of the frequency bands with other devices operating with much more efficient antennas and much higher power than implanted battery-operated systems. Due to body losses and limited battery power, implanted transmitters are typically the weakest link in the telemetry link and require a quiet, protected frequency band to operate consistently and effectively. The FCC created MICS to provide this.

For external to external and body-worn to external medical telemetry communications that are intended to operate outside healthcare facilities, the options are more limited. Primarily FCC Part 15 and the ISM bands provide the frequency solutions. Using these frequency bands is complicated by the different international radio regulations that have different frequencies and power levels between the United States and Europe. This makes a common worldwide solution very difficult. Different products often need to be designed for different countries.

The future of wireless biomedical telemetry will require systems that operate on common worldwide frequency bands to accommodate travelers. Better use of frequency spectrum and avoidance of interferers will also be necessary as wireless devices proliferate. Clear channel access protocol, and listen-before-talk protocols along with cognitive radio concepts and automatic frequency agility can be utilized to mitigate interference and channel-loading concerns.

3.14.2.5 Enabling Power Technologies for Medical Applications

Medical devices, particularly implanted devices, have highly specialized power source requirements. These include small size, appropriate shape factor, long service life (longevity), and predictable end-of-service characteristics. Small size and shape factor are directly related to the biocompatibility and comfort of the device. A long service life enables reduction of the negative consequences of device replacement, namely the necessity of a surgical procedure, and the associated risk and expense. A predictable end-of-service characteristic allows for some type of warning to be issued by the device (such as by RF telemetry), when the power source is approaching depletion.

Figure 43 shows a comparison of power and current needs for several families of implanted devices. The needs vary by several orders of magnitude, depending on the application.

For any given power drain, the characteristics of high longevity and small size of the power source are directly opposed to each other. This is because both the amount of stored energy of a battery and its rate of release (i.e., power) scale with its size. Therefore, any implanted power source represents a balance between the two. It is always desirable to employ a power source design that maximizes energy density, to help reduce the impact of this compromise.

Until recently, and with a few exceptions, the industry standard power sources for implanted devices over the past three decades have been primary (i.e., single use) batteries using lithium (Li) metal anodes. Different versions of these Li-based cells have been widely

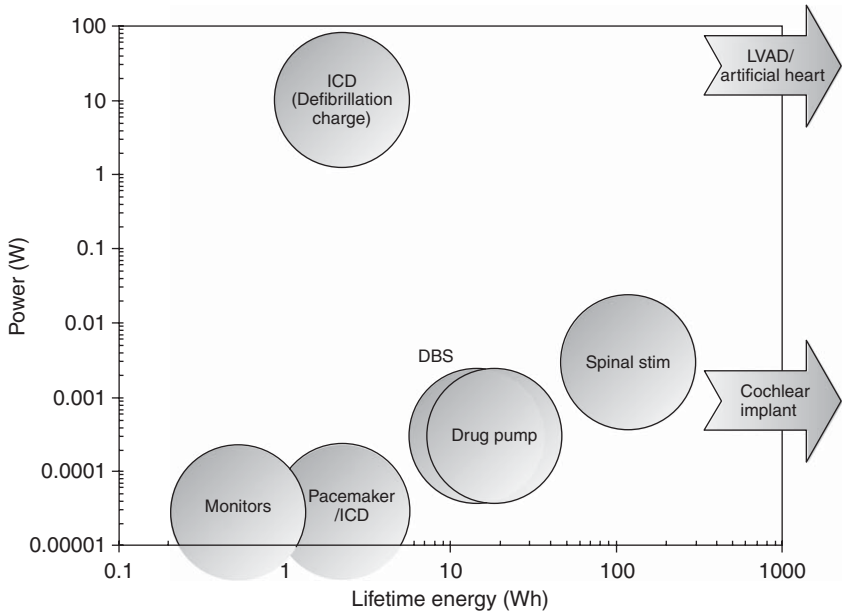


Figure 43 Power and energy map for several implantable device classes: monitors (e.g., heart rate), pacemakers and implantable cardioverter/defibrillators, deep brain stimulators (DBS), drug pumps, spinal stimulators, cochlear implants, left-ventricular-assist devices (LVAD), and artificial hearts. Values are approximate. The artificial heart/LVAD and cochlear implants are shown as arrows to the right, indicating that at present these devices are powered essentially in a continuous fashion by transmitted energy from an external source. (Courtesy: Medtronic.)

deployed (e.g., lithium/iodine, lithium/thionyl chloride, lithium/silver vanadium oxide) (Greatbatch and Holmes 1992, Takeuchi *et al.* 1998). Li is the preferred anode material due to its strong electropositive character (leading to high-voltage cells), high capacity density, and its stability in a variety of electrolyte systems. With a state-of-the-art cathode (a hybrid combination of silver vanadium oxide and carbon monofluoride), and in medium-rate configurations, lithium primary batteries can achieve a packaged energy density of $1\,\mathrm{W\,h\,cm^{-3}}$ (Chen *et al.* 2005), which is as high as any battery currently available.

Since about the year 2000, another type of lithium-based battery, the lithium-ion rechargeable cell, has been used for implanted applications (MacLean *et al.* 1995, Okamoto *et al.* 2002). Initial use was in the field of artificial hearts and LVAD, but more recently an even broader application has emerged in the form of neurostimulators (e.g., the Restore™ device by Medtronic Inc., the Precision™ by Advanced Bionics, and the EON™ by Advanced Neuromodulation Systems). Lithium-ion batteries offer a variety of advantageous characteristics, including high lifetime energy density, relatively high power capability, absence of memory effect, and a low rate of wear-out (also known as capacity fade) (Okamoto *et al.* 2002).

For a detailed description of batteries and general, as well as implanted power sources, along with historical perspective, the reader is referred to Linden and Reddy (2002), Untereker *et al.* (2000), and Scott *et al.* (2004).

The use of microtechnology in battery components includes engineered particulate active materials and conductive additives that range from tens of micrometers to nanometers, as well as microporous polymer separators that permit ion flow while preventing electrical contact between the anode and the cathode.

In terms of high miniaturization at the full battery level, perhaps the broadest use of batteries in very small devices is that of the zinc–air hearing batteries, which are button cells in the size range of a few millimeters in diameter and thickness (e.g., Rayovac size LZ10A). Recent rechargeable batteries that have potential applications in small medical devices include the pin-type BION battery from Quallion (Sylmar, CA, USA) and thin-film-type lithium-based batteries being pursued by a number of start-up companies (Bates *et al.* 2000, Schmidt 2001).

Power sources other than electrochemical batteries have also been considered, but not commercialized. These include thermoelectric generators (MacDonald 2002, Mano *et al.* 2002), β-electric generators, and

biofuel cells. If successful, all are likely to use some aspects of microtechnology in their design. However, at this point it is too early to speculate on the likelihood of commercial adaptation of these technologies.

3.14.2.6 Biocompatible Packaging Technologies for Medical Applications

The electrical systems of implantable devices are not generally biostable or biocompatible. To protect both patient and device, coatings or hermetic packaging are employed. The most common case material for hermetically sealed devices such as pacemakers and implantable defibrillators is titanium. It is biocompatible, formable, weldable, and has a good strength-to-weight ratio. These titanium cases are typically designed with feedthroughs for transmission of electrical signals for sensing and therapy delivery. These feedthroughs are similar to commercial versions except that the materials are selected specifically for biocompatibility. The ferrules are usually titanium or a titanium alloy. The feedthrough wires may be niobium or other suitable material and the insulator is typically glass ceramic or sapphire (**Figure 44**).

Implantable devices may also use glass to provide hermetic protection and allow easy RF communication or power transfer. Glass packages could use wafer bonding techniques from the semiconductor or MEMS industries to form a cover to a package base and enclose the electronics. Other embodiments include glass tubes with sealed ends. Laser joining

Figure 44 Medtronic KAPPA® KUDD901 pacemaker. (Courtesy: Medtronic.)

Figure 45 Trovan ID100A human implantable transponders. (Courtesy: Trovan®)

and other methods have been developed to allow glass–glass and glass–silicon sealing with minimal heating of nearby sensitive components. If electrical connections need to be made through the glass package, standard MEMS techniques can be used to create metal contacts through vias (**Figure 45**).

Acute sensors or therapy devices may not require complete hermeticity. Polymers such as liquid crystal polymer (LCP) with very low gas and water permeability are employed in near-hermetic packages in commercial products and could provide short- to medium-term protection for medical devices (SBIR 2006). This could allow greater shape flexibility including unique features for securing the devices in the body. Research is continuing on methods to improve the performance of polymers in the body by surface modification and coatings (**Figure 46**).

Future developments in materials and biosciences are likely to advance the function of medical device packages from being simply biotolerable to

Figure 46 Bravo™ pH monitoring system. (Courtesy: Medtronic.)

bioneutral – where the device is practically invisible to the body, and eventually make the packages biofunctional with the ability to respond to changes in chemistry, mechanical loading, and other stimuli.

References

About Deafness 2006 http://deafness.about.com/cs/cochlearimplants/a/cifacts.htm

About Vision 2006 http://vision.about.com/od/retinitispigmentosa/a/artificialvisio.htm

Advanced Bionics 2006 http://www.cochlearimplant.com/

All About Vision 2006 http://www.allaboutvision.com/resources/anatomy.htm

American Academy of Ophthalmology 2006 http://www.aao.org

American Academy of Otolaryngology Head and Neck Surgery 2006 http://www.entnet.org/healthinfo/ears/cochlear-implant.cfm

American Foundation for the Blind 2006 http://www.afb.org/

American Optometrists Organization 2006 http://www.aoa.org/

Anderson V C, Burchiel K J 1999 A prospective study of long-term intrathecal morphine in the management of chronic nonmalignant pain. *Neurosurgery* **44(2)**, 289–300

Arena A, Boulougoura M, Chowdrey H S, Dario P, Harendt C, Irion K M, Kodogiannis V, Lenaerts B, Menciassi A, Puers R, Scherjon C, Turgis D 2005 Intracorporeal videoprobe (IVP). In: Bos L (ed.) *Medical and Care Compunetics*, 2nd edn. IOS Press, London, p. 167

Arndt A 2003 Correction for sensor creep in the evaluation of long-term plantar pressure data. *J. Biomech.* **36**, 1813–17

Auditory and Vestibular Pathways 2006 http://thalamus.wustl.edu/course/audvest.html

Bai Qing, Wise K D, Anderson D J 2000 A high-yield microassembly structure for three-dimensional microelectrode arrays. *IEEE Transactions on Biomedical Engineering* **47(3)**, 281–9

Baker S 2002 Building reliability into your digital telemetry system. *Health Management Technology*, www.healthmgttech.com

Bartic C, De Keersmaker K, Parton E 2003 *Neuronen auf dem Chip. Elektronik* **2003(26)**, 22–4

Bates J B, Dudney N J, Neudecker B, Ueda A, Evans C D 2000 Thin-film lithium and lithium-ion batteries. *Solid State Ionics* **135**, 33–45

Bauerdick S, Burkhardt C, Kern D P, Nisch W 2003 Substrate-integrated microelectrodes with improved charge transfer capacity by 3-dimensional micro-fabrication. *Biomedical Microdevices* **5(2)**, 93–9

Bech G J, De Bruyne B, Pijls N H, de Muinck E D, Hoorntje J C, Escaned J, Stella P R, Boersma E, Bartunek J, Koolen J J, Wijns W 2001 Fractional flow reserve to determine the appropriateness of angioplasty in moderate coronary stenosis a randomized trial. *Circulation* **103**, 2928–34

Bell T E, Wise K D, Anderson D J 1997 Flexible micromachined electrode array for a cochlear prosthesis. *Int. Conf. on Solid-State Sensors and Actuators*, Proceedings IEEE, Piscataway, NJ, USA, pp. 1315–18

Benabid A L, Bennazzouz A, Hoffmann D, Limousin P, Krack P, Pollak P 1998 Long-term electrical inhibition of deep brain targets in movement disorders. *Mov. Disord.* **13(3)**, 119–25

Berdondini L, Van Der Wal P D, De Rooij N F, Koudelka-Hep M 2004 Development of an electroless post-processing technique for depositing gold as electrode material on CMOS devices. *Sensors and Actuators, B: Chemical* **99(2–3)**, 505–10

Bergmann G, Graichen F, Siraky J, Jendrzynski H, Rohlmann A 1988 Multichannel strain gauge telemetry for orthopaedic implants. *J. Biomech.* **21**, 169–76

Berkeley Sensors and Actuators Center 2006 http://www.bsac.eecs.berkeley.edu/

Birch M 2004 Engineering biomaterial surfaces to dictate the biological response. *Healthy Aims Dissemination Day 1*. http://www.healthyaims.org

Boston Business Journal 2004 Boston scientific to buy advanced bionics for $740M, http://www.bizjournals.com/boston/stories/2004/05/31/daily12.html

Bouchard C, Tremblay A, Despres J P, Nadeau A, Lupien P J, Theriault G, Dussault J, Moorjani S, Pinault S, Fournier G 1990 The response to long-term overfeeding in identical twins. *N. Engl. J. Med.* **322(21)**, 1477–82

Bourge R C, Aranda J M, Aaron M F, Abraham W T, Magalski A, Smith A L, Roettiger A D, Sparks B. Tallaj J 2005 Clinical outcome of a heart failure management strategy based on continuous hemodynamic monitoring. *Circulation* **112**, II-639

Braunwald E, Zipes D P, Libby P 2001 *Heart Disease: A Textbook of Cardiovascular Medicine*, 6th edn. W. B. Saunders Company, Philadelphia, PA

Braunwald E, Antman E M, Beasley J W, Califf R M, Cheitlin M D, Hochman J S, Jones R H, Kereiakes D, Kupersmith J, Levin T N, Pepine C J, Schaeffer J W, Smith E E III, Steward D E, Theroux P, Gibbons R J, Alpert J S, Faxon D P, Fuster V, Gregoratos G, Hiratzka L F, Jacobs A K, Smith S C Jr. 2002 ACC/AHA 2002 guideline update for the management of patients with unstable angina and non-ST-segment elevation myocardial infarction: A report of the American College of Cardiology/American Heart Association task force on practice guidelines. American College of Cardiology; American Heart Association, Committee on the Management of Patients With Unstable Angina. *J. Am. Coll. Cardiol.* **401**, 366–74

Buchwald H, Avidor Y, Braunwald E, Jensen M D, Pories W, Fahrbach K, Schoelles K 2004 Bariatric surgery: A systematic review and meta-analysis. *JAMA* **292(14)**, 1724–37

Burny F, Donkerwolcke M, Moulart F, Bourgois R, Puers R, Van Schuylenbergh K, Barbosa, Paiva O, Rodes F, Beguertet J, Lawes P 2000 Concept, design and fabrication of smart orthopedic implants. *Med. Eng. Phys.* **22**, 469–79

Canham L 2003 *Biomedical MEMS: Clinical Applications of Silicon Technology*. QinetiQ and IoP Publishing

Canon 2006 http://www.usa.canon.com/html/industrial_medeq/ophthalmic_index.html

Cazeau S, Leclercq C, Lavergne T, Walker S, Varma C, Linde C, Garrigue S, Kappenberger L, Haywood G A, Santini M, Bailleul C, Daubert J-C 2001 Effects of multisite biventricular pacing in patients with heart failure and intraventricular conduction delay. *N. Engl. J. Med.* **344**, 873–80

Chen K, Merritt D R, Howard W G, Schmidt C L, Skarstad P M 2005 Hybrid cathode lithium batteries for implantable medical applications. *J. Power Sources*, in press, available online 22 August 2005

Chow A Y, Pardue M T, Perlman J I, Ball S L, Chow V Y, Hetling J R, Peyman G A, Liang C, Stubbs E B Jr, Peachey N S 2002 Subretinal implantation of semiconductor-based photodiodes: durability of novel implant designs. *J. Rehabil. Res. Dev.* **39(3)**, 313–21

Cleveland Clinic Spine Institute 2006 http://cms.clevelandclinic.org/spine/

Cochlear 2006 http://www.cochlearamericas.com/index.asp

Cristofolini L, Marchetti A, Cappello A, Viceconti M 2000 A novel transducer for the measurement of cement–prosthesis

interface forces in cemented orthopaedic devices. *Med. Eng. Phys.* **22**, 493–501
D'Lima D, Townsend C, Arms S, Morris B, Colwell C 2005 An implantable telemetry device to measure intra-articular tibial forces. *J. Biomech.* **38**, 299–304
Davy D T, Kotzar G M, Brown R H, Heiple K G, Goldberg V M, Heiple K G Jr., Berilla J, Burstein A H 1988 Telemetric force measurements across the hip after total arthroplasty. *J. Bone Joint Surg.* **70**, 45–50
EBI Medical 2006 http://www.ebimedical.com/products/detail.cfm?p=0D
Elliott M R, Thrush A J 1996 Measurement of resolution in intravascular ultrasound images. *Physiol. Meas.* **17**, 259–65
Eye Conditions 2006 http://www.stlukeseye.com/Conditions/ConditionsIndex.asp
Eye Diseases and Conditions 2006 http://www.mdsupport.org/diseases.html
FCC 2006 http://wireless.fcc.gov/services/index.htm
FDA 2006 http://www.fda.gov/opacom/lowlit/hearaid.html
Facts on Hearing Loss 2006 http://www.shhh.org/html/hearing_loss_fact_sheets.html
Fass R (ed.) 2002 *New Frontiers in Gastroesophageal Reflux Disease*. W. B. Saunders Company, London
Flum D R, Salem L, Elrod J A, Dellinger E P, Cheadle A, Chan L 2005 Early mortality among medicare beneficiaries undergoing bariatric surgical procedures. *JAMA* **294(15)**, 1903–8
Fraunhofer 2006 http://www.fraunhofer.de/fhg/Images/mag1-2003-28_tcm6-11223.pdf
Fromherz P, Offenhauser A, Vetter T, Weis J 1991 A neuron-silicon junction: A retzius cell of the Leech on an insulated Gate field effect transistor. *Science* **252(5)**, 1290–2
Fromherz P, Stett A 1995 Silicon neuron junction: capacitive stimulation of an individual neuron on a silicon chip. *Phys. Rev. Lett.* **75(8)**, 1670–3
Gallo J, Kaminek P, Ticha V, Rihakova P, Ditmar R 2002 Particle disease. A comprehensive theory of periprosthetic osteolysis: A review. *Biomed. Pap. Med. Fac. Univ. Palacký Olomouc Czech.* **146(2)**, 21–8
Gastrotarget Corp 1994 *US Pat.* 5 279 607
Gianchandani Y B, Najafi K 1992 A bulk silicon dissolved wafer process for microelectromechanical devices. *J. Microelectromech. Syst.* **1**, 77–85
Gingerich M D, Hetke J F, Anderson D J, Wise K D 2001 A 256-site 3D CMOS microelectrode array for multipoint stimulation and recording in the central nervous system. TRANSDUCERS '01. EUROSENSORS XV. *11th International Conference on Solid-State Sensors and Actuators. Digest of Technical Papers* **1**, 416–19
Gosselin B, Simard V, Sawan M 2004 Low-power implant able microsystem intended to multichannel cortical recording. Proceedings – IEEE International Symposium on Circuits and Systems Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ, USA, p.5
GP Notebook 2006 http://www.gpnotebook.co.uk/simplepage.cfm?ID=2073690151
Greatbatch W, Holmes C F 1992 The lithium–iodine battery: A historical perspective. *PACE* **15**, 2034–6
Griss P, Enoksson P, Tolvanen-Laakso H K, Merilainen P, Ollmar S, Stemme G 2001 Micromachined electrodes for biopotential measurements. *Journal of Microelectromechanical Systems* **10(1)**, 10–16
Griss P, Stemme G 2003 Side-opened out-of-plane microneedles for microfluidic transdermal liquid transfer. *Microelectromechanical Systems Journal* **12(3)**, 296–301
Gross G W, Rhoades B K, Azzazy H M, Wu M C 1995 The use of neuronal networks on multielectrode arrays as biosensors. *Biosens. Bioelectron.* **10(6–7)**, 553–67
Grumet A E, Wyatt J L Jr., Rizzo J F III 2000 Multi-electrode stimulation and recording in the isolated retina. *J. Neurosci. Methods* **101(1)**, 31–42
Hammerle H, Kobuch K, Kohler K, Nisch W, Sachs H, Stelzle M 2002 Biostability of micro-photodiode arrays for subretinal implantation. *Biomaterials* **23**, 797–804
Harris L M, Faggioli G L, Fiedler R, Curl G R, Ricotta J J 1991 Ruptured abdominal aortic aneurysms: Factors affecting mortality rates. *J. Vasc. Surg.* **14**, 812–18
Herber S, Eijkel J, Olthuis W, Bergveld P, Van Den Berg A 2004a Study of chemically induced pressure generation of hydrogels under isochoric conditions using a microfabricated device. *J. Chem. Phys.* **121(6)**, 2746–51
Herber S, Olthuis W, Bergveld P, Van Den Berg A 2004b Exploitation of a pH-sensitive hydrogel disk for detection. *Sens. Actuators B Chem.* **103(1–2)**, 284–9
Hoover A 2006 UF engineer develops tiny, easily mass-produced motion sensor. *University of Florida News,* http://news.ufl.edu/2006/02/09/physical-sensor/
Hoversland R 2006 Histology and embryology: Anatomy of the ear, http://alpha.ipfw.edu/histo-embryo/histear.html
Howorth D J 2002 Feasibility study for a micromachined glaucoma drainage device. M.Sc. thesis, Cranfield University
Jeffers J, Browne M, Taylor M 2005 Damage accumulation, fatigue and creep behaviour of vacuum mixed bone cement. *Biomaterials* **26**, 5532–41
Kane W 1988 Direct current electrical bone growth stimulation for spinal fusion. *Spine* **13(3)**, 363–5
Karoub J 2003 Patent promises to place smart scalpels in surgeons hands. *Small Times*, http://www.smalltimes.com/document_display.cfm?document_id=5405
Katholieke Universiteit Leuven 2006 http://www.kuleuven.ac.be/english/
Kazmers A, Perkins A J, Jacobs L A 2001 Aneurysm rupture is independently associated with increased late mortality in those surviving abdominal aortic aneurysm repair. *J. Surg. Res.* **95**, 50–3
Kewley D T, Hills M D, Borkholder D A, Opris I E, Maluf N I, Storment C W, Bower J M, Kovacs G T A 1997 Plasma-etched neural probes. *Sensors and Actuators, A: Physical* **58(1)**, 27–35
Klein S, Burke L E, Bray G A, Blair S, Allison D B, Pi-Sunyer X, Hong Y, Eckel R H 2004 Clinical implications of obesity with specific focus on cardiovascular disease: A statement for professionals from the American Heart Association Council on Nutrition, Physical Activity, and Metabolism: Endorsed by the American College of Cardiology Foundation. *Circulation* **110(18)**, 2952–67
Kovacevic N 1993 *US Pat.* 5 197 488
Kovacs G 1998 *Micromachined Transducers Sourcebook*. WCB/McGraw-Hill
Krahn A D, Klein G J, Yee R, Takle-Newhouse T, Norris C 1999 Use of an extended monitoring strategy in patients with problematic syncope. *Circulation* **99**, 406–10
Kuiken T A, Dumanian G A, Lipschutz R D, Miller L A, Stubblefield K A 2005 Targeted muscle reinnervation for improved myoelectric prosthesis control. *Proc. 2nd Int. IEEE EMBS Conf. Neural Engineering*, Arlington, VA, USA, March 16–19, 2005, pp. 155–7
Langenfeld H, Krein A, Kirstein M, Binner L 1998 Peak endocardial acceleration-based clinical testing of the ''BEST'' DDDR pacemaker. *Eur. PEA Clin. Invest. Group Pacing Clin. Electrophysiol.* **21**, 2187–91
Lapatki B G, van Dijk J P, Jonas I E, Zwarts M J, Stegeman D F 2003 A thin, flexible multielectrode grid for high-density surface EMG. *J. Appl. Physiol.* **96**, 327–36
Leclercq C, Cazeau S, Ritter P, Alonso C, Gras D, Mabo P, Lazarus A, Daubert J C 2000 A pilot experience with

permanent biventricular pacing to treat advanced heart failure. *Am. Heart J.* **140**, 862–70
Lee W A, Carter J W, Upchurch G, Seeger J M, Huber T S 2004 Perioperative outcomes after open and endovascular repair of intact abdominal aortic aneurysms in the United States during 2001. *J. Vasc. Surg.* **39**, 491–6
Lemos P A, Hoye A, Goedhart D, Arampatzis C A, Saia F, van der Giessen W J, McFadden E, Sianos G, Smits P C, Hofma S H, de Feyter P J, van Domburg R T, Serruys P W 2004 Clinical, angiographic, and procedural predictors of angiographic restenosis after sirolimus-eluting stent implantation in complex patients. An evaluation from the rapamycin-eluting stent evaluated at Rotterdam Cardiology Hospital (RESEARCH) study. *Circulation***109**,1366–70
Levine J A, Eberhardt N L, Jensen M D 1999 Role of nonexercise activity thermogenesis in resistance to fat gain in humans. *Science* **283(5399)**, 212–14
Linden D, Reddy T B (eds.) 2002 *Handbook of Batteries*, 3rd edn. McGraw-Hill, New York
Loizou P 1998 Mimicking the human ear. *IEEE Signal Process. Mag.* **15(5)**, 101–30
Long J W 2001 Advanced mechanical circulatory support with the HeartMate left ventricular assist device in the year. *Ann. Thorac. Surg.* **71**, S176–82
MacDonald S G 2002 *US Pat.* 6 640 137
MacLean G K, Aiken P A, Adams W A, Mussivand T 1995 Preliminary evaluation of rechargeable lithium-ion cells for an implantable battery pack. *J. Power Sources* **56**, 69–74
Madou M 2005 A 3D carbon world. In: Reichl H. (ed.) *Micro System Technologies.*, Franzis Verlag GmbH, Poing (DE), p. 21
Mano N, Mao F, Heller A 2002 A miniature biofuel cell operating in a physiological buffer. *J. Am. Chem. Soc.* **124**, 12962–3
Med-El 2006 http://www.medel.com/
Merriam 1993 *Merriam Webster's Collegiate Dictionary*, 10th edn. Merriam-Webster, Inc., Springfield, MA
Mesotec 2006 http://www.mesotec.com/english/start.htm
Mirowski M, Reid P R, Mower M M, Watkins L, Gott V L, Schauble J F, Langer A, Heilman M S, Kolenik S A, Fischell R E, Weisfeldt M L 1980 Termination of malignant ventricular arrhythmias with an implanted automatic defibrillator in human beings. *N. Engl. J. Med.* **303**, 322–4
Mokwa W 2004 MEMS technologies for epiretinal stimulation of the retina. *J. Micromechanics and Microengineering* **14(9)**, 12–16
Morris B, D'Lima D, Slamin J, Kovacevic N, Arms S, Townsend C, Colwell C Jr. 2001 e-Knee: Evolution of the electronic knee prosthesis: Telemetry technology development. *J. Bone Joint Surg.* **83**, 62–6
Morrison L K, Harrison A, Krishnaswamy P, Kazanegra R, Clopton P, Maisel A 2002 Utility of a rapid B-natriuretic peptide assay in differentiating congestive heart failure from lung disease in patients presenting with dyspnea. *J. Am. Coll. Cardiol.* **39**, 202–9
Muller J E, Abela G S, Nesto R W, Tofler G H 1994 Triggers, acute risk factors and vulnerable plaques: The lexicon of a new frontier. *J. Am. Coll. Cardiol.* **23**, 809–13
Murakami K 2005 A miniature confocal optical scanning microscope for endoscope. *Progress in Biomedical Optics and Imaging – Proceedings of SPIE*. International Society for Optical Engineering, Bellingham, WA, pp. 119–31
Naghavi M, Libby P, Falk E, Casscells W, Litovsky S, Rumberger J, Badimon J J, Stefanadis C, Moreno P, Pasterkamp G, Fayad Z, Stone P H, Waxman S, Raggi P, Madjid M, Zarrabi A, Burke A, Yuan C, Fitzgerald P J, Siscovick D S, de Korte C L, Aikawa M, Airaksinen K E J, Assmann G, Becker C R, Chesebro J H, Farb A, Galis Z S, Jackson C, Jang I-K, Koenig W, Lodder R A, March K, Demirovic J, Navab M, Priori S G, Rekhter M D, Bahr R, Grundy S M, Mehran R, Colombo A, Boerwinkle E, Ballantyne C, Insull W Jr., Schwartz R S, Vogel R, Serruys P W, Hansson G K, Faxon D P, Kaul S, Drexler H, Greenland P, Muller J E, Virmani R, Ridker P M, Zipes D P, Shah P K, Willerson J T 2003 From vulnerable plaque to vulnerable patient. *Circulation* **109**, 1664–72
Najafi K, Wise K D, Mochizuki T 1985 High-yield IC-compatible multichannel recording array. *IEEE Transactions on Electron Devices* **ED-32(7)**, 1206–11
National Eye Institute 2006 http://www.nei.nih.gov/
National Institute on Deafness and Other Communication Disorders 2006 http://www.nidcd.nih.gov/health/hearing/hearingaid.asp
National Institutes of Health 2006 http://www.nih.gov/
Neidlinger T, Harendt C, Gloeckner J, Schubert M B 2000 Novel device concept for voltage-bias controlled color detection in amorphous silicon sensitized CMOS cameras. *Materials Research Society Symposium – Proceedings*, San Francisco, CA, USA, pp. 285–90
Noonan K, Farnum C, Leiferman E, Lampl M, Markel M, Wilsman H 2004 Growing pains: Are they due to increased growth during recumbency as documented in a lamb model? *J. Pediatr. Orthop.* **24(6)**, 726–31
Nordhausen C T, Rousche P, Normann R A 1993 Chronic recordings of visually evoked responses using the Utah intracortical electrode array. *Proc. Ann. Conf. on Engineering in Medicine and Biology,* Publ. by IEEE, Piscataway, NJ, pp. 1391–2
O'Donnell P D 1997 *Urinary Incontinence*. Mosby, London
Okamoto E, Watanabe K, Hashiba K, Inoue T, Iwazawa E, Momoi M, Hashimoto T, Mitamura Y 2002 Optimum selection of an implantable secondary battery for an artificial heart by examination of the cycle life test. *ASAIO J.* 495–502
Optobionics 2006 http://www.optobionics.com/
Pan Y, Xie H, Fedder G K 2001 Endoscopic optical coherence tomography based on a microelectromechanical mirror. *Opt. Lett.* **26(24)**, 1966–8
Panescu D 2005 An imaging pill for gastrointestinal endoscopy. *IEEE Eng. Med. Biol. Mag.* **24(4)**, 12–14
Parodi J C, Palmaz J C, Barone H D 1991 Transfemoral intraluminal graft implantation for abdominal aortic aneurysms. *Ann. Vasc. Surg.* **5**, 491–9
Peeters E, Puers B, Sansen W, Gybels J, De Sutter P 1991 Two-wire, digital output multichannel microprobe for recording single-unit neural activity. *Sensors and Actuators, B: Chemical* **B4(1–2)**, 217–23
Penner A, Doron E 2003 *US Pat.* 6 628 989
Porciani M C, Dondina C, Macioce R, Demarchi G, Pieragnoli P, Musilli N, Colella A, Ricciardi G, Michelucci A, Padeletti L 2005 Echocardiographic examination of atrioventricular and interventricular delay optimization in cardiac resynchronization therapy. *Am. J. Cardiol.* **95**, 1108–10
Prohaska O J, Olcaytug F, Pfunder P, Dragaun H 1986 Thin-film multiple electrode probes. possibilities and limitations. *IEEE Trans. Biomed. Eng.* **BME-33(2)**, 223–9
Puers R 2005 Implantable sensor systems. *Symposium of the Delft University Research Center of Intelligent Sensor Microsystems, DISens*, 3 edn. Delft University of Technology, Delft, pp. 66–79
Puers R, Catrysse M, Vandevoorde G, Collier R, Louridas E, Burny F, Donkerwolcke M, Moulart F 1999 A telemetry system for the detection of hip prosthesis loosening by vibration analysis. *Sens. Actuators, A: Physical* **85(1)**, 42–7
Puers B, Sansen W, Vereecken R, H P Kimmich (ed.), Klewe H J (ed.) 1984 *Development Considerations of a Micropower Control Chip and Ultraminiature Hybrid for Bladder Pressure Telemetry*, 8th edn. pp. 328–32
Quirk W, Clair D 2005 The Orthopedic Industry: Trends and Opportunities. RBC Capital Markets

Receveur R A M, Marxer C R, Woering R, Larik V C M H, deRooij N F 2005 Laterally moving bistable MEMS DC switch for biomedical applications. *J. Microelectromechanical Systems* **14(5)**, 1089–98

Reichert Tonopen 2006 http://www.reicherttonopen.com/xl.html

Rizzo J F, Wyatt J III, Loewenstein J, Kelly S, Shire D 2003 Methods and perceptual thresholds for short-term electrical stimulation of human retina with microelectrode arrays. *Invest. Ophthalmol. Vis. Sci.* **44(12)**, 5355–61

Roberts C, Voor M, Rose M, Glassman S 1998 Spinal ligament loading during axial distraction: A biomechanical model. *Am. J. Orthop.* **27(6)**, 434–40

Roberts D 2006 Microchip Implantation, http://www.mdsupport.org/library/chip.html

Robinson P 1997 Radiology's Achilles' heel: Error and variation in the interpretation of the Rontgen image. *Br. J. Radiol.* **70**, 1085–98

Rodriguez F J, Ceballos D, Schuttler M, Valero A, Valderrama E, Stieglitz T, Navarro X 2000 Polyimide cuff electrodes for peripheral nerve stimulation. *J. Neurosci. Methods* **98(2)**, 105–18

Rose E A, Gelijns A C, Moskowitz A J, Heitjan D F, Stevenson L W, Dembitsky W, Long J W, Ascheim D D, Tierney A R, Levitan R G, Watson J T, Meier P 2001 Long-term mechanical left ventricular assistance for end-stage heart failure. *N. Engl. J. Med.* **345**, 1435–43

Ross R 1999 Atherosclerosis: An inflammatory disease. *N. Engl. J. Med.* **340**, 115–26

Rousche P J, Normann R A 1999 Chronic intracortical microstimulation (ICMS) of cat sensory cortex using the Utah Intracortical Electrode Array. *IEEE Transactions on Rehabilitation Engineering* **7(1)**, 56–68

Roxhed N, Griss P, Stemme G 2005a Generic leak-free drug storage and delivery for microneedle-based systems. *Proceedings of the IEEE International Conference on Micro Electro Mechanical Systems (MEMS),* Publ. by IEEE, Piscataway, NJ, pp. 742–5

Roxhed N, Griss P, Stemme G 2005b Reliable in-vivo penetration and transdermal injection using ultra-sharp hollow microneedles. *Int. Conf. on Solid-State Sensors, Actuators and Microsystems*, Publ. by IEEE, Piscataway, NJ, pp. 213–16

Rutten W L C, Frieswijk T A, Smit J P A, Rozijn T H, Meier J H 1995 3D neuro-electronic interface devices for neuromuscular control: design studies and realisation steps. *Biosensors & Bioelectronics* **10(1–2 pt 2)**, 141–53

Sachs H G, Schanze T, Brunner U, Sailer H, Wiesenack C 2005 Transscleral implantation and neurophysiological testing of subretinal polyimide film electrodes in the domestic pig in visual prosthesis development. *Journal of Neural Engineering* **1**, S57–S64

Sadat A, Qu H, Yu C, Yuan J, Xie H 2005 Low-power CMOS wireless MEMS motion sensor for physiological activity monitoring. *IEEE Trans. Circuits Syst.* **52(12)**, 2539–51

Sansen W, Puers B, Vereecken R 1984 Realization of a pressure telemetry capsule for cystometry. *IEEE/Engineering in Medicine and Biology Society Annual Conference*. IEEE, New York, NY, USA, pp. 711–14

Sansen W, Vereecken R, Puers B, Folens G, Van Nuland T 1987 Closed loop system to control the bladder function. *IEEE/Engineering in Medicine and Biology Society Annual Conference*. IEEE, New York, NY, USA, pp. 1149–50

Sawan M, Duval F, Jin-Sheng L, Hassouna M, Elhilali M M 1993 A new bladder stimulator-hand-held controller and miniaturized implant: Preliminary results in dogs. *Biomed. Instrum. Technol.* **27(2)**, 143–9

SBIR 2006 http://sbir.gsfc.nasa.gov/SBIR/successes/ss/3-019text.html

Schepel S J, Koning G, Oeseburg B, Langbroek A J M, Zijlstra W G 1987 Performance of a pH monitoring system *in vivo*. *Med. Biol. Eng. Comput.* **25(1)**, 63–7

Schiele F, Meneveau N, Vuillemenot A, Zhang D D, Gupta S, Mercier M, Danchin N, Bertrand B, Bassand J P 1998 Impact of intravascular ultrasound guidance in stent deployment on 6-month restenosis rate: A multicenter, randomized study comparing two strategies – with and without intravascular ultrasound guidance. *J. Am. Coll. Cardiol.* **32**, 320–8

Schmidt E M, Bak M J, Hambrecht F T, Kufta C V, O'Rourke D K, Vallabhanath P 1996 Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex. *Brain* **119(Pt 2)**, 507–522

Schmidt C 2001 *US Pat.* 6 782 290

Schneider A, Stieglitz T 2004 Implantable flexible electrodes for functional electrical stimulation. *Medical Device Technology* 16–18

Schutz R, Dorschel K, Muller G J 1997 Confocal microscanner technique for endoscopic vision. *Proc. SPIE – The International Society for Optical Engineering*, San Remo, Italy, pp. 223–33

Schwarz M, Ewe L, Hauschild R, Hosticka B J, Huppertz J, Kolnsberg S, Mokwa W, Trieu H K 2000 Single chip CMOS imagers and flexible microelectronic stimulators for a retina implant system. *Sensors and Actuators A: Physical* **83(1–3)**, 40–6

Scott E, Schmidt C, Skarstad P M 2004 Power sources for implanted medical devices. In: Wnek G, Bowlin G (eds.) *Encyclopedia of Biomaterials and Biomedical Engineering. Dekker Encyclopedias*. Taylor & Francis Group, LLC, New York, NY

Second Sight 2006 http://www.2-sight.com

Slater D K, Hlatky M A, Mark D B, Harrell F E Jr., Pryor D B, Califf R M 1987 Outcome in suspected acute myocardial infarction with normal or minimally abnormal admission electrocardiographic findings. *Am. J. Cardiol.* **60**, 766–70

Smit A, Pollard M, Cleaton-Jones P, Preston A 1997 A comparison of three electrodes for the measurement of pH in small volumes. *Caries Res.* **31(1)**, 55–9

Smit J P A, Rutten W L C, Boom H B K 1999 Endoneural selective stimulating using wire-microelectrode arrays. *IEEE Transactions on Rehabilitation Engineering* **7(4)**, 399–412

Stangel K, Kolnsberg S, Hammerschmidt D, Hosticka B J, Trieu H K, Mokwa W 2001 A programmable intraocular CMOS pressure sensor system implant. *IEEE J. Solid-State Circuits* **36**, 1094–100

Stieglitz T 2001 *Implantable Microsystems for Monitoring and Neural Rehabilitation, Part II. Medical Device Technology* Dec. 2–6

Stieglitz T 2002 *Implantable Microsystems for Monitoring and Neural Rehabilitation, Part I. Medical Device Technology* January, 2–4

Surgeon M 2004 *The Use of the Acoustic Emission Technique for Materials Research and Structural Integrity Monitoring*. Vallen-Systeme GmbH, Germany, www.vallen.de

Szivek J, Roberto R, Slack J, Majeed B 2002 An implantable strain measurement system designed to detect spine fusion. *Spine* **27(5)**, 487–97

Takeuchi E S, Quattrini P J, Greatbatch W 1988 Lithium/silver vanadium oxide batteries for implantable defibrillators. *Pacing Clin. Electrophysiol.* **11**, 2035–9

Technology for Industry 2005 *Micro and Nanotechnology in Healthcare and Life Sciences Market Sector Report*. Technology for Industry Ltd.

The Cleveland Clinic 2006 http://cms.clevelandclinic.org/heartbrain/

The Doheny Eye Institute 2006 http://www.usc.edu/hsc/doheny/

The Foundation Fighting Blindness 2006 http://www.blindness.org/content.asp?id=45

The Hearing Planet 2006 http://www.hearingplanet.com/index.html

The Retinal Implant Project 2006 http://www.deafblind.com/implant.html

Thiebaud P, De Rooij N F, Koudelka-Hep M, Stoppini L 1997 Microelectrode arrays for electrophysiological monitoring of hippocampal organotypic slice cultures. *IEEE Transactions on Biomedical Engineering* **44(11)**, 1159–63

Thom T, Haase N, Raymond W, Howard V J, Rumsfeld J, Manolio T, Zheng Z J, Flegal K, O'Donnell C, Kittner S, Lloyd-Jones D, Goff D C Jr., Hong Y, Members of the Statistics Committee and Stroke Statistics Subcommittee, Adams R, Friday G, Furie K, Gorelick P, Kissela B, Marler J, Meigs J, Roger V, Sidney S, Sorlie P, Steinberger J, Wasserthiel-Smoller S, Wilson M, Wolf P 2006 Heart disease and stroke statistics – 2006 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* **113**, e85–151

Tooley C Dissemination Day 2004 *First Healthy Aims Artificial Sphincter and Sphincter Sensor*, 1st edn, http://www.healthyaims.org/

University of Massachusetts Memorial Medical Center 2006 http://www.umassmemorial.org/ummhc/hospitals/med_center/services/cochlear/cochlear_implant.cfm

Untereker D F, Shepard R B, Schmidt C L, Crespi A M, Skarstad P M 2000 Power systems for implantable pacemakers, cardioverters and defibrillators. In: Ellenbogen K A, Kay G N, Wilkoff B L (eds.) *Clinical Cardiac Pacing and Defibrillation*, 2nd edn. W.B. Saunders Co., Philadelphia, PA, pp. 167–93

Uppsala University 2006 http://info.uu.se/fakta.nsf/sidor/uppsala.university.id5D.html

US Preventive Services Task Force 2005 Screening for abdominal aortic aneurysm: Recommendation statement. *Ann. Intern. Med.* **142**, 198–202

Van Buyten J 2006 Neurostimulation for chronic neuropathic back pain in failed back surgery syndrome. *J. Pain Symptom Manage.* **31(4)**, S25–9

Van Halteren A R 2003 *Urine incontinentie bij volwassenen*. College voor zorgverzekeringen, Amstelveen, HM200023a 03/147

Vogelzang N J, Ruane M, Ratain M J, Dhowlatshahi K, Chodak G W 1987 A programmable and implantable pumping system for systemic chemotherapy: A performance analysis in 52 patients. *J. Clin. Oncol.* **5**, 1968–76

Walsh P C, Retik A B, Vaughan E D Jr., Wein A J, Kavoussi L R, Novick A C, Partin A W, Peters C A 2002 *Campbell's Urology*, 8th edn. W. B. Saunders Company, London

WebMD 2001 Disc Problems. WebMD, www.webmd.com/content/pages/4/1661_50311.htm

Weil E H J 2000 Clinical and experimental aspects of sacral nerve neuromodulation in lower urinary tract dysfunction. Ph.D. thesis, University of Maastricht

Weiland J D, Fujii G Y, Mahadevappa M, Greenberg R J, Tameesh M, Guven D, de Juan E Jr., Humayun M S 2002 Chronic electrical stimulation of the canine retina. *Conference Proceedings. Second Joint EMBS-BMES Conference 2002. 24th Annual International Conference of the Engineering in Medicine and Biology Society*, Annual Fall Meeting of the Biomedical Engineering Society (Cat. No.02CH37392) IEEE, Houston, TX, USA, pp. 2051–2

Weiss E H, Weaver L, Fang M, Tamhidi L, Kang P, Tomassoni G, Gallagher P, Hesselson A, Sharma A, O'Neill G, Ackerman S, Turk K, Rodrigues D 2006 Evaluation of a subcutaneous photoplethysmography sensor in mature ICD pockets. *Heart Rhythm* **3**, S205

Wikipedia 2006 http://en.wikipedia.org/wiki/Image:HumanEar.jpg

Wilkinson M, Hitchings D 2003 *Microsystems in Biomedical Engineering*, 5th edn. FSRM Swiss Foundation for Research in Microtechnology, Switzerland

Wireless Integrated Microsystems 2006 http://www.wimserc.org/

Wong W M, Bautista J, Dekel R, Malagon I B, Tuchinsky I, Green C, Dickman R, Esquivel R, Fass R 2005 Feasibility and tolerability of transnasal/per-oral placement of the wireless pH capsule vs. traditional 24-h oesophageal pH monitoring – A randomized trial. *Aliment. Pharmacol. Ther.* **21(2)**, 155–63

World Health Organization 2006 http://www.who.int/mediacentre/factsheets/fs282/en/

Yamada T, Alpers D H (eds.) 2006 *Textbook of Gastroenterology*, 4th edn. Lippincott Williams & Wilkins, London

Young C G, Austrin M G 1979 *Learning Medical Terminology Step by Step*, 4th edn. The C.V. Mosby Company, St. Louis, MO, USA

Zeijlemaker B Y W, Van Swol C F P, Noordzij J W 2006 http://www.urolog.nl

Zrenner E, Miliczek K D, Gabel V P, Graf H G, Guenther E, Haemmerle H, Hoefflinger B, Kohler K, Nisch W, Schubert M, Stett A, Weiss S 1997 The development of subretinal microphotodiodes for replacement of degenerated photoreceptors. *Ophthalmic Res.* **29(5)**, 269–80

Biographies

Sarah Audet received her Bachelor of Science degree from the State University of New York at Upstate Medical Center in the field of Medical Technology, her Master's of Science Degree in Electrical Engineering from Boston University, and her Ph.D. in Electrical Engineering from the Technical University of Delft, The Netherlands. Sarah has worked in the field of sensor design, development, and manufacture since 1990 at several companies including AT&T, Princeton Gamma-Tech, Motorola, and Medtronic. Sarah is currently the Director of Technology Planning at Medtronic.

In addition to her direct responsibilities, Sarah is involved with mentoring local high school students, mentoring women in science fields, and volunteering to assist with external programs to educate students in the field of science.

James Carney is a Medtronic Technical Fellow and serves as Senior Principal System Engineer working in Medtronic's Cardiac Rhythm Disease Management Division in Minneapolis, MN. At Medtronic, he has led the development of sensors and diagnostic monitors that are placed in the human body to enable the management of chronic diseases or to provide input to therapy-giving devices. His responsibilities also include identifying and assessing new technologies that might be incorporated into implantable medical products. Jim studied Electrical Engineering at the University of Notre Dame and then obtained a Ph.D. in Electrical Engineering from the Massachusetts Institute of Technology. Prior to joining Medtronic, Dr. Carney worked at Honeywell's Sensors and System Development Center where he led research teams in a number of high-technology programs including semiconductor lasers, GaAs ICs, and silicon MEMS.

Greg Haubrich is a Medtronic Technical Fellow and serves as a Senior Principal RF Engineer in Medtronic's Cardiac Rhythm Disease Management Division in Minneapolis, MN. At Medtronic, he has led the design and development of RF Communication Systems and Circuits. Greg has a Bachelor's degree in Electrical Engineering from the University of Minnesota Institute of Technology, and is a licensed Professional Engineer. Prior to joining Medtronic, Greg worked at Honeywell's Military Avionics Division where he designed microwave radar altimeters and at Motorola, Schaumburg, IL, designing cellular mobile and portable cell phone RF circuitry. Greg is also a member of the executive board of the Minnesota Astronomical Society, and regularly volunteers for the Boy Scouts of America as a merit badge councilor.

Erik Herrmann received his BA degree in Physics and Physics-Microcomputing from Ithaca College, Ithaca, NY, in 1990, MBA degree in Technology Management from the University of Phoenix, Phoenix, AZ, in 2001, and MS in Electrical Engineering from Arizona Statue University, Tempe, AZ, in 2003.

Prior to joining Medtronic in 2002, he held positions in product engineering and engineering management with responsibilities for the design and manufacture of electronic products involving analog signal processing, power amplification, embedded digital systems, and RF communications. Since joining Medtronic, he has worked in product development as IC Product Engineer and as a manager in the Technology Group.

Vincent C.M.H. Larik was born in Vaals, The Netherlands, on September 2, 1954 and received his BSEE degree from the HTS, Heerlen, The Netherlands, in 1976. His responsibilities at the Medtronic Bakken Research Center, Maastricht, The Netherlands, include the management of the Advanced Concepts group, and research in the novel technologies and methodologies applicable to invasive medical devices for therapy and diagnosis.

Rogier A.M. Receveur was born in Capelle aan den IJssel, The Netherlands, on February 9, 1969 and holds a Master's degree in Applied Physics from Delft University of Technol-ogy, Delft, The Nether-lands, in 1993. He has worked at Océ Tech-nologies and is now a Senior R&D Engineer at Medtronic Bakken Research Center, Maastricht, The Netherlands. His research interests include microfabrication technology, and sensors for biomedical applications. Ir. Receveur is a member of The Netherlands' Physical Society and the Royal Dutch Society of Engineers.

Erik Scott is Senior Manager of Battery Research and Technology at the Medtronic Energy and Component Center (MECC). He has spent eight years developing battery technology for implantable medical devices, and he has also worked for six years at ALZA Corporation, developing techniques for delivering medicines across skin using iontophoresis. Erik received his Ph.D. in Materials Science and Engineering from the University of Minnesota.

Glen Vaughn holds Bachelor's and Master's Degrees in Electrical Engineering from Arizona State University, along with a Doctorate in Economics and Finance. Glen has held a wide variety of positions in semiconductor and information technology over the last 40 years, first with Motorola and now with Medtronic. In addition to these responsibilities, Glen also serves on the Board of Directors of the Arizona Bioindustry Association and is a member of the Dean's Advisory Council at the Ira A. Fulton School of Engineering at Arizona State University.

3.15 Micro-Combustion

Christopher M. Spadaccini[1] **and Ian A. Waitz**[2], [1]Center for Micro- and Nanotechnology, Lawrence Livermore National Laboratory, Livermore, CA, USA, [2]Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, USA

Nomenclature

Roman

A	Arrhenius pre-exponential factor
a	Arrhenius exponent
a_v	surface area-to-volume ratio
b	Arrhenius exponent
Bi	Biot number for heat transfer
C_b	molar density (mol m^{-3})
C_p	constant pressure specific heat (J kg^{-1} K^{-1})
D	diffusion coefficient (cm^2 s^{-1})
Da_h	homogeneous Damköhler number
Da_2	diffusion-based Damköhler number
d_h	hydraulic diameter (m)

E_a	activation energy
f	Darcy friction factor
fa	fuel–air ratio
h	convective heat transfer coefficient ($W m^{-2} K^{-1}$) or enthalpy
j_D	j-factor for mass transfer
j_H	j-factor for heat transfer
k	thermal conductivity ($W m^{-1} K^{-1}$) or rate constant
k_m	mass transport coefficient ($m s^{-1}$)
L	length (m)
LHV	lower heating value
M	molecular weight
$\dot{m}$	mass flow rate ($kg s^{-1}$)
P	static pressure ($N m^{-2}$)
Pe	Péclet number
Pr	Prandtl number
Q	heat (W)
R	gas constant ($J kg^{-1} K^{-1}$) or reaction rate
Re	Reynolds number
Sc	Schmidt number
S_f	shape factor
T	temperature (K)
V	volume
v	velocity ($m s^{-1}$) or diffusion volume
w	thickness (m)
Y	mole fraction
z	axial location (m)
Greek	
α	porosity
ϕ	equivalence ratio
η	efficiency
μ	viscosity ($N s m^{-2}$)
ρ	density ($kg m^{-3}$)
τ	characteristic time (s)
Subscripts	
a	air
b	bulk flow
chem	chemical
d	diameter (m)
f	fuel
S	surface
therm	thermal
0	initial
1	inlet
2	exit

3.15.1 Introduction

Until the advent of advanced microfabrication techniques and microelectromechanical systems (MEMS), combustion at small size scales was largely an academic topic revolving around the question of quenching diameter. However, as the MEMS field grew, subdisciplines such as power MEMS, micropropulsion, and microreactor technologies were spawned. Devices in each of these areas require a heat source, typically a combustor due to the high energy density of liquid hydrocarbon fuels. Thus, combustion in small volumes has become an important area of research.

Because of various microcombustion applications, several types of combustors have been studied. They can be divided into two major categories: (1) homogeneous gas-phase microcombustors and (2) heterogeneous catalytic microcombustors. The first type is similar to the combustion systems that are found in larger-scale gas turbines, whereas the second type is more characteristic of a chemical reactor with a surface catalyst to promote reaction rates. The type of combustor chosen for a given application will depend on the required fuel, materials constraints, the thermodynamic cycle, and available fabrication techniques; however, the fundamental physics, limitations, and constraints for all of these combustors are the same.

This chapter provides a general understanding of microcombustion fundamentals and challenges, shows some microcombustor operational data, and provides some design guidelines. First, an introduction to basic combustion principles and definitions is presented, followed by a discussion of fundamental issues unique to combustion systems at the microscale. Next, a detailed presentation of homogeneous gas-phase microcombustors is provided, including operational data and guidelines for microcombustor designers. This is followed by a detailed discussion of heterogeneous catalytic microcombustors with data and design suggestions. Finally, this chapter concludes with a discussion of micrcombustor applications ranging from micro heat engines to microscale methanol reformers and some perspectives on microcombustor research needs.

3.15.2 Basic Combustion Concepts

Before discussing issues pertaining specifically to microcombustors, the reader should be familiar with some basic combustion concepts. These include the fundamental chemical equations of combustion reactions, the concept of equivalence ratio (normalized fuel–air ratio), reaction rate and Arrhenius expressions, and various definitions of efficiency.

3.15.2.1 Combustion Chemistry

Combustion consists of a chemical reaction between a fuel (typically hydrogen or a hydrocarbon species) and an oxidizer leading to heat release. In general, the overall chemical equation for complete combustion of hydrocarbons with air can be written as

$$C_xH_y + \left(x + \frac{y}{4}\right)(O_2 + 3.76N_2) \rightarrow xCO_2 + \frac{y}{2}H_2O + 3.76\left(x + \frac{y}{4}\right)N_2 \quad [I]$$

Typical fuels include hydrogen, propane, and gasoline. Choice of fuel often depends on fuel storage constraints, reaction rates, and emissions requirements. The combustion chemistries for hydrogen and propane with air, respectively, are

$$H_2 + \frac{1}{2}O_2 + 1.88N_2 \rightarrow H_2O + 1.88N_2 \quad [II]$$

$$C_3H_8 + 5O_2 + 18.8N_2 \rightarrow 3CO_2 + 4H_2O + 18.8N_2 \quad [III]$$

The only products of complete combustion are water and carbon dioxide (in the ideal case nitrogen is inert); however, in reality this is not the case. Trace quantities of pollutants such as oxides of nitrogen (NO_x) and carbon monoxide (CO), as well as some unburned hydrocarbons, are usually present. Incomplete combustion may occur if the reactants are not well mixed or if there is insufficient residence time (as is often the case with microcombustors). In these scenarios, there will be significant quantities of reactants in the exit gas stream.

Although combustion chemistries can generally be written as single chemical equations, in reality there are many individual subreactions that combine to give the overall reaction. These sets of chemical equations are referred to as the reaction mechanism.

3.15.2.2 Equivalence Ratio

If propane and air, for example, are provided to a combustion system in a 23.8:1 molar ratio, the mixture is considered stoichiometric. In this case, all fuel and air are consumed in the reaction. If there is excess oxidizer, the mixture is lean, whereas with excess fuel, it is rich. This ratio is referred to as the fuel–air ratio and is usually given on a mass basis, which can be obtained by scaling the molar ratio by the molecular weights of the constituents. In a real combustor, the mass-based fuel–air ratio can be found directly from measurements of the fuel and air mass flow rates.

$$fa = \frac{\dot{m}_f}{\dot{m}_a} \quad [1]$$

The fuel–air ratio can be normalized by the stoichiometric ratio to yield a quantity known as equivalence ratio.

$$\phi = \frac{fa_{actual}}{fa_{stoichiometric}} \quad [2]$$

This is a more commonly used measure of mixture ratio. If $\phi = 1$, the fuel–air mixture is stoichiometric. When $\phi < 1$, the mixture is lean, and if $\phi > 1$, it is rich. Combustion reactions typically peak in temperature around an equivalence ratio of unity and are reduced the more lean or rich the mixture. This is primarily due to the unreacted constituents absorbing some of the heat released from the reaction. The theoretical temperature achieved with an adiabatic boundary condition (all released heat is used to heat the products) is known as the adiabatic flame temperature and is a useful reference point.

3.15.2.3 Kinetics

Chemical kinetics is the study of the rate at which a reaction occurs. Generally, the conditions that can impact the reaction rate include reactant concentration, temperature, and the presence of a catalyst. The reaction rate is expressed as the rate of change in concentration of the reacting substances (a decrease for reactants and an increase for products) per unit time. The rate of change can also be equated to a rate constant multiplied by the concentrations of the reactant species raised to the power of their respective stoichiometric coefficients. For example, an arbitrary reaction may be written as

$$A + 2B \rightarrow C + D \quad [IV]$$

where the reaction rate is

$$-R = \frac{d[A]}{dt} = -k[A][B]^2 \quad [3]$$

with brackets denoting concentration and k the rate constant.

The temperature dependence of reaction rate is contained within the rate constant k. This is embodied in the Arrhenius expression for the rate constant that is generally given as

$$k = A\exp\left(\frac{-E_a}{RT}\right) \quad [4]$$

where E_a is the activation energy for the reaction, R the gas constant, T the temperature, and A the kinetic preexponential factor. The pre-exponential factor accounts for molecular collisions and the steric factor, and along with the activation energy, it is usually obtained empirically. Typically, the constants are extracted from a linear fit to data on $\ln(k)$ versus $(1/T)$ plots where the slope of the fit is $(-E_a/R)$. Collision theory and more advanced techniques such as transition state theory can be used to calculate pre-exponentials and rate constants; however, these methods tend to overpredict reaction rates.

The activation energy represents an energy barrier that must be overcome in order to convert from reactants to products. This energy is typically provided in the form of thermal energy, and once it is provided, the reaction rapidly proceeds to completion. Figure 1 shows energy as a function of reaction coordinate for a typical exothermic combustion reaction. In the presence of a catalyst, the activation energy is reduced while all other aspects of the reaction remain unchanged.

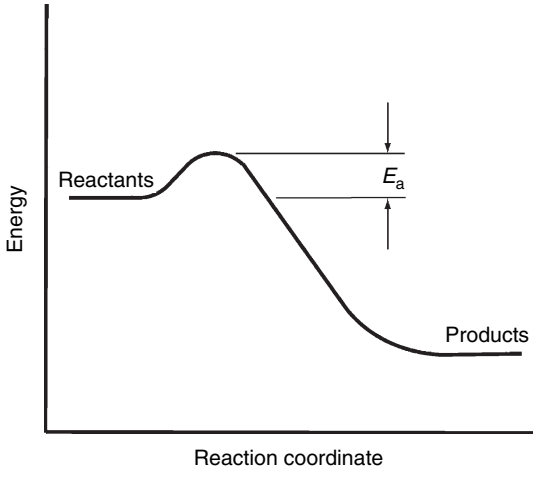


Figure 1 Energy as a function of reaction coordinate for a typical exothermic reaction.

3.15.2.4 Microcombustor Efficiency Definitions

For the purpose of this discussion, overall combustor efficiency will be based on an enthalpy balance across the combustor and is defined as

$$\eta_c = \frac{(\dot{m}_a + \dot{m}_f)h_2 - \dot{m}_a h_1}{\dot{m}_f h_f} \quad [5]$$

where subscript (1) refers to the combustor inlet and subscript (2) refers to the combustor exit. The combustor efficiency can be written as the product of two subefficiencies: a chemical efficiency and a thermal efficiency. These two parameters can be written as

$$\eta_{chem} = \frac{[(\dot{m}_a + \dot{m}_f)h_2 - \dot{m}_a h_1] + \dot{Q}_{loss}}{\dot{m}_f h_f} = \frac{\text{total enthalpy released}}{\text{maximum enthalpy release possible}} \quad [6]$$

$$\eta_{therm} = \frac{(\dot{m}_a + \dot{m}_f)h_2 - \dot{m}_a h_1}{[(\dot{m}_a + \dot{m}_f)h_2 - \dot{m}_a h_1] + \dot{Q}_{loss}} = \frac{\text{enthalpy rise of fluid}}{\text{total enthalpy released}} \quad [7]$$

where $\dot{Q}_{loss}$ is the heat lost to the wall. It is worth noting that alternate definitions of efficiency may be more suitable depending on the application for which a microcombustor is being designed. The above definitions assume that it is desirable to retain enthalpy in the gas stream as opposed to losing heat through the structure. For some applications, a hot wall may be the desired outcome.

3.15.3 Microcombustor Fundamentals and Challenges

There are fundamental concepts and challenges that are unique to microcombustion systems. These include timescale and material constraints as well as heat transfer and fluid–structure thermal coupling. These concepts are outlined below.

3.15.3.1 Timescale Constraints

For many energy conversion applications, power density is the most important metric. Using the gas turbine combustor as a model system, the differences in power density spanning length scales become evident. As summarized in Table 1, the high power density of a microcombustor directly results from high mass flow per unit volume. Because chemical reaction times do not scale with mass flow rate or combustor volume, the

Table 1 Comparison of operating parameters and requirements for a micro gas turbine engine combustor with those estimated for a conventional GE90 gas turbine combustor

	Conventional combustor	*Microcombustor*
Length (m)	0.2	0.001
Volume (m^3)	0.073	6.6×10^{-8}
Cross-sectional area (m^2)	0.36	6×10^{-5}
Inlet total pressure (atm)	37.5	4
Inlet total temperature (K)	870	500
Mass flow rate ($kg\,s^{-1}$)	140	1.8×10^{-4}
Residence time (ms)	~7	~0.5
Efficiency (%)	>99	>90
Pressure ratio	>0.95	>0.95
Exit temperature (K)	1800	1600
Power density ($MW\,m^{-3}$)	1960	3000

Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.
Note: Residence time is calculated using inlet pressure and an average flow temperature of 1000 K.

realization of this high power density is contingent upon completion of the combustion process within a shorter combustor through-flow time.

This fundamental time constraint can be quantified in terms of a homogeneous Damköhler number – the ratio of the residence time to the characteristic chemical reaction time.

$$Da_{\mathrm{h}} = \frac{\tau_{\mathrm{residence}}}{\tau_{\mathrm{reaction}}} \qquad [8]$$

To ensure a Da_{h} greater than unity (and complete combustion), a designer of a microcombustor can either increase the flow residence time or decrease the chemical reaction time. The characteristic combustor residence time is given by estimating the bulk flow rate through the combustor volume.

$$\tau_{\mathrm{residence}} \approx \frac{\text{volume}}{\text{volumetric flow rate}} = \frac{VP}{\dot{m}RT} \qquad [9]$$

Residence time can be increased by increasing the size of the combustion chamber, reducing the mass flow rate, or increasing the operating pressure. A chemical reaction time can be approximated using an Arrhenius-type expression.

$$\tau_{\mathrm{reaction}} \approx \frac{[\mathrm{fuel}]_0}{A[\mathrm{fuel}]^a[\mathrm{O_2}]^b \exp(-E_{\mathrm{a}}/RT_0)} \qquad [10]$$

Eqn [10] indicates that reaction time is primarily a function of fuel properties and the mixture temperature.

Because high power density requirements mandate high mass flow rates through small chamber volumes, the mass flow rate per unit volume cannot be reduced without compromising device power density. Hence, there is a basic trade-off between power density and flow residence time.

$$\text{Power density} \propto \frac{\dot{m}}{V} \propto \frac{\dot{m}_{\mathrm{f}}\mathrm{LHV}}{V} \propto \frac{\rho}{\tau_{\mathrm{residence}}} \qquad [11]$$

For a given operating pressure (and thus density), and assuming a Da_{h} of unity, reducing the chemical reaction time, and thus required residence time, is the only means of ensuring complete combustion without compromising the high power density of the device.

For a catalytic combustor, there is an additional timescale to consider. The reaction time and residence time are still important parameters; however, diffusion time must also be accounted for. The time required for a fuel and oxidizer molecule to come into contact with the active catalytic surface can govern the operation of the device.

As a result, there are several nondimensional parameters that can be examined to determine which process governs performance. Combustor residence time and reaction time can be estimated in the same fashion as done in the gas-phase case. Reaction mechanisms for many catalytic reactions can be found in the literature. The homogeneous Damköhler number is the ratio of these two quantities and indicates their relative importance. With the additional diffusion timescale, a second Damköhler number can be defined. This parameter is the ratio of the diffusion time to the reaction time and indicates the relative rates of these two processes.

$$Da_2 = \frac{\tau_{\mathrm{diffusion}}}{\tau_{\mathrm{reaction}}} \qquad [12]$$

Finally, there is a third ratio of timescales. The ratio of diffusion time to residence time is known as the Péclet number. For high Péclet numbers (>>1), reactants do not have time to diffuse to the active surface and simply pass through the combustor unreacted. A device operating under these conditions is said to be mass transfer controlled or diffusion limited.

$$Pe = \frac{\tau_{\mathrm{diffusion}}}{\tau_{\mathrm{residence}}} \qquad [13]$$

For low Péclet numbers (<<1), the chemical reaction may be controlling the combustion process and the two Damköhler numbers must be examined. If they are both <<1, then the device is kinetically limited or reaction controlled.

For the microscale devices of interest here, power density is the primary performance metric. As discussed previously, high power density requires high mass flow rates through small chamber volumes resulting in relation [11]. At the small volumes of interest, residence time will be correspondingly low while catalytic reaction rates are generally fast. As a result, these devices will likely be diffusion limited. Péclet number must then be considered the governing nondimensional parameter; thus, power density can be rewritten as

$$\text{Power density} \propto \frac{\rho Pe}{\tau_{\text{diffusion}}} \qquad [14]$$

Hence, there is a basic trade-off between power density and diffusion time. Noting that Péclet number can also be written as

$$Pe = \frac{Lv}{D} \qquad [15]$$

where L is reactor length and v velocity, power density can be rewritten as

$$\text{Power density} \propto \frac{\rho Lv}{D\tau_{\text{diffusion}}} \qquad [16]$$

For a given operating pressure and temperature (thus density), and reactant species, reducing diffusion times will increase the power density of the device and improve performance. Diffusion timescales for a given fuel–air mixture generally can be reduced by increasing catalyst surface area-to-volume ratio.

3.15.3.2 Heat Transfer and Fluid–Structure Coupling

Energy loss due to heat transfer at the walls of the combustion chamber in a large-scale flame combustor is typically neglected. However, for a microcombustor, this is an important factor. The surface area-to-volume ratio for a microscale combustor is approximately $500\,\text{m}^{-1}$, or two orders of magnitude greater than that of a typical combustor.

Waitz *et al.* (1998) have shown that the ratio of heat lost to that generated scales with the hydraulic diameter as follows:

$$\frac{Q_{\text{loss}}}{Q} \propto \frac{1}{d_{\text{h}}^{1.2}} \qquad [17]$$

The hydraulic diameter of a microcombustor is on the order of millimeters or less, hundreds of times smaller than that of a typical combustor. Therefore, the ratio of heat lost to that generated may be as much as two orders of magnitude greater than that of a large-scale combustor.

The effect of this large surface heat loss on combustion is twofold. First, large thermal losses have a direct impact on overall combustor efficiency. Therefore, typical large-scale combustor efficiencies greater than 99% are not feasible. Second, the heat loss leads to lower reaction temperatures that can increase kinetic times and narrow flammability limits. This can exacerbate the constraints of short residence time.

In a catalytic device, the heat loss problem becomes worse. The reaction and heat release process occur on the solid surface of the catalyst, so there is a direct conduction path through the structure and out of the device. In the case of a gas-phase microcombustor, there is a thermal boundary layer between the hot reacting gases and the wall, helping to limit heat loss. In a catalytic device, the thermal boundary layer acts to isolate the hot wall from the relatively cooler free-stream flow. In addition, surface area-to-volume ratio is likely to be even higher in a catalytic combustor because of the need to fill the chamber with a monolith or porous substance coated with an active catalytic material.

However, it is important to note that for some applications such as a heat source for a solid-state power generator (e.g., thermoelectric or thermophotovoltaic devices), or for an endothermic reactor (e.g., a methanol reformer for a fuel cell), heat loss through the structure to some other device is desirable.

3.15.3.3 Material Constraints

Choice of materials for microcombustors is often limited due to the available fabrication techniques. Typically, deep reactive ion etching (DRIE) in silicon wafers and aligned fusion bonding are used to generate 2D extruded geometries that serve as combustion chambers, fuel–air mixers, fuel injectors, and other associated components. Consequently, many microcombustion systems are silicon based.

There are several material constraints imposed upon a silicon microcombustion system. The most critical requirement is a wall temperature limit of less than 950 K. At temperatures above this level, silicon begins to soften and lose its structural integrity. However, high surface heat transfer and the high thermal conductivity of silicon are beneficial in this case. Combustor wall temperatures can be kept below the 950 K requirement by conduction of heat through the structure to the ambient.

It is also important to note that silicon-based microcombustors are typically isothermal. This is a result of the combination of the high thermal conductivity of silicon and the short conduction paths inherent in a microsystem. This can be quantified by the heat transfer Biot number

$$Bi = \frac{hL}{k} = \frac{\text{resistance to conductive heat transfer}}{\text{resistance to convective heat transfer}} \quad [18]$$

which is typically <<1 for these systems.

There are several material and fabrication issues that are specific to developing a catalytic microcombustor. First, the catalyst material itself can be problematic. For noble metals such as platinum (a typical hydrocarbon–air catalyst), agglomeration of thin layers can occur at elevated temperatures. This phenomenon has been shown to occur at temperatures on the order of 1000 K (Maskell *et al.* 1987, Wang *et al.* 2003). This is a potential problem not only during the combustion process but also during any high-temperature fabrication steps such as silicon wafer bonding and glass seal packaging processes, both of which occur at temperatures approaching 1350 K. Agglomeration will reduce the active surface area and negatively impact the performance.

The substrate material on which the active catalyst is deposited must also be chosen with care. It should be a material with a temperature limit at least as high as that of the catalyst itself and be inert with respect to the catalyst. Diffusion of the catalyst layer or reaction with its substrate would also further reduce the surface area and the performance. Finally, thermal shock of the catalyst and substrate material must be considered, especially if ceramic substrates are used.

3.15.4 Homogeneous Gas-Phase Microcombustion

For the purpose of the following detailed discussion, a baseline microcombustor will be used as a reference. It should be noted, however, that the conclusions drawn from data generated in this device are generally applicable to other microcombustors.

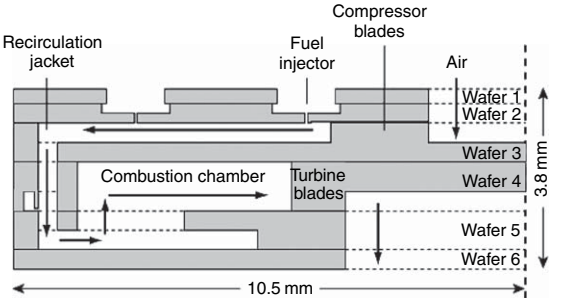


Figure 2 Schematic of a six-wafer microcombustor. (Source: Mehra A 2000 Development of a high power density combustion system for a silicon micro gas turbine engine. Ph.D. thesis, Massachusetts Institute of Technology.)

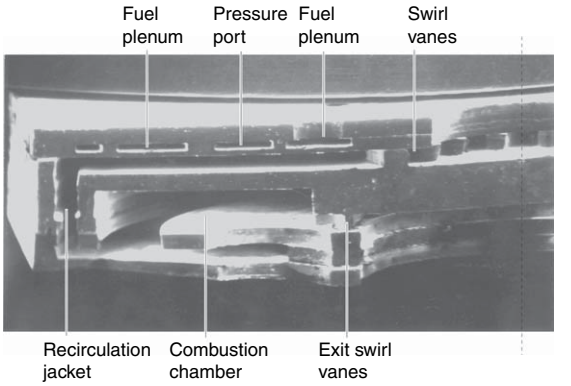


Figure 3 Scanning electron micrograph (SEM) of a six-wafer microcombustor. (Source: Mehra A 2000 Development of a high power density combustion system for a silicon micro gas turbine engine. Ph.D. thesis, Massachusetts Institute of Technology.)

An axisymmetric microcombustor fabricated from six silicon wafers using DRIE and aligned fusion wafer bonding is shown in **Figures 2** and **3**. It consists of a 191-mm^3 radial inflow combustion chamber, various flame holding chamber inlet geometries, a recirculation jacket that serves as both a thermal management system and a fuel–air mixer, and several sets of fuel injectors.

3.15.4.1 Performance of a Baseline Gas-Phase Microcombustor

Tests were performed with premixed hydrogen–air in the radial inflow microcombustor. Plots of combustor exit gas temperature and overall combustor efficiency for constant equivalence ratio are shown in **Figures 4** and **5**, respectively. Mass flow rate was

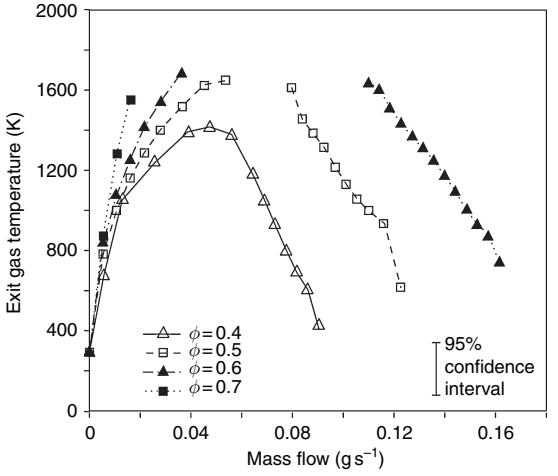


Figure 4 Exit gas temperature versus mass flow rate for an annular six-wafer microcombustor. (Source: Mehra A 2000 Development of a high power density combustion system for a silicon micro gas turbine engine. Ph.D. thesis, Massachusetts Institute of Technology.)

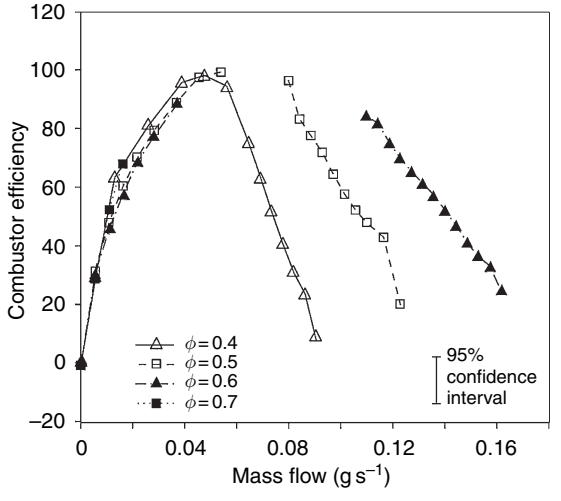


Figure 5 Overall efficiency for an annular six-wafer microcombustor. (Source: Mehra A 2000 Development of a high power density combustion system for a silicon micro gas turbine engine. Ph.D. thesis, Massachusetts Institute of Technology.)

varied by increasing the inlet pressure with a fixed exit area to atmosphere.

The combustor produced exit gas temperatures in excess of 1600 K for a mass flow rate of 0.11 g s^{-1} and an operating pressure of 1.13 atm. The overall efficiency in this case exceeded 90%, and the power density was approximately 1100 MW m^{-3}. The break in the $\phi = 0.5$ and $\phi = 0.6$ exit temperature and efficiency curves is due to a lack of measurement

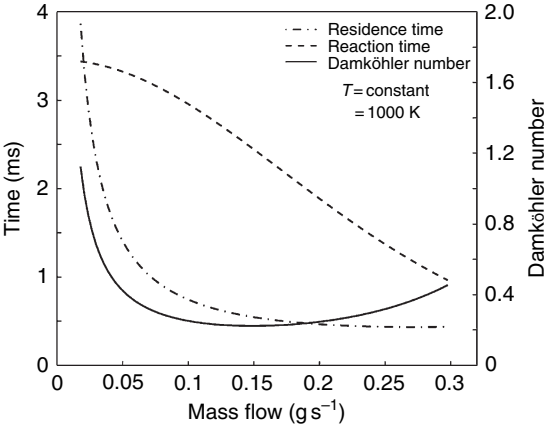


Figure 6 Reaction time, residence time, and Damköhler number (Da_h). (Source: Mehra A 2000 Development of a high power density combustion system for a silicon micro gas turbine engine. Ph.D. thesis, Massachusetts Institute of Technology.)

capabilities for gas temperatures over 1600 K. The $\phi = 0.7$ curve terminates at a mass flow rate of 0.015 g s^{-1} due to upstream burning in the recirculation jacket, which can result in overheating and destruction of the device (Mehra 2000, Mehra *et al.* 2000, Spadaccini *et al.* 2003).

The flame blowout characteristics of the combustor can be explained by examining the change in Da_h over a constant equivalence ratio operating line. Using eqn [9] for residence time, and a simple one-step mechanism for hydrogen–air combustion,

$$\tau_{\text{reaction}} = \frac{[H_2]_0}{A[H_2]^2[O_2]\exp(-E_a/RT)} \qquad [19]$$

with $E_a = 10\,950$ cal mol^{-1} and $A = 1.62 \times 10^{18}$ (cm^3 mol^{-1})^2 s, an estimate of Da_h can be obtained (Mehra 2000). **Figure 6** shows a plot of reaction time and residence time along with Da_h versus mass flow rate. Residence time decreases more rapidly than reaction time, resulting in a Da_h less than one. This leads to chemical inefficiency at high mass flow rates and flame blowout.

3.15.4.2 Recirculation Zones and Flame Stability

Two types of combustion chamber inlets were utilized in this device to understand the impact of different flame holding recirculation zones. The first is a simple annular inlet 1.2 mm wide. The second is an array of 60 slots 2.2 mm long intended to create multiple small recirculation zones for more

rapid and uniform ignition of the incoming flow. Scanning electron micrographs (SEMs) of both inlets are shown in **Figure 7** (Mehra 2000).

Combustor efficiency for premixed hydrogen–air tests in the slotted-inlet geometry is plotted in **Figure 8**. Performance is similar to that of the annular-inlet combustor. However, the slotted-inlet device produced higher exit temperatures and efficiencies in the high mass flow, kinetically limited regime of the operating line. This is attributed to the presence of multiple recirculation zones at the inlet, which more rapidly and uniformly ignite the incoming mixture. The sharp drop in performance for the slotted combustor is due to the rapid extinction of the ignition zones. In the annular case shown in **Figures 5** and **6**, the single large recirculation zone stays lit over a wider range of mass flow rates. However, it also results in lower chemical efficiency (Mehra 2000).

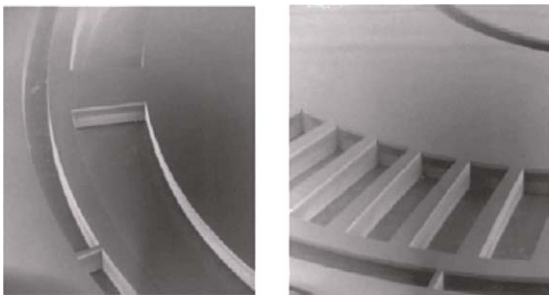


Figure 7 Scanning electron micrographs (SEMs) of inlet geometries. (Source: Mehra A 2000 Development of a high power density combustion system for a silicon micro gas turbine engine. Ph.D. thesis, Massachusetts Institute of Technology.)

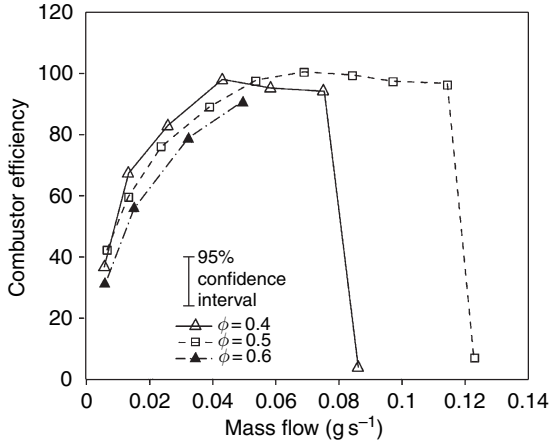


Figure 8 Overall efficiency for a slotted-inlet six-wafer microcombustor. (Source: Mehra A 2000 Development of a high power density combustion system for a silicon micro gas turbine engine. Ph.D. thesis, Massachusetts Institute of Technology.)

Without internal measurements or flow visualization, numerical simulations can be used to understand this performance difference. **Figure 9(a)** shows temperature contours generated from computational fluid dynamics (CFD) software, for the combustor with an annular inlet. In **Figure 9(b)**, contours of temperature are plotted for a 2D cut through the combustor inlet in the $\theta - z$ plane. These contours show a temperature gradient from the bottom of the combustion chamber to the top and indicate the primary ignition zone is in the upper right corner of the combustor. **Figure 10(a)** and **10(b)** show the same contour plots for a combustor with a slotted inlet. The temperature contour in the $\theta - z$ plane shows significantly higher temperatures in the lower regions of the combustion chamber and a higher overall temperature. The shape of the contours indicates that there are small, hot ignition zones near the slotted inlet (Spadaccini 2004, Spadaccini *et al.* 2003).

3.15.4.3 Fuel Injection and Fuel–Air Mixing

Fuel injection and mixing were evaluated in these devices by testing three sets of fuel injectors at different locations within the recirculation jacket and comparing overall combustor efficiencies to a premixed case. The premixed operation consisted of fuel–air injection through the primary air inlet only. **Figure 11** plots efficiency for different equivalence ratios at a mass flow rate of 0.045 g s^{-1}. The injectors were located at radii of 4.8 and 8.0 mm into the recirculation jacket while a set of sidewall injectors were at the base of the recirculation jacket just upstream from the combustion chamber. The figure indicates that performance decreases as the injectors get closer to the combustion chamber due to the decreased mixing length and time (Mehra 2000).

3.15.4.4 Primary and Secondary Zones

Typical large-scale combustors utilize a multizonal combustion strategy. An initial primary zone consists of a high-temperature, near-stoichiometric combustion zone followed by dilution jets of air to form a secondary zone with a leaner mixture and lower temperature. Ultimately, most of the fuel is consumed and a more uniform and controlled exit temperature can be achieved.

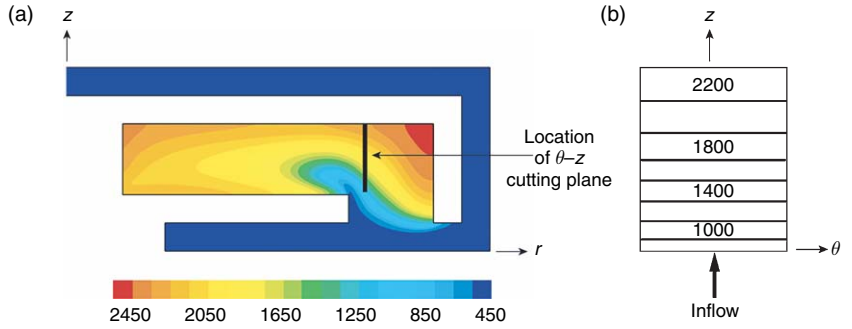


Figure 9 Temperature (K) contours for an annular-inlet microcombustor. Note: FLUENT v. 5.1 with a nine-species, 20-step, hydrogen–air reaction mechanism for the chemical model and an adiabatic wall boundary condition were used in these simulations. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

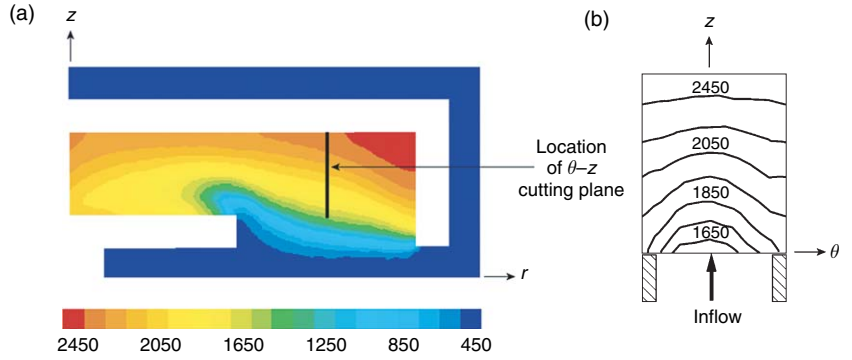


Figure 10 Temperature (K) contours for a slotted-inlet microcombustor. Note: FLUENT v. 5.1 with a nine-species, 20-step, hydrogen–air reaction mechanism for the chemical model and an adiabatic wall boundary condition were used in these simulations. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

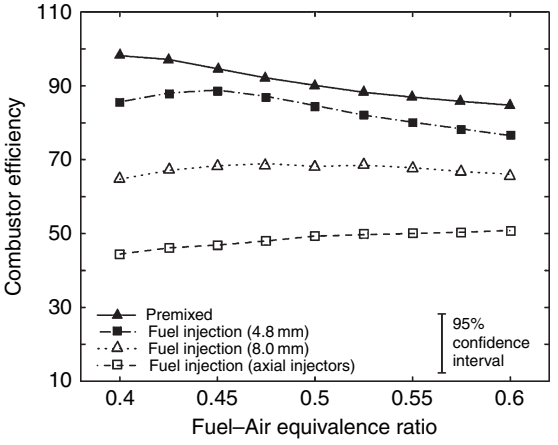


Figure 11 Overall efficiency for fuel injection schemes. (Source: Mehra A 2000 Development of a high power density combustion system for a silicon micro gas turbine engine. Ph.D. thesis, Massachusetts Institute of Technology.)

A multizone microcombustor that operates in a similar fashion may be beneficial for overall performance. This was fabricated by modifying the geometry shown in **Figures 2** and **3**. A series of holes through the inner wall were created to connect the upper recirculation jacket to the combustion chamber. These holes allowed inlet air to bleed into the combustion chamber. This dilution air serves two purposes: (1) it splits the combustion chamber into the two zones and (2) it dilutes the hot combustion products reducing their temperature to the required outlet temperature. In the primary zone, the fuel–air mixture is burned near stoichiometric conditions. By burning at higher temperature, kinetic rates increase and reaction time decreases, effectively increasing Da_h. The dilution jets also serve to strengthen recirculation zones and improve flame stability. Due to the nature of this mode of operation, premixed combustion is not feasible and fuel injection occurs just

downstream of the dilution holes in the cooling jacket. A 3D, adiabatic, reacting flow CFD solution shows the effect of the two zones in **Figure 12**.

The dual-zone combustor exhibited slightly lower exit gas temperatures and efficiencies than the baseline microcombustor. However, its operating range was much wider for hydrogen–air operation. Combustion was stabilized at equivalence ratios as low as 0.2 and mass flow rates as high as 0.2 g s^{-1}. **Figure 13** compares the dual-zone combustor efficiency with the baseline for similar conditions (Spadaccini 2004, Spadaccini *et al.* 2003).

The significant extension in operating range is largely due to strengthened recirculation zones and the hot, stable primary zone. For the same overall efficiency, the dual-zone combustor achieved a 100% increase in mass flow rate before blowout. The ability to operate at low overall equivalence ratio is a result of a higher local equivalence ratio in the primary zone. **Table 2** summarizes the maximum power densities achieved for both devices with various fuels (hydrocarbon fuel data have not been presented here) (Spadaccini 2004).

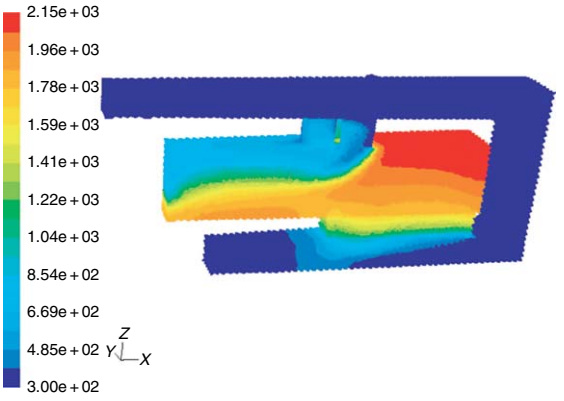


Figure 12 Temperature (K) profile from 3D, reacting flow computational fluid dynamics (CFD) solution for dual-zone microcombustor. (Source: Lee J-W 2000 Numerical simulation of a hydrogen microcombustor. M.S. thesis, Massachusetts Institute of Technology.)

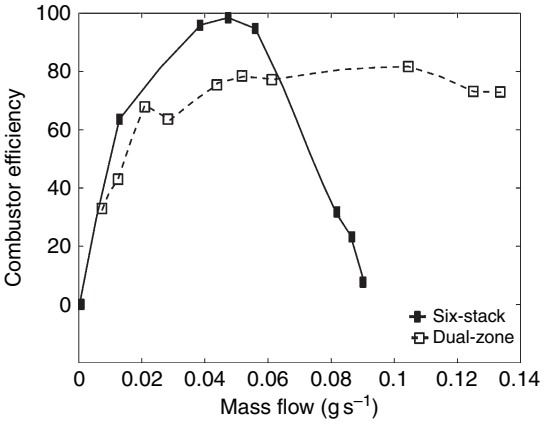


Figure 13 Efficiency comparison of dual-zone and baseline microcombustors with hydrogen–air at ϕ=0.4. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

3.15.4.5 Operating Space

Microcombustor performance is a function of both chemical and thermal losses. Performance can be quantified in terms of nondimensional power density, defined as power density normalized by the maximum possible power density at a given operating condition (the latter being a function of mass flow rate and fuel type). If it is assumed that there is adequate fuel–air mixing and rapid, uniform ignition, this quantity is a function of only two parameters – Damköhler number and nondimensional heat loss. Damköhler number can be estimated as previously described using exit gas temperature for residence time calculations and Chemkin III (a reaction kinetics software package) with detailed reaction mechanisms from the literature for chemical time approximations. Heat loss can be quantified in nondimensional terms by normalizing the heat lost from the device by the maximum possible heat that can be generated at a given operating condition. The heat lost was estimated using a 1D heat transfer model (which is a good approximation in this case due to the low Biot number and isothermal nature of the structure).

Figure 14 plots this nondimensional power density for several microcombustors in three-parameter space. This plot includes multiple fuels as well (hydrogen, ethylene, and propane). The heat loss parameter and homogeneous Damköhler number are on the x- and y-axes, respectively. A second-order, least squares surface fit was used to generate contours of nondimensional power density (the fit explains 65% of the variance in the data) (Spadaccini 2004, Spadaccini *et al.* 2003).

Thermal efficiency in these devices is inversely proportional to heat loss, whereas chemical efficiency is directly proportional to Damköhler number. Hence, optimum performance is achieved at low levels of heat loss and high Damköhler number as indicated in **Figure 14**.

Table 2 Maximum power densities and efficiencies for microcombustors

	Baseline microcombustor		*Dual-zone microcombustor*	
Fuel	*Power density (MW m^{-3})*	*Efficiency (%)*	*Power density (MW m^{-3})*	*Efficiency (%)*
H_2	1150	>99	1400	92
C_2H_4	500	85	100	38
C_3H_8	140	50	75	29

Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.

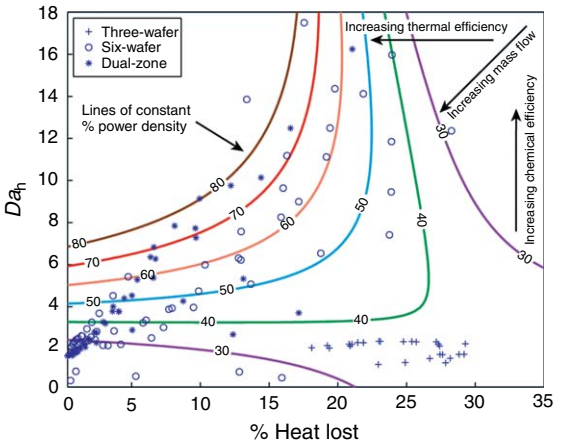


Figure 14 Nondimensional operating space. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

This nondimensional operating space further explains the operating line considerations of a microcombustor. Conditions of increasing mass flow for constant equivalence ratio move from the upper right side of the operating space to the lower left. In the low mass flow region of the operating space, performance is limited by heat loss, whereas at high mass flow, Damköhler number falls and performance is limited by chemical inefficiency. In the area of moderate mass flow, the center of the operating space, the highest power densities are achieved.

3.15.4.6 Suggested Design Methodology and Recommendations

The nondimensional operating space of Figure 14 can also be useful in the design process. For example, a good estimate of the volume required for a propane–air microcombustor suitable for a given application can be obtained. If operating conditions such as a 1600 K exit gas temperature, 1 atm pressure, a mass flow rate of 0.15 g s^{-1}, and an equivalence ratio of $\phi = 0.8$ are required, the nondimensional power density can be computed and is found to be approximately 60%. The heat loss from the device can be assumed to be approximately 5% based on data for similar devices or can be estimated using a heat transfer model. From Figure 14, the desired Damköhler number can be found and is approximately 5.5. The chemical time can be computed using Chemkin III or any other suitable technique and the desired residence time found from the Damköhler number. This results in a residence time of 1.44 ms and a volume of 950 mm^3, or about a fivefold increase in volume compared with the gas-phase microcombustors presented here.

It is important to note that these results are for design point operation only and operating line issues must be kept in mind. Namely, the combustors will likely reach a minimum Damköhler number at which blowout may occur, before reaching the design mass flow rate. To account for this, a gas-phase microcombustor volume should be sized for this off-design operating condition.

The following design trade-offs and recommendations for homogeneous gas-phase microcombustion systems should be considered:

1. The high power density requirements of a microcombustor mandate high mass flow rates per unit volume. However, this results in low residence time and Damköhler numbers of order unity or less. Microcombustion systems are fundamentally limited by this trade-off between high power density and efficient combustion.
2. The large surface area-to-volume ratio of a microcombustor makes nonadiabatic operation intrinsic to its design. The coupling between heat transfer and chemical kinetics is a critical element of the design process. Thermal management strategies such as the recirculation jacket shown in Figure 2 are effective solutions. While keeping

the inner wall cool, it acts as a fuel–air preheater to improve kinetic rates, mitigating the effect of lower wall temperatures.

3. Recirculation zones are critical for rapid and uniform ignition of incoming reactants. However, if these ignition zones are too large, the effective volume of the combustion chamber will be significantly reduced, resulting in decreased residence time and Da_h.
4. Fuel injection for non-premixed operation should occur as far upstream of the combustion chamber as possible to provide adequate mixing length.
5. It is not sufficient to design a microcombustor exclusively for design point operation. Operating line issues need to be considered.

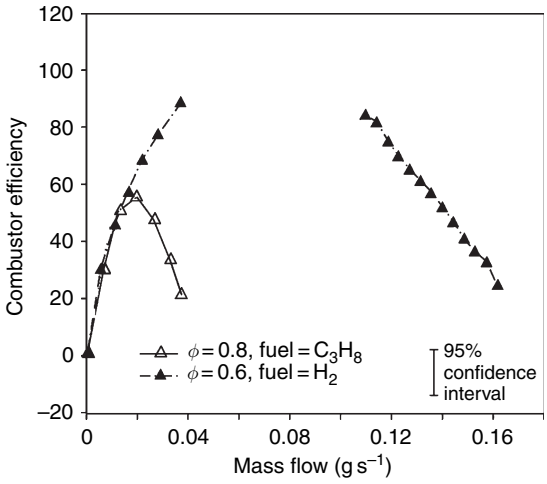


Figure 15 Performance comparison for different fuel types in a gas-phase microcombustor. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

3.15.5 Heterogeneous Catalytic Microcombustion

Thus far, this chapter has focused predominantly on homogeneous gas-phase combustion. We have an understanding of the combustor volume required for a practical device, given the requirements for mass flow, exit gas temperature, and fuel type. This has also provided an understanding of the operating behavior and performance of a microcombustor with homogeneous gas-phase hydrocarbon–air mixtures. Experimental results described by Mehra *et al.* and Spadaccini *et al.*, and those from the performance estimation method presented in Section 3.15.5.6 suggest that a microcombustor operating with a more practical hydrocarbon fuel such as propane will need to be at least five times larger than a device operating with hydrogen. A sample of experimental results is shown in **Figure 15** for both hydrogen–air and propane–air combustion to underscore this point. The factor of five difference in mass flow range for the two fuels is attributable to the difference in chemical kinetic rates.

Ultimately, for most practical applications of microcombustors, the goal is to operate with storable liquid hydrocarbon fuels such as propane or ethylene. To utilize these fuels without increasing combustor volume or temperature, chemical conversion rates must be increased significantly. Surface catalysis is a means of increasing reaction rates. Heterogeneous reactions of hydrocarbon–air mixtures over noble metal catalysts are known to be faster than those occurring in the gas phase. For this basic reason, the development of catalytic microcombustors has been pursued.

A typical large-scale catalytic combustor would consist of a monolithic structure (a set of parallel channels) coated with an active catalyst material. Common catalyst materials include noble metals such platinum and palladium as well as more exotic and proprietary metal oxides. The fuel and air flow through the channels and diffuse to the active surface. When a fuel molecule and an oxygen molecule reach the surface, they are adsorbed and they react, releasing heat. The products of the reaction are then desorbed and they flow out of the channel. The heat that has been liberated from the reaction is transferred both conductively through the solid and convectively to the fluid. Eventually, the remaining fuel–air mixture can get hot enough to ignite in the gas phase. **Figure 16** shows a schematic of a typical catalytic combustor.

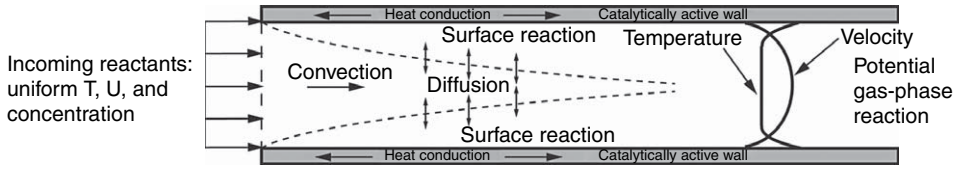


Figure 16 Schematic of a typical catalytic combustor.

For the purpose of illustrating key points and showing operational data, a model catalytic microcombustor, based on the gas-phase device previously presented, will be used. **Figure 2** shows the gas-phase geometry; however in the catalytic case, the combustion chamber is filled with a porous substrate coated with platinum that will act as the active catalytic surface. Although a porous catalyst material was chosen for this application, another common catalyst geometry is the monolithic channel-type structure with catalytically active walls, as previously described.

3.15.5.1 Ignition Techniques and Characteristics

To achieve autothermal operation of a catalytic microcombustor, the catalyst has to be heated to a suitable ignition temperature. This can be accomplished by one of the following three methods:

1. Begin with hydrogen–air operation (which will catalytically ignite at lower temperatures) and preheat the catalyst. The fuel can then be converted to a hydrocarbon fuel once the catalyst has reached adequate temperature (typically >500 K).
2. Externally heat the entire microcombustor with a heater and begin flow of the hydrocarbon–air mixture when the desired preheating temperature is achieved.
3. Heat the catalytic element directly by resistive heating followed by flowing of the fuel–air mixture. This requires that the catalyst itself be fabricated with electrical contacts.

Typically, a catalytic combustor will exhibit an ignition/extinction hysteresis similar to that shown by Schmidt and co-workers (Goralski and Schmidt 1996, Veser and Schmidt 1996, Williams *et al.* 1991). For the test device used as the model here, this is shown in **Figure 17** with two different porous catalyst materials. The entire combustor was heated via an external electrical heater while a hydrogen–air mixture was flowing through the device. Wall temperature is plotted against heater power and ignition occurs at a heater power of approximately 20–30% (around 80–100°C). Upon removal of the external heater, the combustor continues to operate autothermally, resulting in the hysteresis.

3.15.5.2 Performance of a Baseline Catalytic Microcombustor

The baseline catalytic device contained a 95% porous foam material coated with platinum as the active surface. Porosity can be defined as the open volume divided by the total volume.

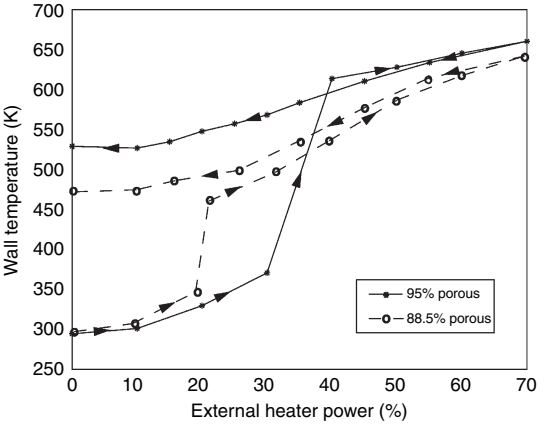


Figure 17 Ignition characteristics for a catalytic microcombustor. (Source: Spadaccini CM 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

$$\alpha = \frac{V_{\text{open}}}{V_{\text{total}}} \qquad [20]$$

Plots of exit gas temperature and overall combustor efficiency for a range of mixture ratios are shown in **Figures 18 and 19**, respectively. The maximum exit gas temperature achieved was approximately 850–900 K over a wide range of flow rates, and efficiencies were up to 30%. The sharp drop in performance at $0.35\,\text{g}\,\text{s}^{-1}$ was a result of transients from the experimental setup that was operating at

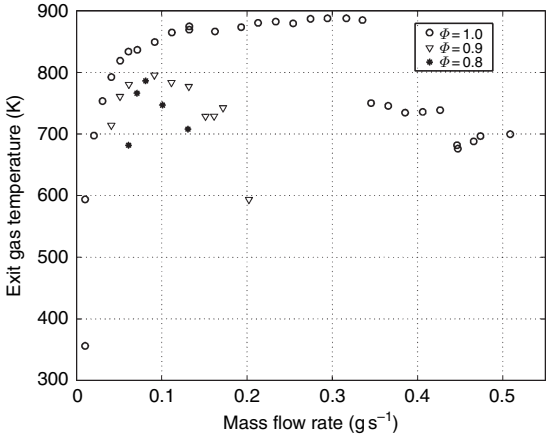


Figure 18 Exit gas temperature plot for a catalytic microcombustor with platinum. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

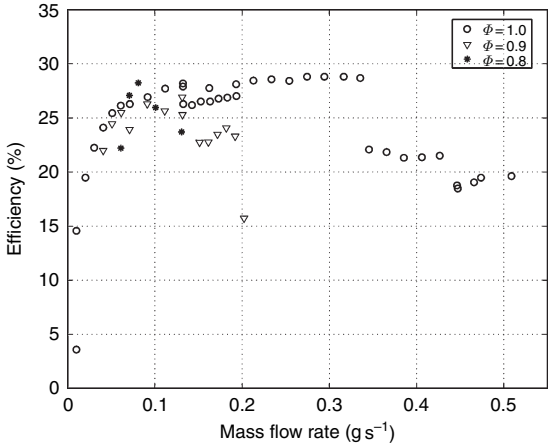


Figure 19 Overall combustor efficiency for a catalytic microcombustor with platinum. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

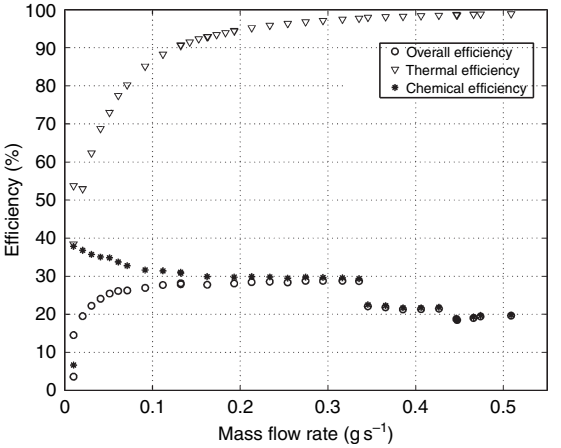


Figure 20 Efficiency breakdown for a catalytic microcombustor with platinum, $\phi = 1.0$. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

maximum flow levels. Although these temperatures and efficiencies are quite low, the mass flow rates achieved were in excess of $0.4\,\mathrm{g\,s^{-1}}$, well beyond those measured for the baseline gas-phase device. The maximum power density achieved was approximately $1050\,\mathrm{MW\,m^{-3}}$, which is a 7.5-fold increase over gas-phase propane–air power densities, and about 95% of that for gas-phase hydrogen–air mixtures (Spadaccini 2004, Spadaccini *et al.* 2007).

The overall combustor efficiency can be divided into its thermal and chemical components. A wall temperature measurement combined with a 1D heat transfer model reveals that heat loss is greater than in the gas-phase case; the losses are dominated by chemical inefficiency, indicating that most of the reactants pass through the device without contacting the active surface. **Figure 20** shows an estimate of the efficiencies in the device for a stoichiometric mixture ratio.

3.15.5.3 Diffusion Characteristics and Timescale Analysis

To determine which phenomena control the combustion process in a catalytic microcombustor, relevant physical timescales can be evaluated. These include reaction time, residence time, diffusion time of the fuel species, and diffusion time of the oxidizer. Residence time can be estimated from the volume, mass flow rate, pressure, and bulk gas temperature as in the gas-phase case (eqn [9]). Reaction rate can be obtained from an Arrhenius-type rate expression. For a propane–air reaction on a platinum catalyst, the following mechanism can be used (Hayes and Kolaczkowski 1997)

$$-R_{C_8H_8} = -k_S[C_3H_8] \qquad [21]$$

where the rate constant is

$$k_S = 2.4 \times 10^5 \exp\left(\frac{-1.08 \times 10^4}{T}\right) \qquad [22]$$

More complex reaction mechanisms can be found in the literature.

Molecular diffusion coefficients for propane and oxygen diffusion through air can be obtained from the Fuller correlation (Hayes and Kolaczkowski 1997):

$$D_{AB} = \frac{1.013 \times 10^{-2} T^{1.75} \left(\frac{1}{M_A} + \frac{1}{M_B}\right)^{1/2}}{P\left[\left(\sum v_i\right)_A^{1/3} + \left(\sum v_i\right)_B^{1/3}\right]^2} \qquad [23]$$

A and B represent the diffusing species and the medium, respectively while M_A and M_B are the molecular weights of the species and v_i is the diffusion volume of each atom in the molecule. Diffusion volumes for carbon, hydrogen, and oxygen are 16.5, 1.98, and 5.48, respectively. A table of diffusion volumes can be found in Hayes and Kolaczkowski or elsewhere in the literature. This correlation is based on collision theory and D_{AB} is in units of square meter per second.

Using estimates of these various timescales, nondimensional parameters can be calculated and used to determine the governing physical phenomena. These parameters and their approximate values for

a catalytic microcombustor are summarized in Table 3. From this simplified timescale analysis, it is clear that the diffusion of reactants to the surface is the controlling parameter. The second Damköhler number indicates that the surface reaction rate is much faster than the rate of diffusion to the surface. The high Péclet number indicates that reactants can flow through the device without coming into contact with the active catalytic surface. The longest timescale in the system is the diffusion time.

If a tubular plug flow reactor is assumed, the Péclet number can be estimated and shown to be a strong function of geometry. For a given set of flow conditions (pressure, temperature, and mass flow rate), the Péclet number can be calculated for a range of diameters (or pore sizes). Figure 21 shows this parameter for $P = 2$ atm, $T = 1000$ K, and a mass flow rate = $0.3\,\mathrm{g\,s^{-1}}$ for both propane and oxygen. For the diameters greater than ~100 μm, the Péclet number is greater than unity. It is also important to note that propane diffuses more slowly than oxygen due to its larger molecular size. This results in a higher Péclet number and indicates that propane diffusion to the active surface is the governing phenomenon.

Table 3 Summary of nondimensional parameters

	Nondimensional parameter	*Range*
Damköhler number for residence time	$Da_h = \frac{\tau_{residence}}{\tau_{reaction}}$	0.5–5
Damköhler number for diffusion time	$Da_2 = \frac{\tau_{diffusion}}{\tau_{reaction}}$	30–500
Péclet number	$Pe = \frac{\tau_{diffusion}}{\tau_{residence}}$	50–130

Sources: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology and Spadaccini C M, Peck J, Waitz I A 2007 Catalytic combustion systems for micro-scale gas turbine engines. *ASME J. Eng. Gas Turbines Power* **129**, 49–60.

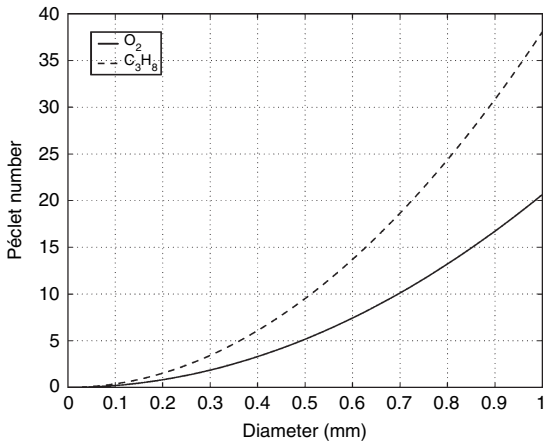


Figure 21 Péclet number versus diameter. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

3.15.5.4 Pressure Loss

For many microcombustor applications, minimizing pressure loss through the device is critical. The addition of a porous catalyst material or a monolithic catalyst support can significantly increase the pressure loss of the system.

Pressure loss through a porous substrate can be estimated from the Ergun equation for Reynolds numbers less than 300 (based on the thickness of a fiber or diameter of a catalyst particle) (Hayes and Kolaczkowski 1997):

$$-\frac{dP}{dz} = \frac{v}{L}\frac{(1-\alpha)}{\alpha^3}\left[\frac{150\mu(1-\alpha)}{L} + 1.75\rho v\right] \quad [24]$$

where μ is the viscosity, v velocity through the porous media, and L a characteristic length scale, usually the width of a foam fiber or diameter of a particle in a packed bed. For a monolithic catalyst support, pressure loss through a single channel can be estimated using a Darcy friction factor (Fay 1994)

$$f = \frac{64}{Re_d} \quad [25]$$

and

$$-\frac{dP}{dz} = f\left(\frac{L}{d_h}\right)\left(\frac{1}{2}\rho v^2\right) \quad [26]$$

3.15.5.5 Isothermal Plug Flow Reactor Model

In most cases catalytic microcombustors can be modeled as isothermal plug flow reactors. The equations developed here have been derived from a control volume analysis of a fluid element in a plug flow reactor. These equations, coupled with an isothermal assumption, yield trends in fuel conversion and bulk gas temperature rise through a reactor as a function of flow conditions and geometry.

The Péclet number analysis indicated that the diffusion of fuel species is the governing phenomenon; therefore, only the mass transport and consumption of

the fuel species is taken into account in subsequent derivations. Homogeneous gas-phase reactions will be neglected, and the catalytic reaction mechanism shown in eqn [21] and [22] will be utilized (propane–air over platinum will be used as a model reaction and for comparison to experiment).

A steady-state gas-phase mole balance across the control volume results in

$$C_b v \frac{dY_{f,b}}{dz} + a_v k_m C_b \left(Y_{f,b} - Y_{f,s}\right) = 0 \quad [27]$$

The surface area-to-volume ratio can be written in terms of porosity and foam fiber thickness as

$$a_v = \frac{4(1-\alpha)}{w} \quad [28]$$

A mole balance performed at the catalyst surface yields

$$k_m C_b \left(Y_{f,b} - Y_{f,s}\right) = (-R_f)_S (1-\alpha) \quad [29]$$

whereas an energy balance across the fluid element provides the final equation

$$-\rho C_p v \frac{dT_b}{dz} + a_v b (T_S - T_b) = 0 \quad [30]$$

Analytical solutions to these equations can be found or an appropriate numerical method can be applied to solve the system. In the case of a non-isothermal structure, a fourth energy balance equation across a differential element in the wall, accounting for conduction in the solid, would be required. Also, a similar set of equations can be obtained for a monolithic channel-type system by modifying eqn [28].

Correlations for gases in packed bed reactors can be used to approximate the heat and mass transfer coefficients in porous substrates. These transport coefficients are given in terms of nondimensional j factors – j_D for mass transfer and j_H for heat transfer (Hayes and Kolaczkowski 1997).

$$j_D = \frac{k_m}{v} Sc^{2/3} \quad [31]$$

$$j_H = \frac{h}{C_p \rho v} Pr^{2/3} \quad [32]$$

These j factors can be estimated using the following correlations (Hayes and Kolaczkowski 1997).

$$j_D = j_H = 0.91 Re_j^{-0.51} S_f \quad Re_j < 50 \quad [33]$$

$$j_D = j_H = 0.91 Re_j^{-0.41} S_f \quad Re_j > 50 \quad [34]$$

S_f is the shape factor, and for this analysis, a value of 0.79 was used (shape factor for ring-type structures) (Hayes and Kolaczkowski 1997). The Reynolds number used here can be found from

$$Re = \frac{\rho v w}{6(1-\alpha)\mu S_f} \quad [35]$$

The boundary conditions required to execute this simplified model include the inlet bulk gas temperature, the inlet bulk gas fuel mole fraction (obtained from equivalence ratio), and the inlet surface mole fraction. The catalyst surface temperature also must be specified, which due to the isothermal condition is constant throughout the reactor. A typical set of boundary and flow conditions is summarized in **Table 4**.

Figures 22 and **23** show temperature and fuel concentration profiles, respectively. Along with substrate porosity of 95% and an average foam fiber

Table 4 Typical boundary and flow conditions for reactor inlet in 1D isothermal plug flow model

Parameter	*Value*
T_b (K)	500
T_S (K)	1000
$Y_{f,b}$ (from ϕ)	0.04 ($\phi = 1.0$)
$Y_{f,S}$	0.00
Pressure (atm)	2
Mass flow (g s^{-1})	0.3

Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.

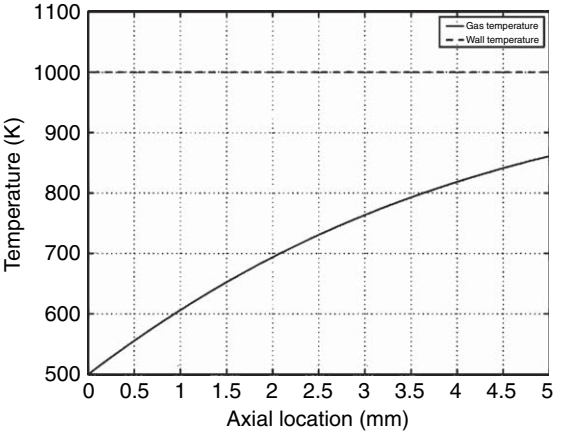


Figure 22 Axial temperature profile in a porous media plug flow reactor. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

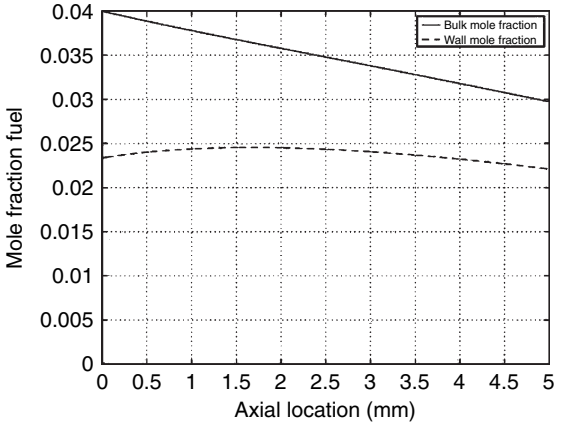


Figure 23 Axial fuel concentration profile in a porous media plug flow reactor. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

thickness of 90 μm, the conditions listed in **Table 4** were used to generate these results.

With this model, the effect of key design variables such as porosity and surface area can be further examined. For constant flow conditions, the model can produce fuel conversion profiles for various porosity and surface area materials. **Figure 24** shows trends in fuel conversion for increasing surface area-to-volume ratios using the conditions listed in **Table 4**. The profiles indicate that higher surface area-to-volume ratios (usually a result of lower porosity for a constant fiber thickness) will significantly improve fuel conversion.

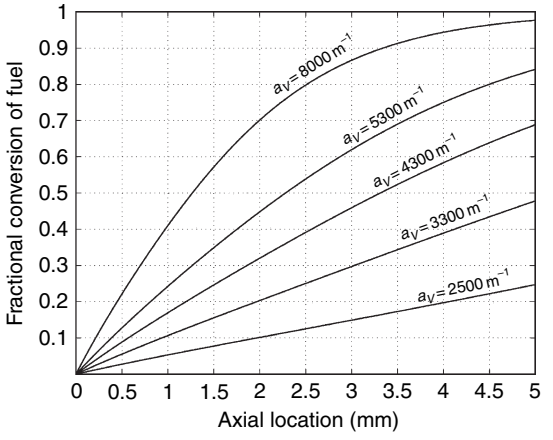


Figure 24 Fuel conversion profiles for various porosities in a porous media plug flow reactor. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

3.15.5.6 Operating Space

The 1D isothermal plug flow model can also be used to visualize the operating space of a catalytic microcombustor. **Figure 25** shows lines of constant power density on a plot of total pressure loss through the device versus catalyst temperature for an equivalence ratio of unity. The maximum power density is obtained from the exit gas temperature predicted by the model. **Figure 25** indicates that at higher pressure loss (higher surface area-to-volume ratio) and higher catalyst temperatures, combustor performance improves. It is also clear that by relaxing a given pressure loss constraint and utilizing a catalyst that can survive at higher temperatures, the available operating space will broaden. A representative constraint of less than 5% total pressure loss is indicated by the black line and gray shaded area. The shaded area above a 1400 K catalyst temperature represents an approximate failure temperature for a catalyst layer.

The catalytic microcombustor operating space can be viewed more generally by plotting nondimensional parameters. **Figure 26** shows lines of constant combustor efficiency on a plot of Péclet number versus thermal efficiency. As Péclet number decreases, combustor efficiency increases due to the time available for diffusion of the reactant species to the active surface. However, as heat is lost from the system (lower thermal efficiency), the overall efficiency decreases.

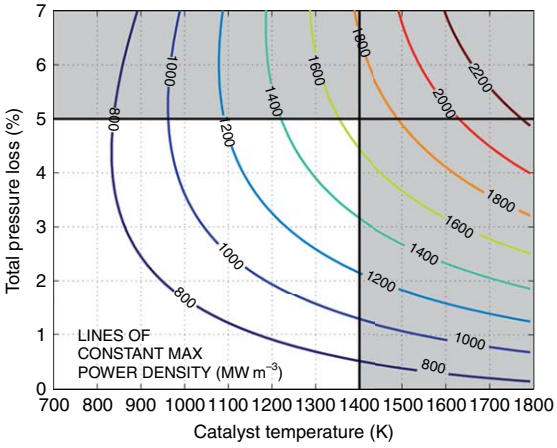


Figure 25 Operating space for a catalytic microcombustor; lines of constant power density. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

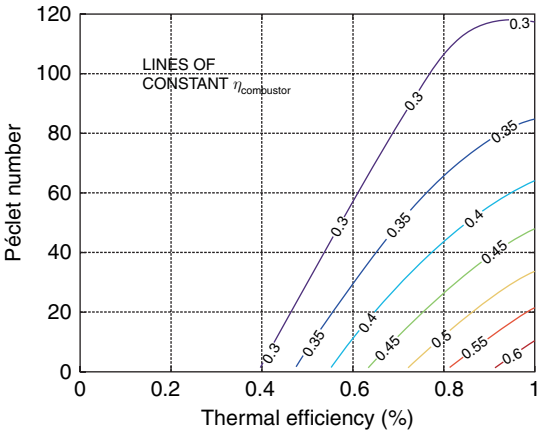


Figure 26 Nondimensional operating space: Péclet number versus thermal efficiency. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

3.15.5.7 Design Recommendations

Design recommendations for heterogeneous catalytic microcombustion systems include the following:

1. High power density catalytic microcombustors are typically diffusion controlled. A designer should seek to approach the high-temperature reaction-controlled regime to maximize performance. This can be achieved by implementing the following:
   a. Utilize the most thermally durable catalytic materials to achieve high-temperature operation.
   b. Utilize the highest surface area-to-volume ratio substrate material available that does not violate the system pressure loss constraint.
   c. Relax the total pressure loss constraint as far as the device performance will permit.
2. Searching for a more active catalytic material is not required unless
   a. the ignition transient is of concern or
   b. the overall design lies in the reaction-controlled regime.
3. High surface area-to-volume ratio catalyst materials result in increased heat loss from the system. As a result, thermal management is more critical than in the gas-phase case. A materials solution such as a thermal barrier combined with a concept similar to the recirculation jacket may mitigate these losses.

3.15.6 Applications – Microcombustors in Practice

There are a large set of applications for microcombustors, most of which involve small-scale power systems or propulsion. These systems can range from microscale heat engines to solid-state power systems and fuel cells. This section will provide a brief review of these applications.

3.15.6.1 Micro Heat Engines

A heat engine can be defined as a device that converts thermal energy into work. The thermal energy results from a temperature difference that is provided by a hot and a cold reservoir. The heat engine utilizes this difference in a thermodynamic cycle. Consequently, there are many types of cycles and heat engines such as a Brayton cycle (gas turbine engine) and an Otto cycle (a spark ignition piston engine). Many heat engines use a combustor as the source for the hot reservoir, and recently, there have been efforts to miniaturize these systems for portable power applications. Combustors for heat engines will be required to provide high-enthalpy gas to the system rather than lose heat through the structure.

The Brayton cycle, or the gas turbine engine, if miniaturized could be used both for portable power by integrating a generator with the system or as a turbojet propulsion device for a small aircraft (Epstein *et al.* 1997, Groshenry 1995). This type of device would require a microcombustor to be integrated with a rotating microscale compressor and turbine. This combustor would need to accept high-pressure air from the compressor and mix and burn fuel, resulting in a high-enthalpy fluid provided to the turbine for work extraction. It is important to note that a combustor for this system should provide this high-enthalpy gas with minimal thermal loss to the structure. An example of a micro gas turbine design is shown in **Figure 27**.

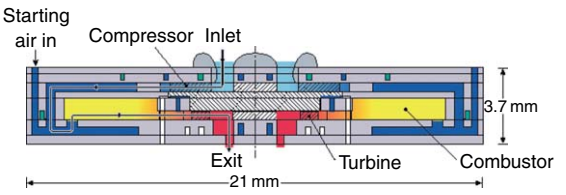


Figure 27 Baseline micro gas turbine engine schematic. (Source: Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology.)

Other microcombustor geometries for Brayton cycles, besides the radial layout shown in **Figure 27**, have been tested as well. These include conventionally machined, small-scale canister geometries. In these designs, combustion efficiencies over 99% have been obtained with hydrogen fuel in a 2-cm^3 chamber and with methane in a 15-cm^3 chamber. The canister devices were also shown to have reduced heat loss when compared with a similarly sized radial design (Isomura *et al.* 2002, Takahashi *et al.* 2002).

Microscale Otto cycle devices have also been studied. Early attempts at implementing these systems typically involved a free piston structure with a spark ignition system at either end of a cylinder. When ignition occurs, hot combustion gases expand, moving the piston to the other end of the cylinder, thus resulting in mechanical work. Again, this type of combustion system would require that enthalpy be retained in the gas and converted to work rather than lost through the structure (Lee *et al.* 2002, Park *et al.* 2002).

Another Otto cycle micro heat engine that has been under development is the micro rotary engine or the micro Wankel engine. In this type of device, there is no reciprocating component, but rather a rotating set of three combustion chambers. Typically, the rotor is triangular in shape and spins in an oval-shaped housing, forming three distinct volumes of gas. Due to the geometry of the rotor and housing, the volume of each chamber changes as the rotor spins. One of these chambers will have a minimum volume and a compressed gas. Fuel is then injected into this volume and ignited. The gas volume expands, causing the rotor to spin. Similar to the piston engine and the gas turbine, the combustor for the rotary engine provides high-enthalpy fluid to drive the system, with the thermal losses through the structure being the primary loss mechanism (Fernandez-Pello *et al.* 2002).

3.15.6.2 Heat Sources for Solid-State Power Generators and Endothermic Reactors

A microcombustor can provide heat for other types of power and chemical systems. Some of these solid-state devices include thermoelectric generators and thermophotovoltaic systems. Thermoelectrics (TEs) are solid-state devices that converts thermal energy into electrical energy by utilizing the Seebeck effect. TEs consist of a p- and n-type semiconductor material connected electrically in series and thermally in parallel. Carriers (electrons in n-type, holes in p-type) transport heat from the hot side of the device to the cold side, which can be used to drive an electrical load. A microcombustor for this type of system would have fundamentally different requirements than those for the fluid power systems discussed in Section 3.15.7.1. A TE device requires that heat be provided through the structure rather than being delivered via a high-enthalpy fluid. The primary loss mechanism is a hot exit gas. For this reason, a catalytic microcombustor that generates heat on the solid and can have a high rate of conduction through the structure is the ideal system (Arana 2003, Schaevitz 2000, Yoshida *et al.* 2006).

A thermophotovoltaic power system uses thermal radiation emitted from a hot structure and converts it to electrical energy with a photocell that is designed to absorb the infrared portion of the spectrum. Like the TE, this solid-state device requires a hot wall rather than a high-enthalpy fluid. Again, the most suitable combustor for this type of system is a catalytic combustor where heat is generated on a surface and conducted through the structure (DiMatteo *et al.* 2001, Wernsman *et al.* 2004).

One other type of system that benefits from heat conduction through the structure is an endothermic microreactor such as a methanol reformer for a fuel cell system. This type of device requires heat to be provided to its walls in order to drive an endothermic chemical reaction (often a heterogeneous catalytic reaction itself). A microcombustor is an ideal heat source for this application. The structures of a catalytic microcombustor and an endothermic microreactor can be thermally coupled. Heat will flow through the wall of the combustor and into the structure of the endothermic system (Park *et al.* 2006).

3.15.6.3 Micro Thrusters

Microscale rockets and thrusters for satellite attitude control, microsatellite repositioning, and even small-scale launch vehicles also may require microcombustors. These devices consist of a high-pressure combustion chamber connected to a nozzle for generating thrust. Fuels can range from liquid hydrocarbons to monopropellants such as hydrazine or hydrogen peroxide and even solid rocket propellants. These combustors do not utilize air as the oxidizer, but instead use stored oxygen. Liquid fueled rockets typically carry liquid oxygen to be injected with the fuel. Monopropellants have oxygen bound within the propellant molecule, and a catalytic reaction is used to break down and burn the propellant species. In solid fuels, the oxidizer is mixed with the fuel grain. In all

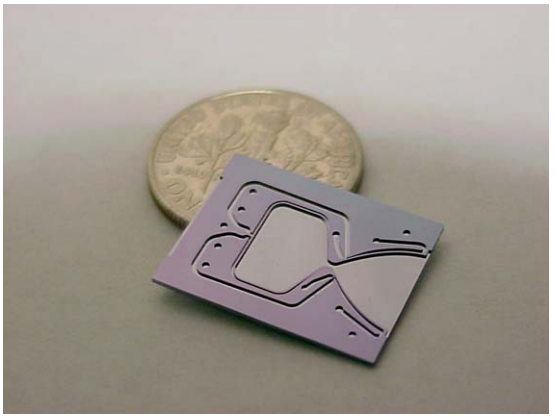


Figure 28 A silicon micromachined regeneratively cooled liquid bipropellant microrocket. (Source: London A P 2000 Development and test of a microfabricated bipropellant rocket engine. Ph.D. thesis, Massachusetts Institute of Technology.)

cases, the combustion process occurs at extremely high temperatures and pressures, resulting in unique requirements for the microcombustion chamber. Usually, a complex active cooling architecture in the structure is required. An example of a regeneratively cooled, liquid bipropellant microrocket is shown in Figure 28. Like the heat engine application, a microcombustor should provide high-enthalpy fluid, in this case to the nozzle for thrust (London 2000).

3.15.7 Future Perspectives

Microcombustors may have broad applications in power MEMS, micropropulsion, and microreactors as these fields mature. Although significant progress has been made in understanding combustion at the microscale, there is still work to be done. Microcombustors must become smaller and more efficient, providing high-enthalpy flow for heat engines and propulsion systems or hot surfaces for solid-state power generators and reactors. In order that these systems be practical, the combustor and its fuel reservoir must not be the limiting component.

To achieve the high power densities required by many applications, a hybrid gas-phase/catalytic microcombustor may be the optimal solution. In this embodiment, the combustor would consist of an initial low-temperature catalytic section that ignites the mixture and raises the temperature to some moderate level. The catalytic reaction would consume some of the fuel and air species but leave enough reactants for a subsequent gas-phase flame-based reaction. It would also raise the temperature to a level at which a gas-phase flame would autoignite, quickly burning the remaining fuel–air mixture and providing a high-enthalpy fluid in a minimum volume.

Other future areas of work will include developing microcombustors that operate with liquid hydrocarbon fuels such as gasoline, JP-8, and kerosene. This will require either injectors that atomize and mix the fuel or prevaporizors that convert the liquid fuel to the gas phase before injection into the combustion chamber. It also will likely require the use of a catalyst to increase reaction rates. Operating at high efficiency and power density with these more commonly used logistics fuels will result in broader applications of micropower systems with combustors.

Another advancement that would result in more practical combustion-based micropower systems is fuel flexibility. A microcombustor that can operate without significant performance change using various fuels could be deployed in scenarios where conventional fuel sources may be limited.

Microcombustors will likely be a part of many small-scale power systems in the future. Similarly, they will play a major role in providing heat to microchemical processors that may have an important industrial impact. Regardless of application, the principles presented here will provide a foundation for the design of microcombustors.

References

Arana L R 2003 High-temperature microfluidic systems for thermally-efficient fuel processing. Ph.D. thesis, Massachusetts Institute of Technology

DiMatteo R S, Greiff P, Finberg S L, Young-Waithe K A, Choy H K H, Masaki M M, Fonstad C G 2001 Enhanced photogeneration of carriers in a semiconductor via coupling across a nonisothermal nanoscale vacuum gap. *Appl. Phys. Lett.* **79(12)**, 1894–6

Epstein A H, Senturia S, Al-Midano O, Anathasuresh G, Ayon A, Breuer K, Chen K-S, Ehrich F E, Esteve E, Frechette L, Gauba G, Ghodssi R, Groshenry C, Jacobson S, Kerrebrock J L, Lang J H, Lin C-C, London A, Lopata J, Mehra A, Mur Mirandi J O, Nagle S, Orr D J, Piekos E, Schmidt M A, Shirley G, Spearing S M, Tan C S, Tzeng Y-S, Waitz I A 1997 Micro-heat engines, gas turbine, and rocket engines. *AIAA 97-1773, 28th AIAA Fluid Dynamics Conf., 4th Shear Flow Control Conference*, Snowmass Village, CO, USA

Fay J A 1994 *Introduction to Fluid Mechanics*. MIT Press, Cambridge, MA

Fernandez-Pello A C, Pisano A P, Fu K, Walther D, Knobloch A, Martinez F, Senesky M, Jones D, Stoldt C, Heppner J 2002 MEMS rotary engine power system. *Power MEMS 2002: Int. Workshop Power MEMS, Technical Digest*, Tsukuba, Japan, pp. 28–31

Goralski C T Jr., Schmidt L D 1996 Lean catalytic combustion of alkanes at short contact times. *Catal. Lett.* **42**, 15–20

Groshenry C 1995 Preliminary design study of a micro-gas turbine engine. Masters thesis, Massachusetts Institute of Technology

Hayes R E, Kolaczkowski S T 1997 *Introduction to Catalytic Combustion*. Gordon and Breach Science Publishers, Amsterdam, The Netherlands

Isomura K, Murayama M, Yamaguchi H, Ijichi N, Saji N, Shiga O, Tanaka S, Genda T, Hara M, Esashi M 2002 Component development of micromachined gas turbine generators. *Power MEMS 2002: Int. Workshop Power MEMS, Technical Digest*, Tsukuba, Japan, pp. 32–5

Lee D H, Park D, Yoon J, Kwon S, Yoon E 2002 Fabrication and test of a MEMS combustor and reciprocating device. *IEEE Micro Electro Mechanical Systems (MEMS) Conf. 2002, Tech. Digest*, Las Vegas, NV, USA, pp. 272–5

Lee J-W 2000 Numerical simulation of a hydrogen microcombustor. M.S. thesis, Massachusetts Institute of Technology.

London A P 2000 Development and test of a microfabricated bipropellant rocket engine. Ph.D. thesis, Massachusetts Institute of Technology

Maskell W C, Sammes N M, Steele B C H 1987 Agglomeration of thin platinum films on a zirconia substrate. *J. Phys. D Appl. Phys.* **20**, 99–104

Mehra A 2000 Development of a high power density combustion system for a silicon micro gas turbine engine. Ph.D. thesis, Massachusetts Institute of Technology

Mehra A, Zhang X, Ayon A A, Waitz I A, Schmidt M A, Spadaccini C M 2000 A six-wafer combustion system for a silicon micro gas turbine engine. *IEEE/ASME J. Microelectromech. Syst.* **9(4)**, 517–27

Park D, Lee D H, Yoon J, Kwon S, Yoon E 2002 Micro reciprocating engine as a power source for microsystems. *Power MEMS 2002: Int. Workshop Power MEMS, Technical Digest*, Tsukuba, Japan, pp. 40–3

Park H G, Malen J A, Piggot T P, Morse J D, Grief R Grigoropoulos CP, Havstad M A, Upadhye R 2006 Methanol steam reformer on a silicon wafer. *J. Microelectromech. Syst.* **15(4)**, 976–985

Schaevitz S B 2000 A MEMS thermoelectric generator. Masters thesis, Massachusetts Institute of Technology

Spadaccini C M 2004 Combustion systems for power-MEMS applications. Ph.D. thesis, Massachusetts Institute of Technology

Spadaccini C M, Mehra A, Lee J, Zhang X, Lukachko S, Waitz I A 2003 High power density silicon combustion systems for micro gas turbine engines. *ASME J. Eng. Gas Turbine Power* **125**, 709–19

Spadaccini C M, Peck J, Waitz I A 2007 Catalytic combustion systems for micro-scale gas turbine engines. *ASME J. Eng. Gas Turbines Power* **129**, 49–60

Takahashi K, Murayama M, Isomura K, Tanaka S 2002 Development of a methane fueled combustor for micro-scaled gas turbine. *Power MEMS 2002: Int. Workshop Power MEMS, Technical Digest*, Tsukuba, Japan, pp. 44–6

Veser G, Schmidt L D 1996 Ignition and extinction in catalytic oxidation of hydrocarbons. *AIChE J.* **42**, 1077–87

Waitz I A, Gauba G, Tzeng Y-S 1998 Combustors for micro gas turbine engines. *ASME J. Fluids Eng.* **20**, 109–17

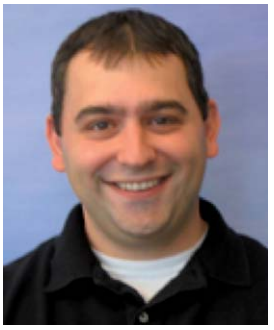
Wang J, Zhao D G, Sun Y P, Duan L H, Wang Y T, Zhang S M, Yank H, Zhou S, Wu M 2003 Thermal annealing behaviour of Pt on n-GaN schottky contacts. *J. Phys. D Appl. Phys.* **36**, 1018–22

Wernsman B, Siergiej R R, Link S D, Mahorter R G, Palmisiano M N, Wehrer R J, Shultz R W, Schmuck G P, Messham R L, Murray S, Murray C S, Newman F, Taylor D, DePoy D M, Rahmlow T 2004 Greater than 20% radiant heat conversion efficiency of a thermophotovoltaic radiator/module system using reflective spectral control. *IEEE Trans. Eelctron Devices* **51(3)**, 512–15

Williams W R, Stenzel M T, Song X, Schmidt L D 1991 Bifurcation behavior in homogeneous–heterogeneous combustion: I. Experimental results over platinum. *Combust. Flame* **84**, 277–91

Yoshida K, Tanaka S, Tomonari S, Satoh D, Esashi M 2006 High-energy density miniature thermoelectric generator using catalytic combustion. *J. Microelectromech. Syst.* **15**, 195–203

Biographies



Christopher M. Spadaccini received a Bachelor of Science and Masters of Science from the Department of Aeronautics and Astronautics at the Massachusetts Institute of Technology (MIT) in 1997 and 1999, respectively. In 2004 he received a Ph.D. from the Department of Aeronautics and Astronautics at MIT for his work on "Combustion Systems for Power-MEMS Applications." Dr. Spadaccini is currently a staff research engineer at the US Department of Energy's Lawrence Livermore National Laboratory (LLNL). As a member of the Center for Micro- and Nanotechnology at LLNL, he pursues research in the areas of micro- and nanoscale power systems, microcombustion, fluid mechanics and microfluidics, heat transfer and thermodynamics, as well as new nanostructured materials and novel microfabrication techniques. He has written over a dozen conference and journal publications and is a part-time faculty member in the Chemical and Materials Engineering Department at the San Jose State University. Dr. Spadaccini's work is applied to a broad range of problems including national security, energy, and environmental issues.

Ian A. Waitz received a Bachelor of Science in Aerospace Engineering from the Pennsylvania State University in 1986. In 1988 he received a Masters of Science in Aeronautics from George Washington University, and in 1991 he completed his Ph.D. in Aeronautics at the California Institute of Technology. Currently he is a professor in the MIT Department of Aeronautics and Astronautics and Director of the Partnership for AiR Transportation Noise and Emissions Reduction (PARTNER), an FAA/NASA/Transport Canada-sponsored Center of Excellence. His principal fields of interest include propulsion, fluid mechanics, thermodynamics, reacting flows, aeroacoustics, and, in particular, aspects of the above that relate to environmental issues associated with aircraft design and operation. Professor Waitz pursues experimental and computational research in these areas. He has written approximately 60 technical publications including a report to the US Congress on aviation and the environment, holds three patents, and has consulted for many organizations. During 2002–2005 he was Deputy Head of the Department of Aeronautics and Astronautics. He was an associate editor of the AIAA *Journal of Propulsion and Power*. In 2003 Professor Waitz received a NASA Turning Goals Into Reality Award for Noise Reduction. He is a fellow of the AIAA, and an ASME and ASEE member. He teaches graduate and undergraduate courses in the fields of thermodynamics and energy conversion, propulsion, fluid mechanics, and environmental effects of aircraft. He was honored with the 2002 Class of 1960 Innovation in Education Award and appointment as a MacVicar Faculty Fellow in 2003.

3.16 Micro-Coolers

Junghoon Yeom and Mark A. Shannon, Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Nomenclature

A	surface area where heat transfer occurs
A_{eff}	cross-sectional area of an MHP
A_{surf}	project geometrical area where boiling occurs
c_p	intensive heat capacity or specific heat
d	fan diameter
h	overall heat transfer coefficient
h_{boil}	boiling heat transfer coefficient
k_{eff}	effective thermal conductivity of an MHP
L_{eff}	effective length between the evaporator and the condenser in an MHP
q	heat flux removed
q_{CHF}	critical heat flux – maximum heat flux in the nucleate boiling regime
R_{th}	effective thermal resistance
S	Seebeck coefficient
T	absolute temperature
ΔT	temperature difference between the surface to be cooled and the ambient

ΔT_{sat}	wall superheat ($=T_{wall}-T_{sat}$)	ρc_p	mass specific capacity
ZT	nondimensional figure of merit	σ	electrical conductivity of thermoelectric material
λ	thermal conductivity of thermoelectric material		

Glossary

CHF Critical Heat Flux
COP Coefficient of Performance
CPL Capillary Pumped Loop
DRIE Deep Reactive Ion Etching
EDIFICE Embedded Droplet Impingement for Integrated Cooling of Electronics
EDM Electrical Discharge Machining
EHD Electrohydrodynamic
HIT Heterostructure Integrated Thermionic
HP Heat Pipe
J–T Joule–Thompson
L–H Linde–Hampson
LHP Loop Heat Pipe
LTS Loop Thermosyphon
MEMS Microelectromechanical Systems
MFD Magnetofluid Dynamic
MHP Micro Heat Pipe
μHEX Microchannel Heat Exchanger
TEC Thermoelectric Cooler
TIM Thermal Interface Material
VCM Variable Capacitance Motor
VIDA Vibration-Induced Droplet Atomization

3.16.1 Introduction

Small to microscale cooling systems exploit an extraordinary number of cooling modalities, ranging from passive to active cooling, air to liquid to solid-state media. These systems are manufactured with an array of technologies including traditional copper machining as well as silicon micromachining for integrated cooling. To begin our discussion of microscale cooling, we need to start with a working definition of a microcooler, which encompasses most of, but not all, the cooling modalities encountered. Miniature systems typically consist of both microscale and mesoscale systems. Mesoscale systems are somewhat larger, on the order of centimeters, with a maximum dimension not to exceed, say, 5 cm. For this chapter, we define microcoolers as devices where the critical heat transfer through the system being cooled occurs at dimensions less than a few hundred microns and where the overall dimension of the system does not exceed a few square centimeters of a planform area. While one may note that the former constraint does not limit the number of devices much, since the heat energy transferred is often limited at interfaces and boundary layers that are less than hundreds of microns, when combined with the second constraint, the task of defining a microcooler becomes easier. The second constraint simply limits the total number of systems considered to a manageable number, while keeping the dimensions small enough that adapting normal or even mesoscale cooling systems to work at the scale becomes a significant technical challenge and is consistent with the definition of microscale cooling by Phelan *et al.* (2001, 2002) as those whose maximum dimensions are on the order of millimeters, such that in principle, they can be incorporated directly at chip level in electronic heat transfer applications to provide localized cooling.

Most microcooling technologies deal with heat transfer from a hotter to a cooler body, and fewer deal with moving heat energy up a temperature gradient or cooling below ambient temperature. Our intent is to review the current state-of-the-art mechanical microscale cooling fundamentals and devices and to discuss the strengths and shortcomings of each. Out of the huge number of cooling technologies, we will also focus on microcooler systems that can be miniaturized and incorporated into packaging at the chip, substrate, or board level, which can be integrated into systems for electronic cooling, and/or lightweight and portable systems used in military or medical applications. We should also note that there are a tremendous number of articles on numerical simulation of microcoolers to better understand and optimize the systems. However, a review of analytical and computation modeling related to microcooling technologies is beyond the scope of this chapter.

The objectives of this chapter are to

(1) critically review available small-scale microcoolers to discuss the potentials and challenges of both conventional and emerging cooling technologies;
(2) assess and suggest future microcooling technologies for two major applications of electronics cooling and portable coolers for military/medical uses;
(3) address the question of whether we can apply some of the well-developed technologies for electronics cooling to less developed portable applications.

We will also discuss how the development of MEMS technology relates to microcooler technology via (McGlen 2004, Munkejord *et al.* 2002, Perret *et al.* 2002, Punch 2005)

(1) integration of microcoolers into silicon chips and the thermomechanical issues that arise when different materials with different CTEs are used at the interface;
(2) advancements that microfluidic devices such as pumps, valves, and microchannels have on microcooler technologies, and some of the remaining outstanding issues.

3.16.1.1 Applications of Microcoolers

Microcooling technologies for electronics cooling have been actively researched and developed in both industry and university research. More than 95% of the papers, proceedings, and proposals address the issues related to electronics cooling. Much less written work has been publicized about portable microcoolers for military and medical purposes, which indicate that most of the ideas and research are still formative and ongoing for other applications.

3.16.1.1.1 Electronics cooling

One might wonder why so much effort has been, and continues to be, expended on developing microcoolers for electronics heat transfer. Current methods of heat removal involving extended surfaces and fan arrays that are frequently used in electronics become insufficient as more powerful electronic components dissipate more heat while the space around these components continues to be reduced due to miniaturization trends of electronic systems (Garimella *et al.* 2002). Exacerbating these macroscopic trends, at the level of the electronic chip, the ever-higher densities of transistors and dissipative elements fabricated on silicon are being interconnected with multiple layers of thin-film structures of dielectrics and conductors, which act to trap heat and slow heat transfer from the active elements. The thermal conductivity of thin films can be one or two orders of magnitude lower than that of the bulk material, so even small increases in power dissipation in a thin-film region may cause serious local heating in microelectronic devices. Many of the important parameters governing the operation of active elements, such as electron and hole mobility, dopant migration, and Schottky barrier, are strongly temperature dependent, and even a few degrees increase can cause errors or premature device failure. In addition, the trend to reduce operating voltages in digital systems reduces the total voltage differential separating a "1" from a "0." Thus, thermal voltage fluctuations are becoming more significant for proper digital system operation. In many instances the standard techniques used today cannot achieve the required cooling performance due to physical limitations in heat transfer capabilities. These limitations are principally related to the low thermal conductivity of air for convection, and the maximum amount of heat that can be conducted by the materials used in the electronic systems.

Figure 1 shows a comparison of various cooling techniques as a function of the attainable heat transfer in terms of the heat transfer coefficient. The heat transfer coefficient of natural cooling in air, relying on both free convection and radiation, is around $10\,\mathrm{W\,m^{-2}\,K}$ ($10^{-3}\,\mathrm{W\,cm^{-2}\,K}$), which limits heat flux removed to air below $0.05\,\mathrm{W\,cm^{-2}}$ at the cooler–air interface. In order to successfully remove heat fluxes at the chip–cooler interface that are expected to be in excess of $120\,\mathrm{W\,cm^{-2}}$ by 2010, and $180\,\mathrm{W\,cm^{-2}}$ by 2018, according to the International Electronics Manufacturing Initiative (iNEMI Report 2005), a very large heat sink or an unacceptably large chip-to-ambient temperature difference would be needed. To accommodate a heat flux of more than $100\,\mathrm{W\,cm^{-2}}$, at a temperature difference of 50 K, requires an effective heat transfer coefficient of $2\,\mathrm{W\,cm^{-2}\,K}$, which is more than three orders of magnitude higher than natural convection to the chip itself can supply. Moreover, it is becoming higher than even forced convection over an enlarged area of an aluminum or copper heat sink can potentially provide. For instance, even with a 100 to 1 area increase and $150\,\mathrm{W\,m^{-2}\,K^{-1}}$ forced heat transfer

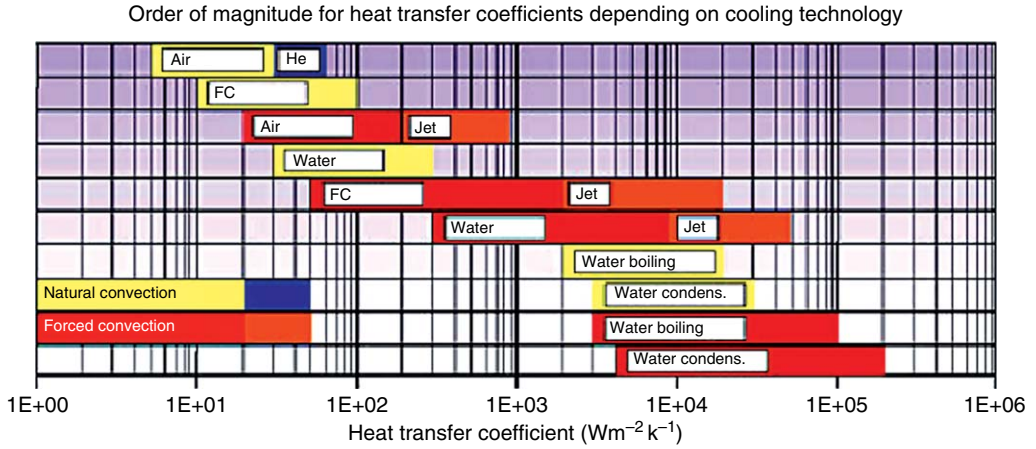


Figure 1 Heat transfer coefficient attainable with natural convection, single-phase liquid forced convection and boiling for different coolants. (Source: Lasance C 1997 Technical data column. *Electronics Cooling* **3(1)**, http://www.electronics-cooling.com/articles/).

coefficient from **Figure 1**, an isothermal (infinite thermal conductivity) heat sink does not quite reach what is needed. From **Figure 1**, it appears that there will be a need for liquid cooling or other heat transfer enhancement, such as microcoolers, in the future of thermal management. The majority of this review will be dedicated to discussing a number of promising thermal management technologies that are emerging for electronics applications. A particular emphasis is placed on so-called on-chip cooling, a result of the integration of a microcooler into the electronic devices to be cooled. Recently, optoelectronic devices that generate significant amounts of heat (100–1000 W cm^{-2}) and often require an operating temperature below the room temperature for better performance need active cooling, which has led to the use of compact refrigeration systems (Burger 2001, Little 1984). Nascent active microcooling system research will also be touched on in this review to meet these subambient temperature cooling applications.

3.16.1.1.2 Microclimate cooling

Protective garments have been proposed to cool military personnel on active duty in hot desert climates or other hazardous environments where personal microclimate control is required. Sealing off military personnel from the surrounding environment with a protective suit made of impermeable layers of material is the only effective barrier to many nuclear, biological, and chemical weapons. However, sealing off the body from transfer of gases and perspiration for any substantial time raises potentially life-threatening problems of heat stroke that needs enhanced cooling. A number of other nonmilitary applications also offer potential application for the microclimate controller fabric, such as use by firefighters, astronauts, surgeons operating in elevated temperature rooms, and users in private sector chemical and industrial hazardous operating environments. There are many nonpersonal cooling applications such as temperature-sensitive package shipping, individual beverage cooling, glove box and car seat cooling, and even thermal signature management. There has been limited research and development of lightweight portable microclimate cooling, which will be discussed briefly in this chapter.

3.16.1.1.3 Drinking water cooler

Maintaining adequate hydration levels for combat soldiers during field operation in hot climates is critical to an individual's performance and even their lives. Under exertion, 1–3 liter of water per hour is lost through evaporation that needs to be replenished. Similarly, outdoor enthusiasts, e.g., hiking, biking, and engaged in sports face similar water demands. Soldiers carry their water supplies on their backs in an ergonomically designed bladder-type reservoir commonly referred to as the On-the-Move Hydration System, and similar systems are available for civilian uses. Although some of the commercial products are designed to provide a means for soldier hydration during operation in extreme cold environments, no system is yet built to offer an active cooling to drinking water warmer

than body temperature in the hot desert environments. Supplying cooled water to soldiers and people under high exertion in hot climates is important for two main reasons. First, drinking cool water lowers body temperature, preventing heat-related injuries, such as heat stroke. Second, people have difficulty consuming large quantities of hot water, which can reach 50°C in desert climates. Therefore, a lightweight, portable, active water cooler unit needs to be developed to cool drinking water below body temperature.

3.16.2 Challenges Related to Microcooling Technologies

In electronics cooling, several challenges exist: (1) remove ever-increasing heat fluxes approaching the flux at the surface of the sun; (2) scale down the overall system to be compact and compatible with the existing electronics; (3) transfer heat to air (ultimately), which has low heat capacity; (4) spot removal of heat energy where the Biot number of the system is small; (5) nonlinear heat removal due to the addition of energy at the junction to be cooled; and (6) low enough cost and high enough reliability to be successful in the marketplace.

For military and medical applications, the issues are tightly related to whether or not the device to be cooled is portable and/or needs to be lightweight and small in volume. The challenges include (1) cooling energy density and power density, providing an overall cooler system (including unit, coolant, and battery if required) that is lightweight and small enough to be portable, which is related to scaling; (2) a cooler that has the desired cooling rate at the temperature that it needs to operate at; (3) energy consumption that does not need high-voltage source or complex electronics; (4) robust operation; and (5) cost.

Microcooler technologies can, in a broader sense, be categorized into two areas: refrigeration cooling and nonrefrigeration cooling (Phelan *et al.* 2001). Refrigeration cooling refers to a system that can lower the target temperature (e.g., the junction temperature in the electronic chip) below the ambient temperature or at least below the temperature that a nonrefrigeration system can accomplish. Since nonrefrigeration cooling uses natural/forced convection to air or liquid and/or the phase change of a liquid, the lowest possible temperature will be limited by the temperature of the medium (air or liquid) used for the low-temperature reservoir, i.e., the ambient temperature. Therefore, nonrefrigeration cooling is useful only for cooling devices hotter than the ambient. It will not, however, be effective to lower the temperature of human bodies or drinking water exposed to hot ambient conditions.

Central to our discussion of microcoolers for the above applications is the ability to remove excess heat from the device. The heat flux carried away from the device being cooled, given simply by $q = hA\Delta T$, is proportional to the heat conductance term (hA), a product of the heat transfer coefficient and the surface area where heat transfer occurs, and the temperature difference between the surface of the device and the environment to which the heat is being rejected. Often, the temperature of the low-temperature reservoir is higher than the surrounding ambient temperature, such as a device rejecting heat within a computer case, in which numerous other heat-producing devices heat up the air within the case. Another example is within tanks and personnel carriers, where significant amounts of heat are generated by weapons, communications, and control systems, which often exceed the ability of air conditioning systems to remove, particularly in hot ambient desert environments. Similarly, for new high-performance military aircraft, the design is being driven more and more by the need to remove heat energy produced by increasing high-powered avionics packages.

In electronics cooling, the surrounding low-temperature rejection reservoir is typically lower than the chip heat rejection surface temperature, allowing heat to be spontaneously transferred out of the chip without supplying active refrigeration cooling. Due to the specification of the electrical components, there is often a maximum allowable temperature at which the device should be operated, capping the maximum attainable ΔT (a typical allowable temperature difference of 60°C between the component surface and the ambient). For consumer electronics such as computer processors, often an 85°C maximum chip case temperature is specified, with the intent of holding the silicon device temperature to less than 125°C. Above this device temperature, significant degradation in performance is often noted for silicon devices. Military and special device specifications often exceed 125°C, and newer high-temperature transistors are being actively developed to significantly raise the maximum operating temperature in order to lessen the relatively stringent absolute temperature cooling systems must maintain. However, the trend toward higher device densities

combined with a higher-frequency operation well into radio and microwave frequencies both act to increase the heat flux that must be removed, regardless of the specific maximum operating temperature. In addition, since electronic devices themselves are being packed with higher densities within the overall system, higher rejection temperatures available due to higher permissible device temperature operation act to increase the overall temperature reservoir to which the system's devices reject to. Therefore, the goals in designing the thermal management systems have been to (1) increase the heat exchange surface area as large as possible and/or (2) increase the overall heat transfer coefficient.

As will be discussed, increasing the surface exchange area is limited by the thermal conductance of the heat sink/exchange system being employed and the total size and weight permitted. Thus, increasing the surface area using normal-scale electronic cooling systems is rapidly reaching practical limits. Similarly, increasing the overall heat transfer coefficient, which is typically proportional to the air velocity raised to a power less than one, by increasing cooling air speed is limited by fan power consumption, which is proportional to the velocity squared. From a system perspective, at some point for any given system, adding more power, even if it is available (which is decidedly not true for portable systems), will potentially generate more heat than is removed. However, both the heat exchange area and the heat transfer coefficient are also coupled together through characteristic length scales of the heat transfer and system. As will be discussed, when the phenomenological length scale associated with the heat transfer, such as the fluid boundary layer and the thermal diffusion length, are on the same order as the physical length scale, enhancements in the total heat transferred are possible. Importantly for microcoolers, several key phenomenological length scales are in the submilliter range, or well within the microscale. Therefore, microscale passive cooling technologies have sprung out of research to address both the goals of thermal management systems.

While nonrefrigeration microscale cooling enhancement has dominated microcooler research and development thus far, there has also been significant, but nascent research in microscale refrigeration cooling. Advances in microscale passive cooling can certainly be helpful but is not sufficient when the ambient temperature is hotter than the target temperature of the system, such as with many optoelectronic applications, and microclimate or drinking water cooling. Active cooling capacity by the addition of energy is needed. Many different types of active cooling technologies have been proposed, both solid-state and mechanical. Solid-state thermoelectric coolers have received the bulk of the research, since the method has been well known and actively developed over several decades, and many thermoelectric coolers have been commercialized for cooling numerous electronic systems, including sensitive low-temperature optical sensors. Since the active heat pumping mechanism is the transport of hot charge carriers across a semiconductor junction, solid-state coolers have the potential to be scaled down to the microscale. Similar thermoionic coolers also have this potential. Integrating solid-state coolers with microfluidic heat exchangers that are being developed for nonrefrigeration microcoolers offers the potential to provide the subambient temperature cooling needed. As will be discussed, the main issues thus far with solid-state cooling are the low coefficient of performance (COP) of solid-state coolers. The COP is the ratio of heat transferred to the power supplied to move the heat. Current fully integrated solid-state microscale systems have COPs less than one, which leads to difficulties with operating temperature regimes and thermal runaway. However, a great deal of research is being expended to increase the thermal efficiency of solid-state coolers, such that the COP of the system will potentially be far above 1, approaching that of large-scale mechanical refrigeration systems. We will briefly discuss some of the ongoing research, particularly related to fabricating micro- to nanoscale thermoelectric elements.

Due to the inherently high COP of normal-scale mechanical vapor compression refrigeration systems, which can be as high as 4–8 depending on the temperature lift and absolute temperature, there has been some research into miniaturizing mechanical refrigerators to the meso- and microscales, as will be briefly discussed. However, meso- to microrefrigeration mechanical cooling technologies for electronics cooling have not yet been successfully commercialized, to the author's knowledge. Mechanical refrigeration cooling for portable devices used in military or medical applications is even further from being realized. While mechanical microcoolers also benefit greatly from the passive microcooler technology developed thus far, the main reason for the lack of adoption of mechanical vapor compression refrigeration is the lack of reliable and/or low-energy consumption gas microcompressors. Vapor

compression cycles work by compressing a refrigerant to a higher pressure, increasing the temperature of the vapor. The vapor condenses into a liquid as heat is rejected to the ambient-temperature heat reservoir. On expansion to a lower pressure in the evaporator, the liquid vaporizes and absorbs heat at the low temperature, thereby providing cooling power. Compression ratios greater than 1 are often required for most potential refrigerants. The problem is that the ability to compress gas does not typically scale well as dimensions are shrunk, particularly to the microscale. While there exist microscale methods to achieve high compression ratios, as will be discussed, the process is highly inefficient, again reducing the overall COP to well below 1. We will discuss the state of fluid compressors development and the open questions that remain to be resolved and other aspects of vapor compression microcoolers.

3.16.3 Microcooling Technologies with Related Principles

As noted earlier, nonrefrigeration cooling is only effective in lowering the temperature of a surface much hotter than ambient and therefore has been exclusively developed for electronics cooling. In fact, most of the existing cooling solutions for microelectronics equipment are based on nonrefrigeration cooling. Discussion on nonrefrigeration cooling, therefore, is limited to the application of electronics cooling. Refrigeration cooling also must transfer heat ultimately to an ambient thermal reservoir, so the techniques discussed for electronic cooling are for the most part completely applicable to active systems as well, with respect to condensers, evaporators, and heat rejection heat exchangers and fans. The first system we consider is the use of the mesoscale fan to induce the convective flow around the heat sink, which is known to reach its limit for cooling of high heat flux electronics. Major research efforts have also been, and continue to be, focused on liquid cooling (using a single-phase or two-phase) that shows commercial viability, both for the present and for the near future. A variety of miniature heat pipes, for instance, can be ubiquitously found in many laptop computers as heat spreaders to provide higher cooling power than microfin-based copper sinks, which will be presented. Finally, active systems will be briefly presented: solid-state thermoelectric, electrokinetic, and electrostatic fluid pumps.

3.16.3.1 Air Cooling

Even though the standard air-cooling technique using a fan with a heat sink appears to reach its limit for cooling of high-power electronics, the idea of using ambient air as a coolant will remain valid and attractive for several reasons. Most of the liquid cooling techniques, including heat pipes and microchannel heat exchangers, play a role as heat spreaders and still rely on air cooling for their terminal dissipation to the surrounding air. Moreover, cooling small portable electronic devices requires a low profile, so direct air-cooling techniques may remain viable when other cooling techniques that use liquid pumping or refrigeration cycles may be too large or energetic to satisfy constraints placed on the system. A heat transfer coefficient based on pure natural convection ranges from 1 to $10\,W\,m^{-2}\,K$ while forced convection with a standard rotary fan can push a heat transfer coefficient up to $150\,W\,m^{-2}\,K$ with acceptable noise levels (Lasance and Simons 2005). A recent review article describes the limitations and key challenges of conventional air-cooling technologies (Rodgers *et al.* 2005). Innovative schemes enhancing air movement have emerged in smaller scales to extend the useful range of air cooling by enhancing hA. We will present recent developments on microfans, piezoelectric fans, and microjet cooling.

3.16.3.1.1 Microfans

The seemingly endless pursuit to miniaturize consumer electronics and pack more functionality into them has led to the commercialization of mesoscale (order of 10 cm) fans to cool high-powered chips. However, there are no commercial fans currently available that are small enough to cool very small, portable, high-powered electronic devices. Laptops now use small axial fans to cool microprocessors with integrated heat sinks attached to the chip casing. However, these fan and heat sink combinations are not sufficient to cool more powerful microprocessors that are used in desktop systems that require more powerful fans, heat pipes, and even active cooling systems to enable their use. The miniaturization effort for axial flow fans faces limitations due to two main reasons: difficulty in fabrication and an unfavorable trend in the aerodynamic performance of the fan. Grimes *et al.* (2003) investigated the issues of scaling of microfans by fabricating a series of geometrically similar 3D microaxial flow fans with representative sizes (120, 40, and 6 mm of fan diameters, d) and quantifying the performance of the

fans in terms of entropy generation. They used micro electrodischarge machining (μEDM) to fabricate the smallest fan tested that had a 6-mm diameter and a 25-μm blade tip clearance. They did not use silicon MEMS technology, with which it is difficult to produce 3D shape of the blades.

In general, a reduction in size of a fan results in a reduction in its efficiency, which is caused by a relative increase in entropy generation (Massey 1983). Further analysis showed that there is a larger increase in the local entropy generation rate as the diameter is reduced ($\sim d^{-1/2}$) in the laminar boundary layer than the entropy generation rate in the turbulent boundary layer, which is only weakly inversely dependent on the length scale ($\sim d^{-1/7}$) (Grimes *et al.* 2003). Since laminar flow dominates with microscale fans, their efficiency and performance become poorer as the diameter of the fan is reduced. Recently, the same group from the University of Limerick reported similar aerodynamic disadvantages for micro radial flow fans fabricated by stereolithography as the fan size is reduced below 10 mm (Hanly *et al.* 2005). They also investigated the integration of the microfans into low-profile portable electronics (Walsh and Grimes 2006). **Figure 2** shows a micro radial flow fan made using stereolithography, and the copper heat sink with a center hole that the fan is inserted into. On spinning, cooler air is drawn by the fan across the evenly spaced radial fins, and the subsequently heated air is exhausted.

Along with the penalty that smaller fans run at a lower efficiency than larger fans, acoustic noise generated by a microfan is another important issue. As the heat flux from the electronic devices increases, the fan has to run at a higher rotational speed, and vibrational noise is typically proportional to the harmonics of rotational speed and the tip-to-edge distance per unit volumetric flow rates. Typically, if the radius of the radial fan is decreased one order of magnitude, the rotational speed needs to increase one order of magnitude to maintain the same volumetric flow rate per unit of planform area. Also, it is often difficult to proportionally decrease the tip-to-edge distance when reducing size. Therefore, both the rotational speed and the bypass tip flow rate, which is a major source of noise and losses, are higher. Of course, the smaller size reduces the overall magnitude. However, it can be hard to control the noise level, particularly at higher frequencies. Similarly, the power consumption per unit volumetric flow rate also typically increases with the rotational frequency squared. Thus, smaller-size systems may consume more power for a given cooling power needed.

3.16.3.1.2 Piezoelectric fans

Radial microfans are not the only type of fan being developed. Piezoelectric bimorph-type fans have been proposed and initial development conducted to provide a small, low-noise, low-power devices in thermal management systems for supplemental cooling in hot spots and other stagnant areas of small portable electronics (such as cell phones) where a rotary fan is ineffective (Bürmann *et al.* 2003). A piezoelectric fan utilizes a piezoceramic patch or several patches bonded to a thin flexible blade (shim material) that flaps back and forth due to an alternating voltage applied to both sides of the piezoelectric patches. **Figure 3** shows a schematic of a piezoelectric fan and the electrical driver circuit. The fan produces a streaming airflow in a direction transverse to the blade with vortices shed off from its

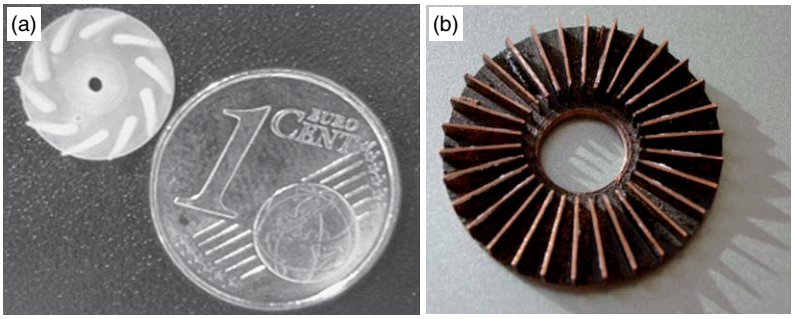


Figure 2 Photographs of (a) a fan and its size relative to a one cent Euro coin and (b) a manufactured heat sink with 30 fins whose thickness is 0.5 mm. (Source: Walsh E, Grimes R 2006 The development of an integrated fan and heat sink solution for thermal management in low profile applications. *Proc. ASME ICNMM2006 4th Int. Conf. Nanochannels, Microchannels, Minichannels*, Limerick, Ireland, June 19–21, 2006).

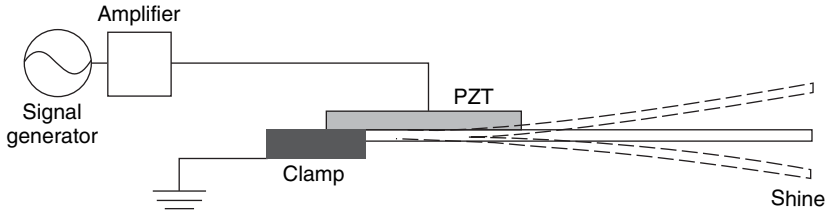


Figure 3 A schematic of a piezoelectric fan and driver system. (Copyright 2004 from Experimental investigation of the thermal performance of piezoelectric fans. *Heat Transf. Eng.* **26(1)**, 4–14 by Açikalin T, Wait S M, Garimella S V, Raman A. Reproduced by permission of Taylor & Francis Group, LLC., http://www.taylorandfrancis.com)

tips. The main features that make the piezoelectric fans well suited for cooling of portable electronic applications are the low acoustic noise levels (via tuning the first mode of resonance outside the range of frequencies audible to the human ear) and low power consumption (via driving the fan at resonant frequency) (Açikalin *et al.* 2004). Of course, air or another fluid will act to damp out the resonance; thus, the design must ensure that the structure is not critically damped or overdamped to exploit resonance driving.

Since Toda first introduced the piezoelectric fans for electronics cooling in the late 1970s (Toda 1979), numerous researchers developed different types of piezoelectric fans and demonstrated the enhanced heat and mass transfer rates for the system interested (Loh *et al.* 2002, Schmidt 1994, Yoo *et al.* 2000). Flow visualization together with the simulation using analytical and computational models has been used to optimize the location of the fans with respect to the cooled surface (Açikalin *et al.* 2004, Bürmann *et al.* 2003, Schmidt 1994). The Garimella group built prototypes of the fans, tested them in enclosures with dimensions similar to a cell phone and a laptop computer, and reported up to a 100% enhancement over natural convection heat transfer (Açikalin *et al.* 2004).

Typical sizes of the fans investigated thus far are rather large, their widths ranging from 1 to 3.5 cm, their lengths from 6.5 to 12.8 cm, which put their overall dimension a bit outside our definition of the microcooler. For a given thickness of the fan (usually tens of microns), smaller sizes of the fans are desired, but their resonant frequency may fall into an audible range. More importantly, high operating voltages (40–140 V) and the complex circuitry needed for the AC power supply have limited this device from being integrated into portable electronic devices, particularly those sensitive to high-voltage-induced noise.

3.16.3.1.3 Microjet cooling

A forced air convection cooling technique using an impinging jet of air is well known and can be found in many industries because of its high heat transfer coefficients with relatively low pressure drops. Martin (1977) summarized early research in macroscale jet impingement heat transfer where the focus was on optimizing the transport processes associated with steady impinging jets. Since this time, it has been proposed to use a single impinging jet to spot-cool critical components in electronic devices, and an array of impinging jets to cool entire circuit boards. Numerous theoretical and numerical computational studies have dealt with microjet impingement cooling of microelectronics (Boeschoten *et al.* 2000, Chiriac *et al.* 2001, Chiriac and Rosales 2003). It should be noted, however, that microjet refers to the size of the jet impinging on the surface, not the overall dimension of the device. These studies commonly report cooling enhancement with the flow transition from laminar to turbulent and from steady to unsteady. Also, the nozzle-to-substrate distance, the geometry of individual nozzles, and array spacing of the nozzles are important design parameters.

Brignoni and Garimella (1999) built a macroscale test setup to experimentally optimize the performance of confined impinging air jets used in conjunction with a pin-fin heat sink. They found that the heat transfer rates are weakly dependent on a nozzle-to-target spacing and are more influenced by the geometry of the impinging substrate and a nozzle diameter for a given Reynolds number. Wu *et al.* (1999) developed a micro impinging-jet heat exchanger integrated with a 8×8 MEMS temperature sensor and heater to facilitate a 2D surface temperature measurements. The MEMS nozzle arrays were made using MEMS fabrication processes, such as photolithographic patterning, KOH etching of bulking silicon, and fusion bonding (**Figure 4**). Using a single microjet (500 μm diameter driven by

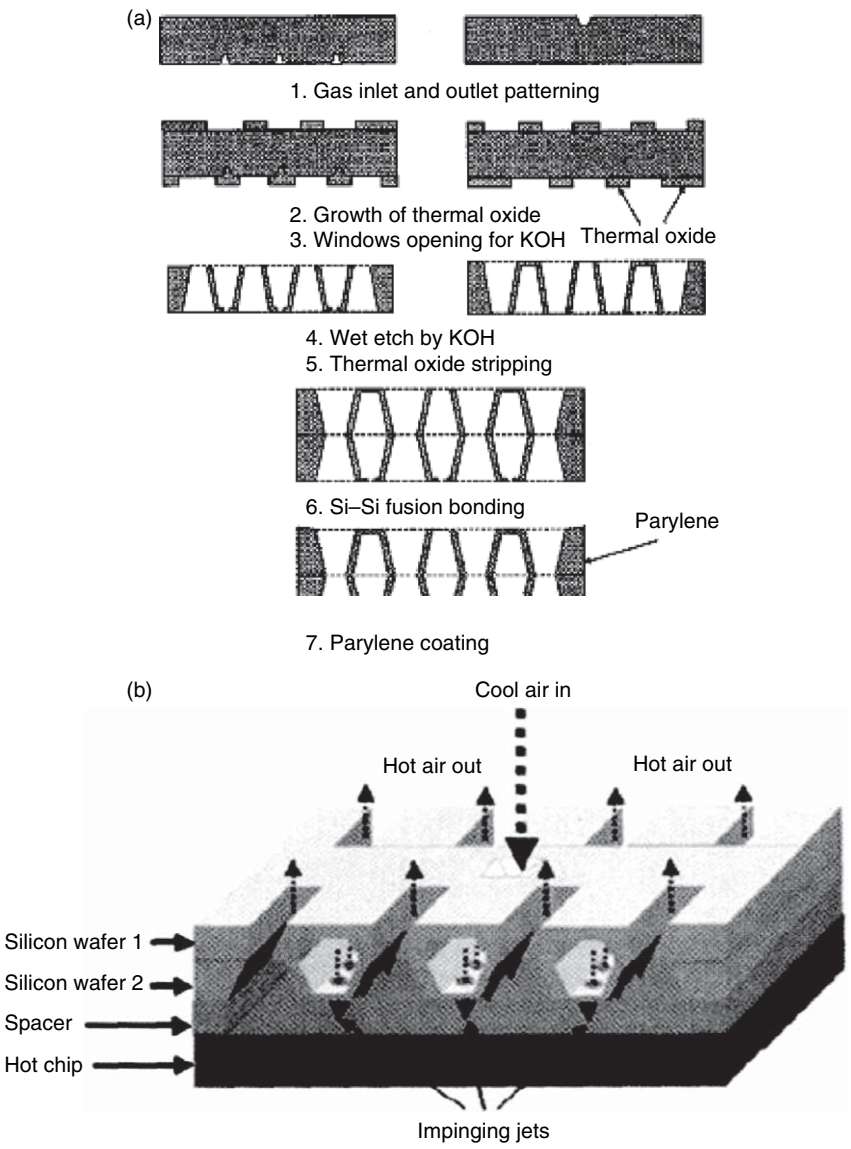


Figure 4 (a) The fabrication sequence. (Source: Wu S, Mai J, Tai Y C, Ho C M 1999 Micro heat exchanger by using MEMS impinging jets. *IEEE 12th Int. Conf. MEMS*, Orlando, FL, USA, January 17–21, 1999, pp. 171–6. (© 1999 IEEE)) (b) A schematic of two MEMS chips bonded together to form an array of impinging jets for cooling tiny surfaces. (Source: Allan R 1999 MEMS micro heat exchanger employ impinging jets to boost cooling efficiency. *Electron. Des.* **47(7)**, 29–30).

35 kPa pressure), Wu *et al.* were able to cool the temperature of the sensor chip (2 cm × 2 cm) from 70 to 33°C with 1.12 W heat removal. Another silicon microjet array was built to cool electronics by stacking four wafers (Stefanescu *et al.* 1999). Deep reactive ion etching (DRIE) techniques were extensively used to create a 18 mm × 18 mm × 0.4 mm stagnation chamber with an array of 0.25-mm diameter nozzles.

The impinging microjet cooling devices discussed thus far, however, need a pressurized gas source such as compressed air, which limits the utilization of technology for cooling many portable electronics. Recently, synthetic (zero net mass flux) jets have drawn considerable attention as they have the advantage of using ambient air, thus doing away with the necessity of plumbing high-pressure channels associated with traditional jets. A synthetic jet is formed from the entrainment of surrounding fluid that is then expelled in a jet. For spot cooling, cool ambient air is entrained in the synthetic jet device and then the expelled jet impinges on the top hot surface to be cooled, as shown in **Figure 5**. Due to the pulsating

nature of the flow, synthetic jets introduce a stronger entrainment than conventional steady jets of the same Reynolds number and more vigorous mixing between the wall boundary layers and the rest of the flow. A group at Georgia Institute of Technology developed a miniaturized synthetic jet device consisting of a vibrating diaphragm (a circular latex membrane with a diameter of 25 mm and a thickness of 225 μm), a small electromagnetic actuator to drive the diaphragm, and a housing with an orifice of various sizes, as shown in **Figure 6** (Campbell *et al.* 1998, Kercher *et al.* 2003). Campbell *et al.* (1998) reported a 22% reduction in the temperature rise in the CPU of the laptop computer when synthetic jet cooling was used instead of a normal fan. Kercher *et al.* (2003) investigated various design parameters of the synthetic jet cooling device and optimized the cooling performance for resonant frequency, orifice-to-chip spacing, orifice size, etc. Mahalingam *et al.* presented a modified version of the synthetic jet cooler similar to conventional jet ejector and integrated it into a cooling module to determine the cooling performance, which turned out to be six to eight times greater than that of comparable conventional turbulent flow (Mahalingam and Glezer 2003, 2005, Mahalingam *et al.* 2004). More recently, Pavlova and Amitay (2005) experimentally investigated the efficiency of synthetic jet impingement cooling using thermocouples and microparticle imaging velocimetry (μPIV) and concluded that synthetic jet cooling is up to three times more effective than steady jets at the same Reynolds number.

Clearly, the synthetic jet cooling technique exhibits its high potential to be utilized in electronics cooling because of its sourceless, simple operation with relatively low power consumption (less than a half watt to drive a solenoid). Further miniaturization of the jets in specific geometries related to specific device placement has real potential to revolutionize many microcooler applications. However, the heat transfer

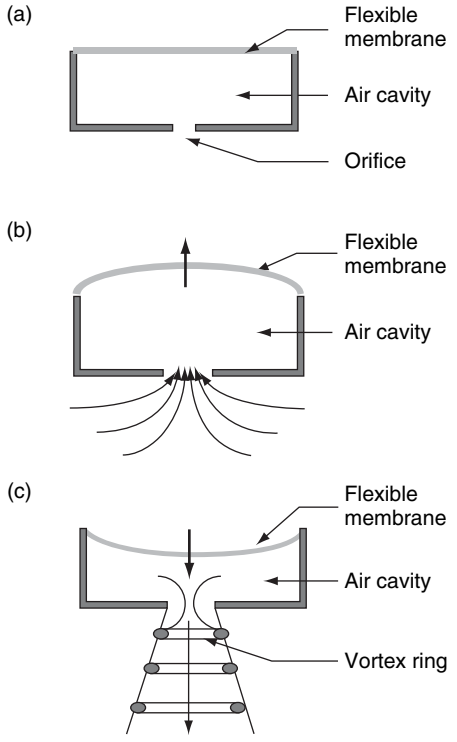


Figure 5 Schematic diagrams of a turbulent synthetic jet device and the generation of synthetic jets. (a) A turbulent synthetic jet device with its basic components: a driving element, a flexible membrane, an air cavity, and an orifice hole. (b) Entrainment of air by deflecting membrane outward followed by (c) rejection of the air by deflecting membrane inward and vortex ring formation. (Source: Kercher D S, Lee J B, Brand O, Allen M G, Glezer A 2003 Microjet cooling devices for thermal management of electronics. *IEEE Trans. Comp. Packag. Technol.* **26(2)**, 359–66. (© 2003 IEEE)).

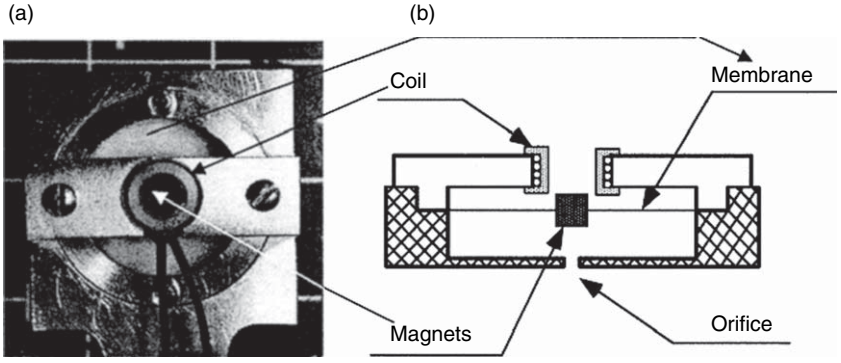


Figure 6 (a) Photograph of top view of the synthetic jet cooling device and (b) its schematic diagram. (Source: Kercher D S, Lee J B, Brand O, Allen M G, Glezer A 2003 Microjet cooling devices for thermal management of electronics. *IEEE Trans. Comp. Packag. Technol.* **26(2)**, 359–66. (© 2003 IEEE)).

coefficients achieved thus far range from 77 to 94 $W\,m^{-2}\,K$ (Kercher *et al.* 2003) compared with about 400 $W\,m^{-2}\,K$ for a large high-performance fan with heat sink. Therefore, further improvement still needs to be made to outperform a conventional fan-based air cooling, perhaps by augmenting liquid thermal management solutions for high heat flux electronics.

3.16.3.2 Liquid Cooling

Tremendous research efforts have been directed to develop small liquid cooling devices to take full advantages of the potential of liquid cooling. Microelectromechanical systems (MEMS) and microfluidics are relatively new technologies that exploit existing microelectronics infrastructure to create complicated micromachines or devices that can move and control fluids. Recent progress in MEMS and microfluidics can potentially provide a microfluidic-based thermal management solution for cooling integrated circuits with high heat fluxes. The eventual goal for liquid thermal management is to integrate miniaturized or microcooling techniques (e.g., flow heat exchangers, liquid jet or spray coolers, and heat pipes) into electronic circuitry on the same substrate using the same fabrication facility, so that low-cost, but highly advanced and functional systems can be created.

Liquid cooling for application to electronics can be categorized in a couple of different ways: (1) indirect or direct liquid cooling, depending on whether the liquid is in a direct contact with the components to be cooled or pass through separate heat exchangers and/or other material structures (Lasance and Simons 2005) and (2) single-phase or two-phase fluidic cooling (Mudawar 2001). The direct cooling method such as immersion cooling can theoretically produce higher cooling power because of the absence of the interfacial thermal resistance but limits the choice of the liquid coolants and the type of the fluid actuation. Two-phase cooling, in general, can produce a large increase in the removal rate of heat flux with only a modest increase in device temperature, but with potential complications in fluid dynamics and system complexity and control. There is, however, no clear distinction of which device belongs to what category. In fact, many of the cooling devices discussed in the literature can be realized in either way; for example, a thermosyphon can be configured in a direct or an indirect mode, and the microchannel heat exchanger can have a single-phase and/or two-phase cooling. In the context of microcooler technology, we discuss the important working principles of cooling for each system, such as boiling heat transfer in two-phase cooling, and focus on examples of liquid cooling devices employing such principles.

Before beginning the discussion, the main motivation of why liquid coolants are heavily researched and developed in microcoolers is the potential for increasing heat flux removal. Compared with that of air, the capacity of most liquids to store energy is considerably higher, i.e., the intensive heat capacity, c_p, of liquid water at room temperature ($\sim$4.2 $J\,g^{-1}{}^{\circ}C^{-1}$) is four times that of air, and the mass specific capacity, ρc_p, is 4000 times greater. In addition, the thermal conductivity at room temperature ($\sim$0.6 $W\,m^{-1}{}^{\circ}C^{-1}$) is 24 times greater than air ($\sim$0.024 $W\,m^{-1}{}^{\circ}C^{-1}$). Of course, water is an extraordinary molecule, and other liquids that are used for their dielectric properties such as fluorocarbons have considerably lower differences but are still far higher than air. Therefore, single-phase liquids can adsorb and transfer significantly more heat. Allowing a liquid-to-vapor phase change increases the amount that can be absorbed and transferred by orders of magnitude more than for a single phase of the same substance. Recent reviews on various liquid coolants and their thermophysical properties are available elsewhere (Mohapatra 2006, Mohapatra and Loikitis 2005). Gambill and Lienhard (1989) estimated the maximum theoretically possible heat flow removed by boiling based on water molecules turning into vapor without influencing each other to be 2000 $k\,W\,cm^{-2}$, although the highest reported experimental value is over 200 $k\,W\,cm^{-2}$, using high velocities and high pressures (Lasance and Simons 2005), which is still nearly three orders of magnitude higher than the highest air-based transfer. Some liquid microcooler devices discussed later can handle a few hundred watts per square centimeter, and there is significant potential by making use of nano- to microstructured geometries of pushing the microscale heat transfer much closer to the theoretical maximum, which is why liquid microcooler research is so promising.

3.16.3.2.1 Passive two-phase heat spreaders

In traditional electronics cooling, heat sinks of high thermally conducting materials are utilized to conduct heat from the electronic chip to the larger surface areas of the heat sink where the heat is rejected to the coolant via natural or forced convection. The merit of heat spreading to lower the heat flux density by increasing heat transfer area can be

compromised, however, as the device size becomes smaller and the thermal resistance across the smaller and/or thinner heat sink is no longer negligible. Passive two-phase heat spreaders such as thermosyphons or heat pipes have been developed and nowadays have found use as heat spreaders for CPU in many computers. The advantages of thermosyphons or heat pipes stem from harnessing the two-phase phenomenon including (1) the latent heat of vaporization of the coolant can take away considerably more heat than the sensible heat change of the fluid and (2) liquid temperature in the cooler is approximately constant, minimizing thermal resistance and entropy generation in heat spreaders. According to Fourier's law, the ability of transporting large amounts of thermal energy between two points (evaporator and condenser) with a small temperature difference is equivalent to having an extra high thermal conductivity (higher than copper), which is analogous to near superconductor under an electrical field (Tien 1975). Thermosyphons and heat pipes are also attractive for microcooler devices because there is no need to drive fluid with active micropumps, making them simpler, lighter, more compact, and less expensive. These advantages become even more important as the device size is miniaturized further, and the space to accommodate a microcooler becomes tighter.

One of the key issues in heat spreader technology is the choice of thermal interface materials (TIMs). The TIMs must possess high thermal conductivity and often require that its coefficient of thermal expansion be similar to silicon in order to eliminate unwanted mechanical failures. Recent developments of TIMs using metallic nanoparticles or carbon nanotubes (CNTs) are an interesting topic, but beyond the scope of this work. Those interested can refer to recent articles (Prasher *et al.* 2003, Wu *et al.* 2005, Xu and Fisher 2006) and reviews (Dani *et al.* 2005, Gwinn and Webb 2003).

3.16.3.2.1.(i) Thermosyphon – Boiling A two-phase closed thermosyphon, also known as a wickless heat pipe, is a hermetically sealed chamber partially filled with a volatile fluid and free of noncondensable gases. Thermosyphons have been under study as thermal management solution for the past 50 years (Tuma 2006). A device to be cooled can be either immersed directly into a pool of inert dielectric liquid or indirectly cooled by transferring heat to the boiler part of the thermosyphon, as shown in **Figure 7**. Once the liquid is heated by the chip, it vaporizes in the boiler. The vapor travels to the condenser and condenses while dissipating heat to the ambient via natural or forced convection. The condensed liquid is returned to the evaporator, typically by gravitational force in classical thermosyphons, completing a closed loop. The operation of thermosyphons is similar to that of heat pipes except that a heat pipe employs a wicking structure for driving a condensate back to the evaporator. Early research on macroscale thermosyphons configured as indirect liquid cooling for microelectronics was reviewed by Palm and Tengblad (1996).

Among the several different types of resistance that limits the performance of a thermosyphon such as boiling resistance, condensation resistance, and vapor drag, the primary limitation in many designs

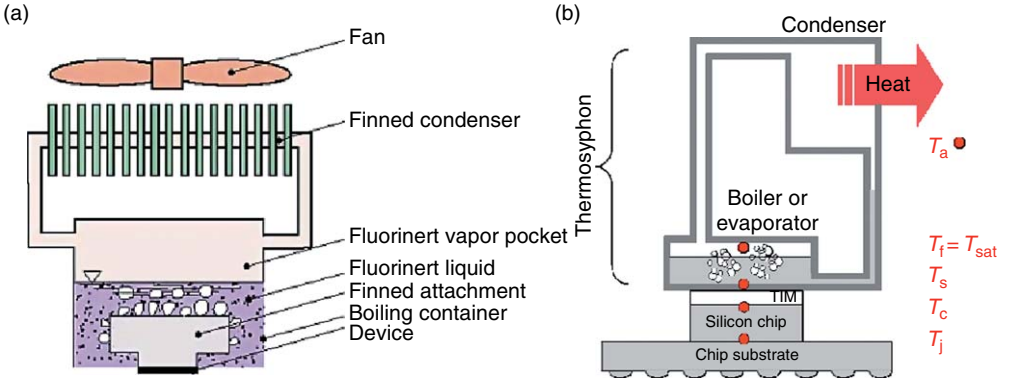


Figure 7 Schematic diagrams of thermosyphon configured with (a) direct immersion cooling (Source: Mudawar I 2001 Assessment of high-heat-flux thermal management schemes. *IEEE Trans. Comp. Packag. Technol.* **24(2)**, 122–41 cooling (© 2001 IEEE)) and (b) indirect cooling modes. (Source: Tuma P E, Mortazavi H R 2006 Indirect thermosyphons for the cooling electronic devices. *Electronics Cooling* **12(1)**, 26–33. http://www.electronics-cooling.com/articles/).

lies in boiling, i.e., evaporation at a solid–liquid interface. Therefore, a basic understanding of boiling heat transfer is required to properly design a thermosyphon and analyze its performance. The type of boiling occurring in a thermosyphon is called saturated pool boiling, often characterized by the critical heat flux (CHF) that is the maximum heat flux, q_{CHF}, in the nucleate boiling regime. A higher heat flux than the CHF lets the vapor bubbles coalesce to form a stable vapor blanket on the surface, effectively limiting the heat transfer. The boiling heat transfer coefficient, h_{boil}, is defined as

$$h_{boil} = \frac{q_{CHF}}{A_{surf}\Delta T_{sat}} \quad [1]$$

where A_{surf} is the projected geometrical area where boiling occurs and $\Delta T_{sat}(= T_{wall} - T_{sat})$ the wall superheat. Therefore, a higher projected heat transfer coefficient can be achieved either by increasing q_{CHF} or by decreasing the excess temperature overshoot. Nakayama and his coworkers (1980a, b, 1982, 1984) utilized a 3D network of interconnected microchannels made of multilayers of perforated copper with relatively large pore sizes and evaluated boiling enhancement, as shown in **Figure 8**. Mudawar and Anderson (1993) used the similar enhanced structures to achieve the CHF greater than 100 W cm^{-2} with FC-72. These enhanced surfaces can provide (1) more nucleation sites for bubbles to initiate and grow and (2) an extended surface from which sensible heat transfer takes place. Along with a manifold increase in the CHF, the enhanced structures also effectively lower temperature overshoot prior to boiling initiation, which should lead to reliable operation of the electronic devices (Mudawar 2001). Surveys of various surface treatment techniques employed for boiling enhancement can be found in Thome (1990) and in Webb and Kim (2005), including 3D cavities, etched grooves, electroplating, attached wire and screen promoter, oxide and ceramic coatings, porous surfaces, and structured rough surfaces. These enhanced surfaces commonly have reentrant cavities, entrapping vapors and creating active nucleation sites for boiling initiation. Examples of the enhanced surfaces from the literature are shown in **Figure 9**. Apart from the surface enhancement effect, other studies reveal that the CHF is also dependent on (1) subcooling of coolants (Bergles 1997, Vandervort *et al.* 1994), (2) the surface orientation (Chang and You 1996, El-Genk and Guo 1993, Howard and Mudawar 1999, Mudawar *et al.* 1997), and (3) external vibration (Park and Bergles 1988).

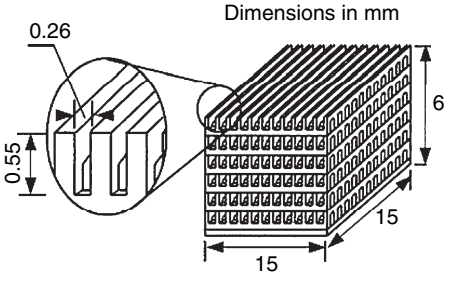


Figure 8 Enhanced evaporator structure with six bonded layers of perforated copper. (Sources: Nakayama W, Nakajima T, Hirasawa S 1984 Heat sink studs having enhanced boiling surfaces for cooling microelectronic components. *ASME Paper No. 84-WA/HT-89*; Pal A, Joshi Y K, Beitelmal M H, Patel C D, Wenger T M 2002 Design and performance evaluation of a compact thermosyphon. *IEEE Trans. Comp. Packag. Technol.* **25(4)**, 601–7. (© 2002 IEEE)).

Recent boiling heat transfer studies whose experiments are configured for electronics cooling are summarized in **Table 1**. With various boiling enhanced surfaces, many researchers employed traditional machining to build macroscale prototype thermosyphons and measure the heat transfer characteristics when cooling microelectronic components over a range of heat fluxes (Pal *et al.* 2002, Webb and Yamauchi 2002). Up to 200 W cm^{-2} of the CHF and 13.5 W cm^{-2} K of h_{boil} have been reported using water or dielectric fluorocarbon fluids as coolants (Klett *et al.* 2004, McGillis *et al.* 1990). A group led by Joshi at the Georgia Institute of Technology has been developing miniature or microthermosyphon systems for on-chip cooling. Ramaswamy *et al.* (2000, 2003) used the 3D boiling enhanced structure similar to Nakayama's to make a compact thermosyphon and investigated the effects of the chamber pressure, subcooling temperature, vapor space confinement, and the orientation of the evaporation structures. Employing a similar microfabricated copper structure, Murthy *et al.* (2000, 2002, 2004) further developed a miniature thermosyphon system as an on-chip cooling system as shown in **Figure 10**. They also demonstrated the orientation-independent performance derived from geometrical designs of the chamber (Murthy *et al.* 2003).

Loop thermosyphons (LTSs) ideally have less hydraulic resistance for the liquid returning to the evaporator compared with conventional thermosyphons because of the absence of the vapor drag from the counter flow. The driving force to maintain the vapor flow in LTS is the pressure head generated from the liquid column in the liquid return channel. In turn, the major pressure drop occurs in the vapor

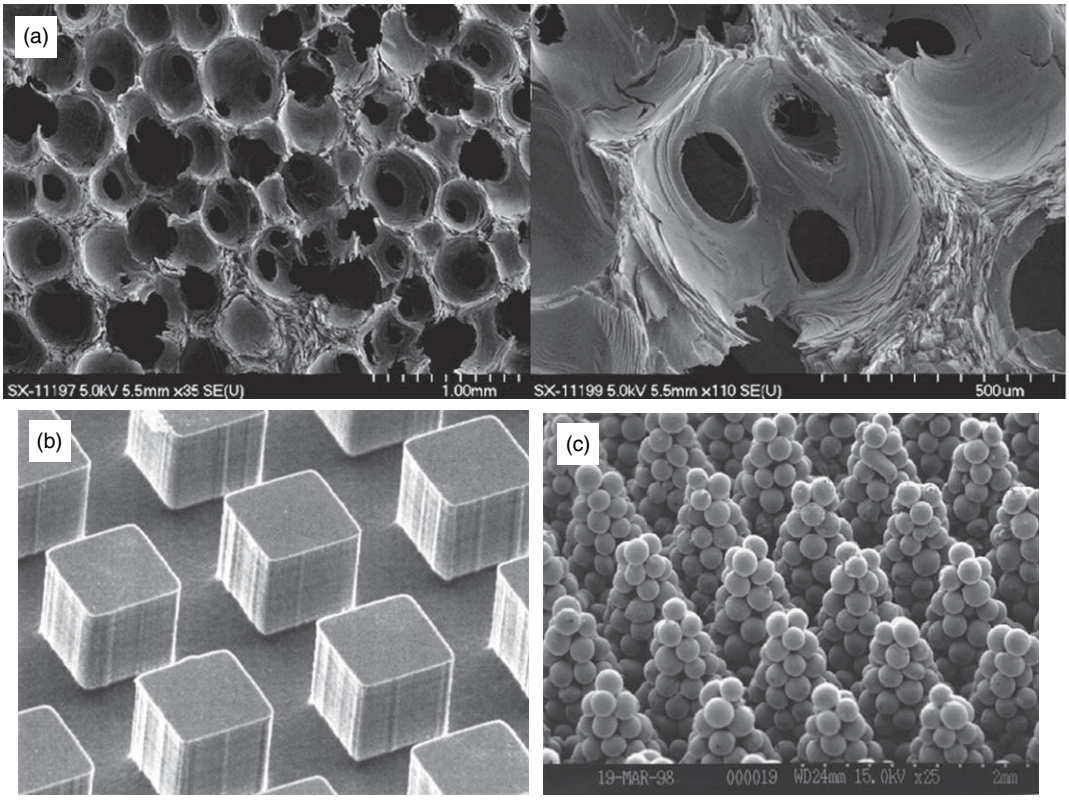


Figure 9 Scanning electron micrographs of various enhanced surface structures to extend boiling regime. (a) Porous graphite foam (Source: Klett J W, Trammell M, Boudreaux P 2004 Parametric investigation of a graphite foam evaporator in a thermosyphon with FluorinertTM and a silicon CMOS chip. *IEEE Trans. Device Mater. Reliabil.* **4(3)**, 626–37. (© 2004 IEEE)), (b) DRIE-etched silicon pin-fin array (Source: Honda H, Takamastu H, Wei J J 2002 Enhanced boiling of FC-72 on silicon chips with micro-pin-fins and submicron-scale roughness. *ASME J. Heat Transf.* **124**, 383–90.), and (c) height-modulated porous-layer coating of spherical copper particles (of diameter 200 μm) (Source: Reprinted from Liter S G, Kaviany M 2001 Pool-boiling CHF enhancement by modulated porous-layer coating: theory and experiment. *Int. J. Heat Mass Transf.* **44**, 4287–311., Copyright 2001, with permission from Elsevier).

channel, giving rise to fluid transport limitation. Chu *et al.* (1999) built a macroscale LTS and experimentally investigated the effects of fill volume, heat load, and condenser airflow rate on overall thermosyphon performance. Average heat transfer coefficients on the flat-plate evaporator ranging from 0.38 to 0.8 $W\,cm^{-2}\,K$ were reported. Khrustalev (2002) reported 12 LTSs, sharing one cold plate to simulate the cooling of multiple electronic components.

With all these extensive studies on pool boiling heat transfer, many of their data were collected in an apparatus or under conditions that are impractical for microcoolers (Tuma 2006). As summarized in **Table 1**, some of the reported heat transfer coefficients for miniaturized thermosyphon systems are much smaller than those of macroscale counterparts (Murthy *et al.* 2002). As the device size is reduced further, the condensation resistance becomes significant as the area for the terminal heat rejection to ambient by subsequent forced-air or forced-liquid cooling decreases. In addition, entrainment of fluid within the flow paths plays a much larger role in limiting performance than for macro- or even mesoscale systems. Another issue in two-phase passive heat spreading is the sensitivity of the system to orientation, which causes dry-out to occur when the condenser outlet and the evaporator inlet lie in the same horizontal plane (Pal *et al.* 2002). Of course, for microscale systems, the effect of gravity as the driver of returning liquid to the evaporator becomes much weaker as the size is reduced, and if the dimensions are too small, gravity becomes a negligible force. Thus, as size is reduced to the microscale, thermosyphons begin to employ wicking structures for the return, at least over part of the path, which leads to our discussion of micro heat pipes (MHPs).

Table 1 Summary of the thermosyphon research on electronics cooling in literature

References	*Test fluid (sat. temp.)*	*Surface enhancement*	*Enhanced surface (cm)*	*q″ (W cm^{-2})*	*h_{boil} (W cm^{-2}-K)*
		Immersed or indirect thermosyphons			
Nakayama *et al.* 1984	FC-72 (56)	Six layers of perforated Cu plates	1 × 1	100	3.5
McGillis *et al.* 1990	Water (29) Water (44)	Cu pin-finned plates	1.27 × 1.27	90 200	5 6
Chang and You 1997a	FC-72 (56)	Diamond particle-coated Cu	1 × 1	14.3	2.05
Chang and You 1997b	FC-72 (56)	Al particle-coated Cu	1 × 1	26.3	2.89
Webb *et al.* 1998	R-134a	Flamed sprayed porous Al	2.5 × 2.5	20.4	8.7
Rainey and You 2000	FC-72 (56)	Al particle-coated, Cu finned block	1 × 1	82.8	3.76
Honda *et al.* 2002	FC-72 (56)	DRIE etched micropin-finned Si	1 × 1	60 30	3 1.03
Ramaswamy *et al.* 2000	FC-72	Stack of perforated Cu plates	1 × 1	70	2.30
Liter and Kaviany 2001	Pentane (36)	Porous-layer coating of spherical Cu particles	Ø 5.08	71.2	4.45
Pal *et al.* 2002	PF5060	Six layers of perforated Cu plates	1.27 × 1.27		
Murthy *et al.* 2002	FC-72	Microfabricated Cu structure	1.5 × 1.5	42.5	0.74
Rainey *et al.* 2003	FC-72 (56)	Al particle-coated, Cu finned block	1 × 1	102	2.49
Wei *et al.* 2003a	Water	Sintered metal particle	1.7 × 1.7	80	4
Ramaswamy *et al.* 2003	FC-72	Stack of perforated Cu plates	1 × 1	80	3.48
Mukherjee and Mudawar 2003	FC-72 (56) PF5050 (56)	Microfinned Cu channels	2.13 × 2.13	62 113.3	2.5 3.91
Roh 2003	PF5060 (30) FC-72 (56)	Graphite foam Graphite foam	3 × 3	98 118	1.78 4.9
Klett *et al.* 2004	FC-87	Slotted graphite foam	1 × 1	149	13.5
El-Genk *et al.* 2005	FC-72 (52)	Porous graphite	1 × 1	85.6	8.33
Chien and Chen 2005	Water (60)	Cross-grooved Cu plates	1.4 × 1.4	44	4.1
		Loop thermosyphons			
Chu *et al.* 1999	Water	Cu pin-finned plates	12.6 × 12.6	N/A	0.8
Khrustalev 2002	Methanol	Cu tube with sintered Cu wick	N/A	70	N/A

3.16.3.2.1.(ii) Micro heat pipes The working principle of the heat pipes is similar to that of the thermosyphons (i.e., near-isothermal evaporation–condensation cycle) and differs only in that the liquid condensed in the condenser returns to the evaporator via the capillary force of the wick. Since the thermosyphon is a gravity-dependent device that can be used neither for space applications nor for orientation-dependent applications, conventional heat pipes have provided the viable thermal management in numerous applications including terrestrial applications. The concept of an MHP was introduced for electronics cooling by Cotter (1984) who defined the MHP as "one so small that the mean curvature of the vapor–liquid interface is necessarily comparable in magnitude to the reciprocal of the hydraulic radius of the total flow channel." Due to their size, material, and manufacturing process, instead of traditional wick structures, MHPs often employ microchannels with a cusped cross section (e.g., polygon), having a diameter in the range of about 10–500 μm. The surface tension gradient that drives the capillary liquid flows is generated at the sharp corners in MHPs as shown in **Figure 11**. Due to lesser liquid in the evaporator section (**Figure 11**, A–A), the radius of curvature of the meniscus is smaller at the vapor–liquid interface, leading to a lower liquid pressure compared to that in the conderser (**Figure 11**, D–D) and thus the liquid moving from the conderser to the evaporator.

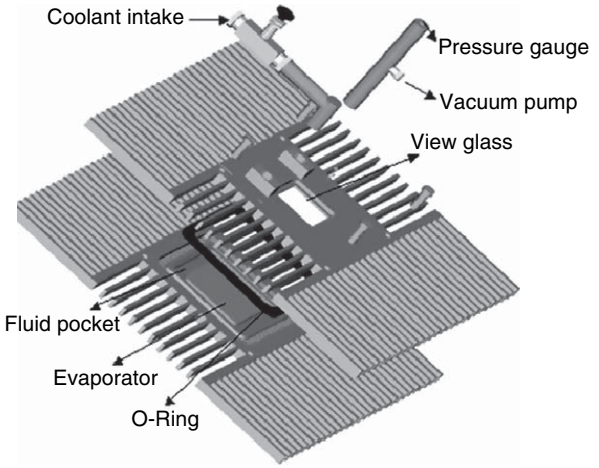


Figure 10 Exploded view of a single thermosyphon two-phase heat spreader. (Copyright 2004 from Two-phase heat spreaders utilizing microfabricated boiling enhancement structures. *Heat Transfer Eng.* **25(1)**, 26–36 by Murthy S, Joshi Y, Nakayama W. Reproduced by permission of Taylor & Francis Group, LLC., http://www.taylorandfrancis.com)

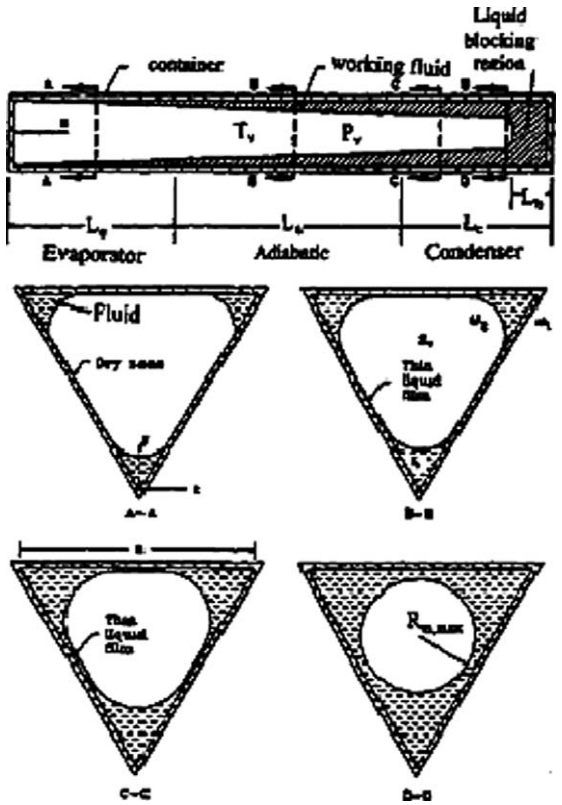


Figure 11 Schematic of micro heat pipes. (Source: Khrustalev D, Faghri A 1994 Thermal analysis of a micro heat pipe. *ASME J. Heat Transf.* **116**, 189–97).

Excellent reviews on the MHP by Faghri (1995) and Peterson (1992, 1996) encompass the early efforts in the development and characterization, emphasizing numerical analysis, fabrication, and testing of MHPs, manufacturing as standalone devices or fabricated as an integral part of the silicon wafers. Since then, great advances in research related to arrays of MHPs, flat-plate MHPs, and capillary loop heat pipes have preceded commercial success, and state-of-the-art MHPs can be found in many current laptop computers as primary heat spreaders (Wuttijumnong *et al.* 2004). Here we summarize the efforts on MHPs appearing after these reviews and focus on the direct integration of MHPs into silicon substrates. The readers can also refer to recent reviews on micro and miniature heat pipes as electronic component coolers (Groll *et al.* 1998, Vasiliev 2005, 2006).

Table 2 summarizes the materials, manufacturing, and performance of various miniature and micro heat pipes to date. An objective parameter characterizing the system efficiency is the effective thermal resistance (R_{th}) determined by the ratio of the temperature difference (ΔT) between the heat source (evaporator) and the heat sink (condenser) to the value of the dissipated capacity (q_{in}) (Pastukhov *et al.* 2003). Some researchers reported the effective thermal conductivity (k_{eff}) of their MHPs, which is related to the effective thermal resistance by

$$k_{eff} = \frac{1}{R_{th}} \cdot \frac{L_{eff}}{A_{eff}}, \quad R_{th} = \frac{\Delta T}{q_{in}} \qquad [2]$$

where L_{eff} is the effective length between the evaporator and the condenser and A_{eff} the cross-sectional area of the MHP. The heat removed by the MHPs ranges from a few watts per square centimeter to over $300\,W\,cm^{-2}$. A water–copper heat pipe system, manufactured by traditional

Table 2 Summary of the recent research on miniature and micro heat pipes for electronics cooling in literature

References	*Type*	*Wick structure*	*Materials/ processing*	*Overall size (cm × cm × cm)*	*Evap. area (cm × cm)*	*Working fluid*	*Parametric study*	*R_{th} (°C W^{-1}) k_{eff}(W m^{-1} K^{-1})*	*Maximum heat removed (W)*
Cao and Gao 1996	Flat MHP	EDM machined, axial Cu groove	Cu/EDM	8.2 × 0.7 × 0.3	1.9 × 1	Water	N/A	40 times copper	20.6
Park *et al.* 2001	Flat MHP array	KOH etched Si groove	Silicon and glass/ KOH etching, EDM machining	2.4 × 1.6 × 0.125	0.5 × 0.88	Water	Liquid fill ratio	N/A	5.9
Kirshberg *et al.* 2000, Pettigrew *et al.* 2001	Micro CPL	Wet etched glass channels	Si and glass/wet and DRIE etching, anodic bonding	3.5 × 0.3 × 0.1	0.1 × 0.2	Water	N/A	0.82°C W^{-1}	8.5
Chung *et al.* 2001	Capillary pumped LHP	Cu or Al capillary	Cu or Al/machining	10 × 7.35 × 0.4	2.5 × 2.5	Water	Material orientation	0.64°C W^{-1}	640
Cytrynowicz *et al.* 2002	Loop heat pipe	Porous silicon	Si and glass/KOH etching	N/A	N/A	Water	Pressure drop wick structure	N/A	N/A
Ponnappan 2002	Mini heat pipe (HP)	Mesh Cu screen Cu axial groove	Cu/machining	10 × 1.27 × 0.635	1.85 × 1.27	Water	Liquid fill ratio wick structure	0.25°C K^{-1} (2.2 W cm^{-2} K^{-1})	150
Kang *et al.* 2002	Flat-plate MHP array	KOH etched, radial grooved Si	Three Si layers/KOH etching, Au eutectic bonding	5 × 5 × 0.2	1.1 × 1.1	Water	Liquid fill ratio	1°C W^{-1}	25
Kang and Huang 2002	Flat-plate MHP array	KOH etched, star-shaped Si groove	Three Si layers/KOH etching, Au eutectic bonding	2.5 × 2.5 × 0.1	0.6 × 2.5	Methanol	Geometry of capillary groove, liquid fill ratio	289 W m^{-1} K^{-1}	20.6

Pastukhov *et al.* 2003	Miniature LHP	Ti wick	Al pipe and Al radiator plate	25 × Ø 0.5	2.5 × 2.5	Ammonia	Geometry of loop	0.3°C W^{-1}	40
Khandekar *et al.* 2003	Flat-plate MHP	Cu thread	Cu/machining	4 × 4 × 0.09	4 × 1	Water	Orientation	N/A	22
LeBerre *et al.* 2003	Flat-plate MHP array	KOH etched, triangular Si groove	Three Si layers/KOH etching, Si–Si direct bonding	2 × 2 × 0.1	N/A	Methanol	Liquid fill ratio use of artery	600 W m^{-1} K^{-1}	4
Kang *et al.* 2004	Flat-plate MHP array	Etched Cu groove, 100-mesh Cu screen	Three Cu foil layers/ chemical etching, vacuum diff. bonding	3.1 × 3.1 × 0.27	1.4 × 1.4	Water	Liquid fill ratio	1.05°C W^{-1}, 1.6°C W^{-1}	45
Launay *et al.* 2004	Flat-plate MHP array	KOH etched artery groove	Two or three Si layers/KOH etching, Si fusion bonding	2 × 0.5 × 0.1	2 × 0.5	Methanol	Liquid fill ratio	0.074°C W^{-1} (900 W m^{-1} K^{-1})	2
Lai *et al.* 2004	Flat MHP	DRIE etched Si groove	Three Si layers/DRIE etching, fusion bonding	6 × 2 × 0.15	1 × 1	Water	Liquid fill ratio temp of coolant	0.8°C W^{-1} (1200 W m^{-1} K^{-1})	32
Moon *et al.* 2004	MHP	Drawn Cu pipe with polygon X-sec.	Oxygen-free Cu/ precision drawing	5 × Ø 0.15	1 × ph 0.15	Water	Geometry of capillary groove	2.2°C W^{-1}	7
Tsai *et al.* 2005	Flat-plate LHP	Comb-grooved Cu	Cu pipes, evaporator	16 × 12 × 0.5	1.4 × 1.4	Methanol	Various coolants	N/A	68
Singh *et al.* 2005	Flat mini HP	Cu composite fiber Cu axial groove	Cu/machining	10 × 0.8 × 0.15	0.7 × 0.7	Water	Wick structure	0.25°C W^{-1}	3

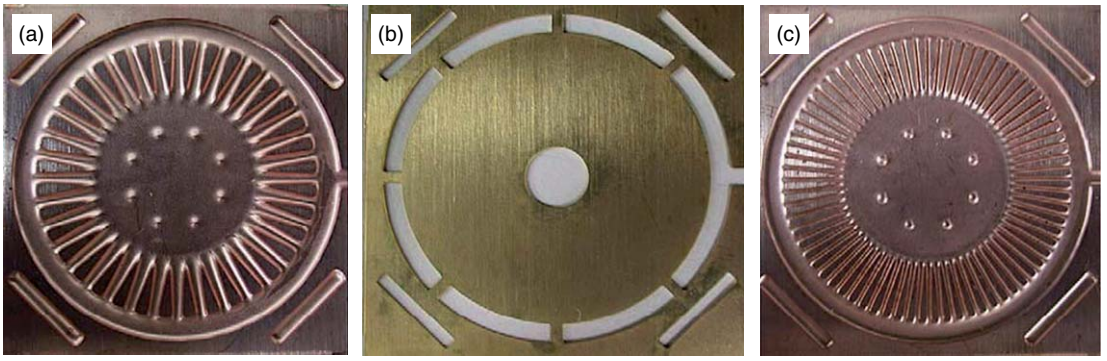


Figure 12 Photographs of each layer of a groove-shaped micro heat pipe (MHP) heat spreader: chemically etched Cu grooves for gas-phase (a) and for liquid-phase (c) fluids and the interface Cu layer (b). (Source: Reprinted from Kang S-W, Tsai S-H, Ko M-H 2004 Metallic micro heat pipe heat spreader fabrication. *Appl. Therm. Eng.* **24**, 299–309, Copyright 2004, with permission from Elsevier).

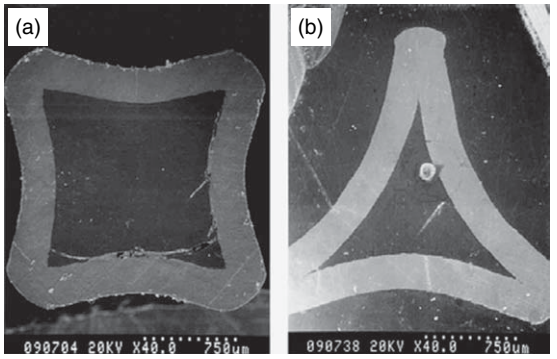


Figure 13 Cross-sectional views of a micro heat pipe (MHP) with a (a) curved rectangular type and (b) curved triangular type of microchannels. (Source: Reprinted from Moon S W, Hwang G, Ko S C, Kim Y T 2004 Experimental study on the thermal performance of micro-heat pipe with cross section of polygon. *Microelectron. Reliabil.* **44**, 315–21, Copyright 2004, with permission from Elsevier).

machining or EDM, dominates in the MHP literature and shows a heat transfer capacity of 5–100 W cm^{-2}. Kang *et al.* (2004) used a chemical etching technique to shape radial grooves in copper and stacked three Cu layers to build a flat-plate MHP array as shown in **Figure 12**. Moon *et al.* (2004) utilized a high-precision drawing process to fabricate the copper tube with a polygon cross section. **Figure 13** depicts the cross section of their MHPs, having sharper edge in the corners for higher capillary forces.

Together with conventional machining, MEMS fabrication processes such as bulk silicon micromachining have been increasingly implemented to fabricate MHPs with hydraulic diameters on the order of 50–500 μm as an integral part of the silicon substrate. The use of directionally dependent etching of silicon was proposed in the earlier work of Peterson and coworkers (Mallik *et al.* 1992, Peterson 1992) and has been experimentally verified (Peterson *et al.* 1993). Since then, an extensive investigation has been directed toward refining the fabrication procedure and developing flat-plate MHP arrays. **Figure 14** illustrates the fabrication processes for the star-grooved MHP arrays in which each corner with more acute angles provides better capillarity (Kang and Huang 2002). Many other silicon-based MHPs follow the similar process sequence with a modification in the bonding step, such as Si–Si direct fusion bonding instead of Au–Au eutectic bonding (**Table 2**). Le Berre *et al.* (2003) coupled typical KOH-etched triangular channels with arteries, as shown in **Figure 15**, for the separate liquid transport and demonstrated 300% enhancement in the effective thermal conductivity compared with the MHPs without the arteries. DRIE of silicon has also been used to create an array of capillary grooves and the MHP prototype proving the comparable performance (Lai *et al.* 2004). **Figure 16** shows DRIE etched grooves and a schematic of a flat-plate silicon MHP. A polymer-based MHP system using SU-8 photocurable resin and low-temperature (less than 110°C) bonding have been used as the thermal management solution to InP/InGaAs-based devices, which often cannot withstand a high-temperature wafer-to-wafer bonding (over 500°C) process (Liu *et al.* 2003). The effective thermal conductivity of these MHPs reported in the literature varies from 250 to 16 000 W m^{-1}°C^{-1} as summarized in **Table 2**. Many of these MHPs can conduct a relatively high load of heat farther away than either copper or diamond (**Table 3**).

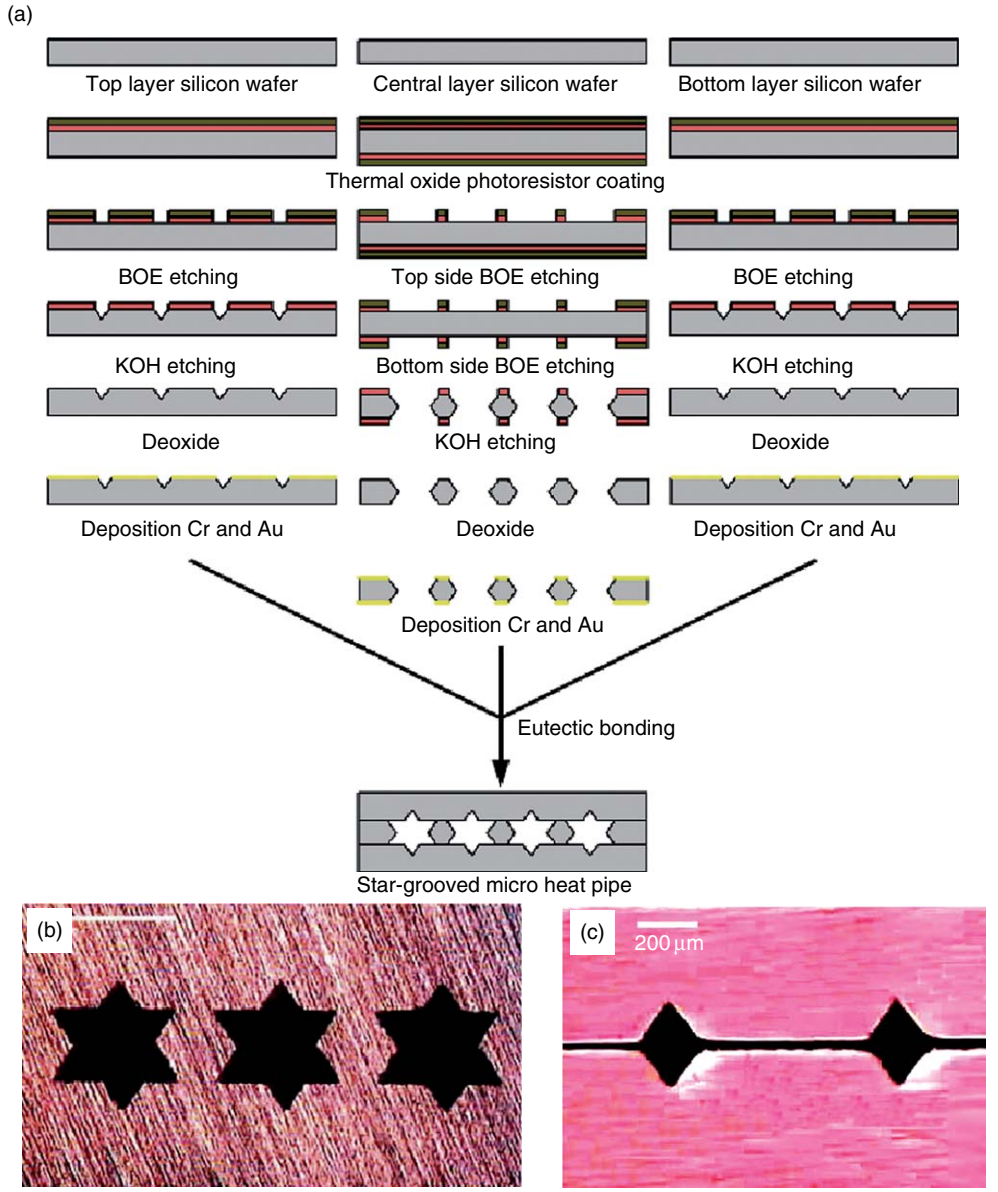


Figure 14 (a) Fabrication process for star-grooved micro heat pipes (MHPs) and (b, c) optical micrographs of star-shaped and rhombus-shaped grooves. (Source: Kang S-W, Huang D 2002 Fabrication of star grooves and rhombus grooves micro heat pipes. *J. Micromech. Microeng.* **12**, 525–31).

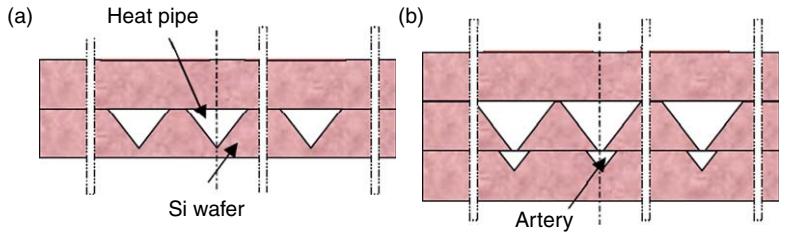


Figure 15 Transverse cross-sectional diagram of a micro heat pipe (MHP) array (a) with triangular channels (two wafers) and (b) with triangular channels coupled with arteries (three wafers). (Source: LeBerre M, Launay S, Sartre V, Lallemand M 2003 Fabrication and experimental investigation of silicon micro heat pipes for cooling electronics. *J. Micromech. Microeng.* **13**, 436–41).

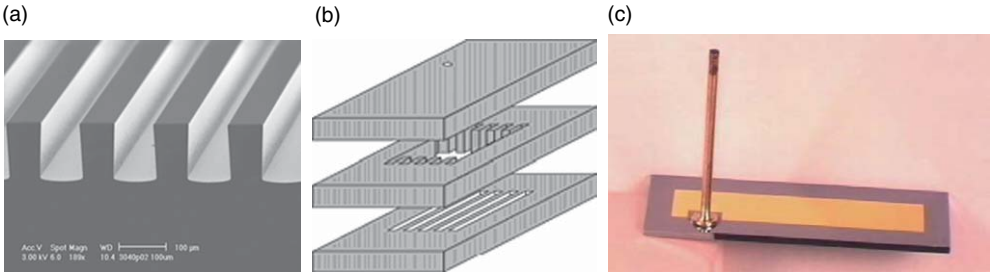


Figure 16 (a) DRIE etched Si groove, (b) schematic of Si grooved heat pipe structure, and (c) a picture of a filling-ready prototype. (Source: Lai A, Gillot G, Ivanova M, Avenas Y, Louis C, Schaeffer C, Founier E 2004 Thermal characterization of flat silicon heat pipes. *IEEE 20th SEMI-THERM Symp.*, San Jose, CA, USA, pp. 21–5, (© 2004 IEEE)).

Table 3 Mechanical properties of metals and ceramics with high thermal conductivity

Material	*Density (g cm^{-3}) at 300 K*	*Thermal conductivity at 300 K (W m^{-1} K^{-1})*	*CTE at 300 K (ppm °C^{-1})*
Silicon	2.3	151	4.2
Aluminum	2.7	238	23.6
AlN	3.3	170–200	4.5
Beryllia	3.9	250	7.6
Copper	8.9	398	17.8
CuW (10–20% Cu)	15.7–17.0	180–200	6.5–8.3
CuMo (15–20% Mo)	10	160–170	7.0–8.0
Diamond	3.5	2500	1.1

Source: Copyright 2004 From Two-phase heat spreaders utilizing microfabricated boiling enhancement structures by Murthy S, Joshi Y, Nakayama W. Reproduced by permission of Taylor & Francis Group, LLC., http://www.taylorandfrancis.com

The performance limitation of the MHPs can originate from many different phenomena such as flow within the capillary structures, sonic choking of the vapor velocity, entrainment of refrigerant, and boiling regimes similar to the conventional heat pipe case. It has been shown, however, that the MHPs primarily suffer from capillary and boiling limits (Hopkins *et al.* 1999). A capillary limit is reached when the capillary pressure difference in the liquid phase is not large enough to overcome the sum of the pressure drops associated with the liquid flow, vapor flow, and phase change phenomena, to return the fluid back to the evaporator. A boiling limit occurs when intense heat in the evaporator region gives rise to vapor bubbles, obstructing the flow of liquid into the evaporator. In addition, Cao and Faghri (1994) identified the vapor continuum limitation associated only with MHPs. As the size of the MHPs decreases, the mean free path of the vapor is comparable with the minimum dimension of the vapor passage, where the vapor may lose its continuum characteristics. Due to these limitations, the effectiveness of an MHP, in general, decreases with decreasing lengths. For MHPs with a length less than about 1 cm, the performance of a solid piece of metal such as copper is often comparable (Lasance and Simons 2005), whereas for larger heat pipes, the effective thermal conductivity can be many times that of copper. The effect of the gravity on the operation of the MHPs has been investigated analytically (Sugumar and Tio 2006) and experimentally, and it is shown that the orientation (i.e., the angle of inclination) can significantly influence their performance and that the degree of the gravity effect depends on the cross-sectional shape of the MHPs. Detailed descriptions of thermal analysis and heat transfer limitations of MHPs and LHPs are not discussed here but can be found elsewhere (for MHP: Faghri 1995, Khrustalev and Faghri 1994, Lefèvre and Lallemand 2006, Suman *et al.* 2005, Tio *et al.* 2000; for loop heat pipe (LHP): Hamdan *et al.* 2003, Kim and Golliher 2002).

Recently, LHPs or capillary pumped loops (CPLs) have gained more interest. The advantages of the LHP or CPL over the conventional HPs or flat-plate MHPs are (1) greater potential of geometric freedom in design, (2) higher heat loads due to the concurrent nature of the vapor and liquid flows, as opposed to the countercurrent flows in the conventional HPs, and (3) the potential to easily integrate new components (e.g., valves and pumps) for further performance enhancement (Kirshberg *et al.* 2000). The capacity of these LHPs or CPLs to accommodate a heat flux was

shown to be as high as 625 W cm^{-2} theoretically by Kim and Golliher (2002) for a particular configuration, and a flux of 104 W cm^{-2} was experimentally found by Chung *et al.* (2001).

As promising as these enhanced heat transfer microcooler technologies are, it is important to note that an MHP, or two-phase coolers in general, cannot provide additional cooling beyond that of the ambient environment. MHPs often are a part of an integrated system and require additional cooling techniques such as forced air cooling or liquid cooling to minimize condensation resistance and to enhance heat transfer to the surroundings. In addition, there are several engineering and reliability issues, such as charging the two-phase coolant into the device chamber while excluding condensable gases and keeping condensable gases from diffusing into the channels and the refrigerants from diffusing out. The later issue becomes more important as the size is reduced, since the total charge becomes quite small compared with miniature and mesoscale systems, and the surface area-to-volume ratios can become huge.

3.16.3.2.2 Liquid impinging jet and spray cooling

Liquid jet impingement cooling and spray cooling are well-established techniques typically found in applications that need to dissipate highly concentrated heat loads, such as quenching metal alloy parts. For electronics cooling applications, jet impingement or spray cooling with dielectric fluorocarbon coolants have been widely used to maintain relatively low, steady temperature in ultra-high heat flux (above 10 kW cm^{-2}) devices such as lasers, X-ray anodes, and modules in supercomputers (Mudawar 2001, Pautsch and Bar-Cohen 2001). Both jet impingement and spray cooling involve spraying a liquid on the hot surface, which spreads as a thin film and evaporates (Webb and Kim 2005). This active two-phase cooling provides essentially the advantages equivalent to passive two-phase devices but offers higher cooling power (faster heat removal rates). There is a subtle difference between the two techniques in that spray cooling is specifically denoted for the liquid impinging on the surface as small droplets.

Recent studies about the heat transfer characteristics of liquid jet and spray impinging on relatively small areas with a nozzle diameter larger than 1 mm are summarized together with the electronics cooling experiments in Wolf *et al.* (1996) and Mudawar (2001). A heat flux removal of more than 400 W cm^{-2} was reported with a wall superheat of 30–40°C and flow rates of 120 l min^{-1} (Wolf *et al.* 1996), making this technique among the highest reported heat removal method. Since the majority of cooling from a single jet is confined within the impact zone, an array of jets is commonly used to spread out the cooling concentration and achieve relatively uniform cooling over the larger areas of chip surfaces. With a 3×3 array of rectangular jet impingement, Mudawar and Wadsworth (1991) developed a multichip cooling module that was able to supply an FC-72 coolant independently to each chip. The dependence of the CHF on jet width, jet velocity, subcooling, and a distance between nozzles and impact surfaces was also investigated with a conclusion that the CHF was mostly influenced by fluid temperature and the jet velocity (Estes and Mudawar 1995, Ma and Bergles 1986, Mudawar and Wadsworth 1991, Vader *et al.* 1992). Furthermore, the same surface enhancement structures used in pool boiling research can improve jet impingement cooling by offering higher heat transfer rates and lower overshooting of the temperature beyond the boiling temperature (Wadsworth and Mudawar 1992), which typically improves the overall performance.

Apart from the nonuniform distribution of cooling on chips, high-speed jet impingement can be aggressive and damage the heated surface with a large change in momentum between the impinging liquid, which should be avoided for delicate devices (Mudawar 2001). Spray cooling can diminish this impact and flow rate requirement and simultaneously increase the cooling uniformity by distributing droplets over the larger surface area and delaying the liquid separation from the wall during boiling, as depicted in **Figure 17**. However, there are disadvantages with spray cooling, particularly at the microscale, including (1) high pressure drop at the nozzles, requiring high-pressure sources, (2) susceptibility of micronozzel clogging, which impacts the reliability of long-term operation, and (3) difficulties in producing the predictable and repeatable impact patterns (Mudawar 2001). Recent investigations into spray boiling for electronics cooling can be found elsewhere (Estes and Mudawar 1995, Mudawar and Estes 1996, Yang *et al.* 1996).

Typical flow rates associated with macrojets are too large to be practically implemented for integrated VLSI chip cooling system due to the necessary pumping capability and the size constraints. Equipment accompanying jet impingement cooling and spray cooling is rather bulky and difficult to miniaturize. Therefore, as promising as the method is, there are relatively few instances of this technique being used

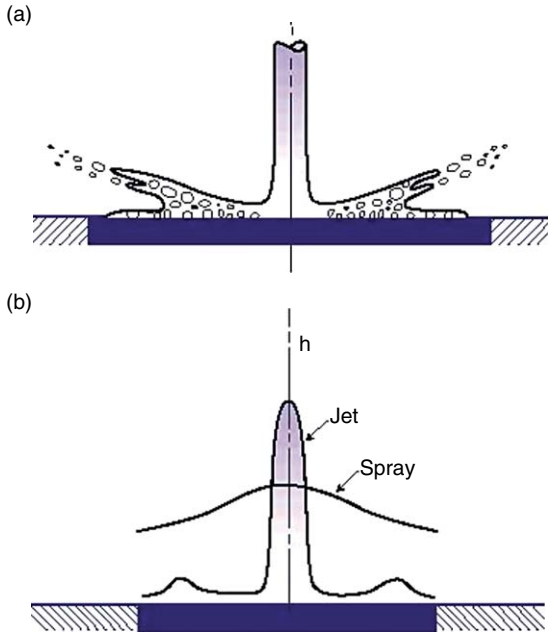


Figure 17 (a) Liquid wall layer splashing and separation at CHF in free circular impinging jets and (b) a qualitative comparison of heat transfer coefficient spatial variation for sprays and free jets. (Source: Mudawar I 2001 Assessment of high-heat-flux thermal management schemes. *IEEE Trans. Comp. Packag. Technol.* **24(2)**, 122–41, (© 2001 IEEE)).

in microcooler technology. A group at the Carnegie Melon University developed a droplet impingement cooling device which they called an embedded droplet impingement for integrated cooling of electronics (EDIFICE), to remove chip heat fluxes higher than $100\,W\,cm^{-2}$ by capitalizing on two-phase cooling with dielectric fluids (Amon *et al.* 2001, 2005, Yao *et al.* 2001). Micron-sized droplets (50–100 μm) of dielectric coolants were dispensed through an array of micronozzles and impinged on the microstructured surfaces to provide a direct cooling for chips in portable and desktop electronics. To make these impingement devices, microfabrication techniques such as DRIE have been extensively utilized to create (1) micronozzles, swirl nozzles, and microinjectors for jet breakup, (2) microstructured surfaces for increasing the spreading of liquid and promoting boiling heat transfer, and (3) electrostatic microvalves to regulate the coolant flow rates. **Figure 18** shows a schematic of the EDIFICE design and how it is connected to the entire system. Similar to the LHP, the vapor generated at the heated component travels through the system to the heat sink or air-cooled condenser, where the condensed coolant returns to the EDIFICE by the micropumps. The heat removed can also be stored in a detachable heat storage unit that contains an organic phase-change material with a low melting temperature of 50 to 60°C (Vesligaj and Amon 1999). A CHF of $42.1\,W\,cm^{-2}$ with HFE-7200 on a microstructured surface 2.5 cm × 2.5 cm in area was demonstrated with a relatively small flow rate of $4.64\,g\,cm^{-2}$ min but with a high pressure of 500 kPa.

In order to eliminate the need for high-pressure pumps and the associated fluid routing, Glezer's group developed vibration-induced droplet atomization (VIDA) coolers utilizing a piezoelectric actuator to vibrate a plate at the resonant frequency, which produces small liquid droplets (diameter 70–800 μm) and propels them onto a heated surface (Heffington *et al.* 2000, 2002). These droplets form a thin liquid film that evaporates on the chip surface. The vapor can be taken to a remote condenser or condensed on the internal surface of the chamber wall, as shown in **Figure 19**. The condensed liquid travels back to the piezoelectric actuator via gravity, which makes this cooler orientation dependent (Heffington *et al.* 2002). Fundamental studies and numerical simulations of VIDA phenomenon are available in James (2003a, b), and the spray characterization in Vukasinovic *et al.* (2004). Over $200\,W\,cm^{-2}$ of heat flux with a 100°C temperature rise was transferred with a chilled water heat sink as the external heat removal device (Heffington *et al.* 2002). To overcome the orientation dependence, a new gravity-independent version of the VIDA coolers was constructed using a piezoelectric diaphragm pump to supply a constant stream of liquid to the VIDA atomizer (Heffington and Glezer 2004). Moreover, the addition of a synthetic jet was made to the new design, as shown in **Figure 19(b)**, accelerating the transport of atomized droplets and the vapor and thus increasing heat transfer characteristics. It was shown that the VIDA–synthetic jet combination was capable of removing heat fluxes as high as $420\,W\,cm^{-2}$ at a surface temperature of 133°C, but it needed on the order of 3 W of input power, which may limit it for many microcooler applications.

To test different types of microfabricated liquid jet coolers, Wang *et al.* (2004) experimentally investigated the performance of single-jet and multijet arrays with an orifice diameter ranging from 40 to 76 μm in a confined geometry for liquid-phase and two-phase heat transfer. **Figure 20** shows a schematic of the microjet device bonded to the microfabricated heater and temperature sensors and images of a jet orifice and a heater die. With DRIE, an array of orifices was micromachined in the jet chip, which

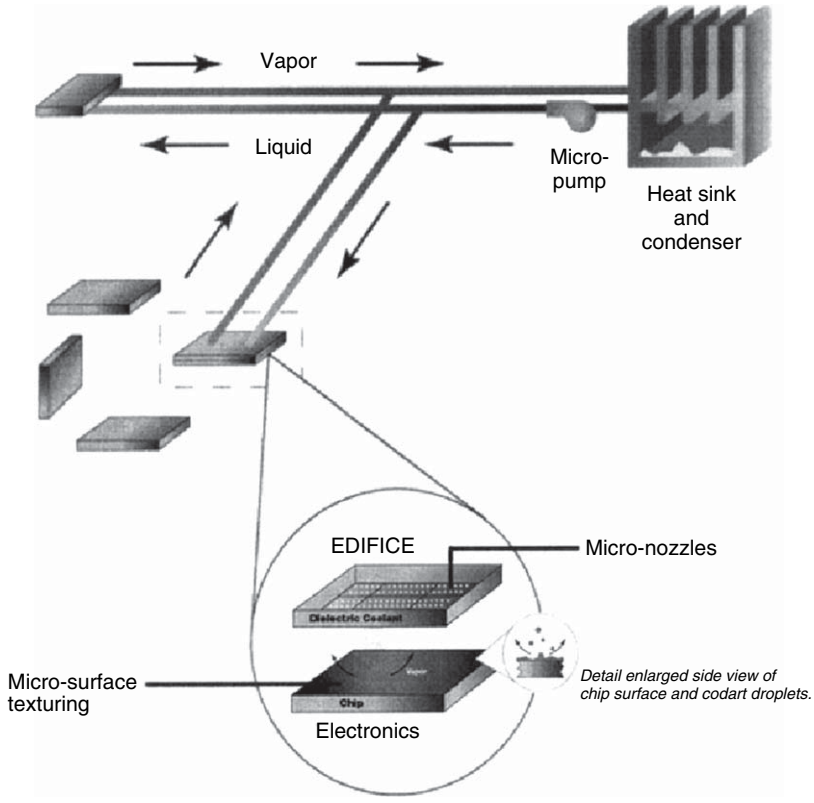


Figure 18 Schematic of EDIFICE, a liquid jet array cooling system, with the overall system. (Source: Amon C H, Yao S C, Wu C F, Hsieh C C 2005 Microelectromechanical system-based evaporative thermal management of high heat flux electronics. *ASME J. Heat Transf.* **127**, 66–75).

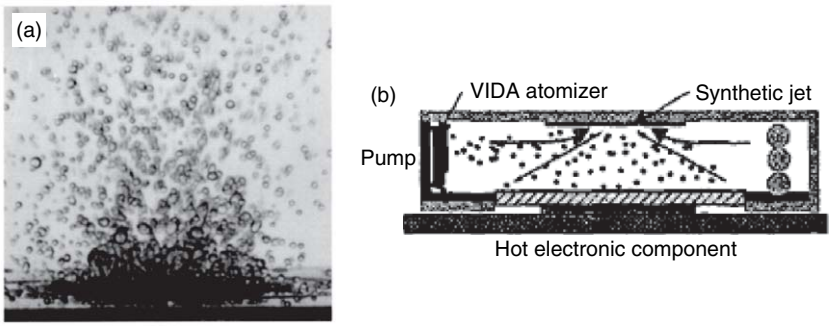


Figure 19 (a) Spray formation during the vibration-induced atomization of a single 0.1 ml water drop forced at 1030 Hz (Source: Reprinted with permission from Vukasinovic B, Smith M K, Glezer A 2004 Spray characterization during vibration-induced drop atomization. *Phys. Fluid* **16**, 306–16, Copyright 2003, American Institute of Physics) and (b) a schematic of a vibration-induced droplet atomization (VIDA) cooler cell with integral synthetic jet. (Source: Heffington S N, Glezer A 2004 Two-phase thermal management using a small-scale, heat transfer cell based on vibration-induced droplet atomization. *Proc. Intersoc. Conf. Therm. Phenom., ITHERM2004*, pp. 90–4, (© 2004 IEEE)).

was anodically bonded to a glass wafer for sealing. A cooling rate of 92 W cm^{-2} with a 100°C temperature rise was achieved with a flow rate of only 8 ml min^{-1} and a pressure drop of 50 kPa. Although not yet demonstrated, a relatively low flow rate requirement in this microjet cooling may permit the integration of a closed-loop electroosmotically pumped cooling system in the near future (Yao *et al.* 2003).

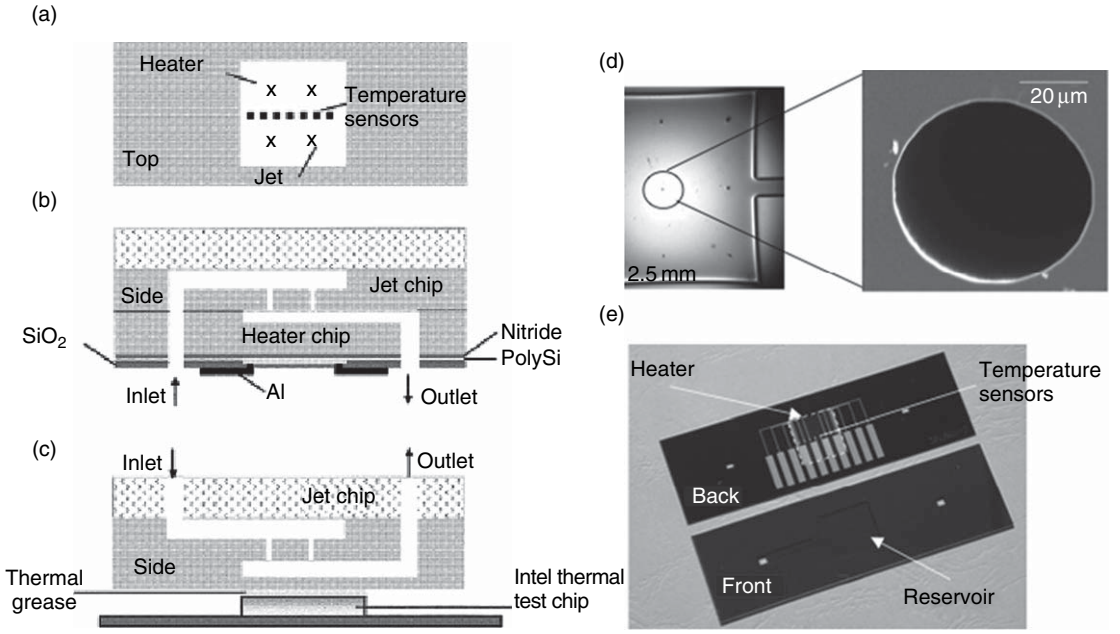


Figure 20 Schematics of (a) the top view and (b, c) cross-sectional views of the bonded jet and heater test devices, (d) a scanning electron micrograph of the 50 μm jet orifice, and (e) a photograph of a fabricated tester. (Source: Wang E N, Zhang L, Jiang L, Koo J-M, Maveety J G, Sanchez E A, Goodson K E, Kenny T W 2004 Micromachined jets for liquid impingement cooling of VLSI chips. *J. Microelectromech. Syst.* **13**, 833–41, (© 2004 IEEE)).

3.16.3.2.3 Microchannel heat exchangers

Not all devices can permit direct liquid contact of the coolant. Therefore, another heavily researched and developed liquid-based heat transfer method is via microchannel heat exchangers (μHEX). Microchannels are typically defined as micromachined channels having hydraulic diameters of 10 μm to a few hundred micrometers. Thus, most liquid flow in microchannels is the low Reynolds number, laminar flow, which in the absence of phase change gives rise to a heat transfer coefficient that is inversely proportional to the hydraulic diameter (Mudawar 2001). Unfortunately, the pressure drop increases with the inverse of the second power of the channel width for a fixed flow rate. Thus, the smaller the channel dimension is, the more the heat transfer is enhanced, but at the expense of a more severe pressure drop and a higher pumping cost. So, unlike passive two-phase heat spreaders, active pumping is needed to drive liquid flow in the μHEX. Therefore, the main limitation in μHEXs has to do with the development of micropumps capable of overcoming the pressure drop. Moreover, small microchannels with a hydraulic diameter less than 200 μm were shown to be easily clogged, hampering the robust operation of the device and thus setting up the practical minimal channel dimensions (Mudawar 2001).

Despite these issues, μHEXs have been heavily reported, in part due to the ability to directly integrate μHEX with silicon chips to maximize heat transfer from electronic devices made in Si. More generally, Tuckerman and Pease (1981) pioneered the development of the μHEX and reported heat removal of 790 W cm^{-2} with a water flow rate of 516 ml min^{-1}, a pressure drop of 207 kPa, and a temperature increase of 71°C using single-phase liquid convection in the microchannels. Since then, tremendous research focusing on μHEX has been conducted with and without boiling. For flow boiling in the microchannels, Bowers and Mudawar (1994) demonstrated 200 W cm^{-2} of heat removal with flow rates less than 65 ml min^{-1} and pressure drops of less than 35 kPa. Comprehensive reviews are available for thermal and hydraulic performance of the single-phase microchannel devices by Ho and Tai (1998), Mehendale *et al.* (2000), Garimella and Sobhan (2003), Garimella and Singhal (2004), and Kandlikar (2005) and for boiling and two-phase flow in microchannels by Ghiaasiaan and Abdel-Khalik (2001), Mudawar (2001), and Bergles *et al.* (2003). Fabrication techniques involved in making

microchannel heat sink are reviewed by Joshi and Wei (2005). Correlations developed for macrochannels have been used to estimate the Nussel numbers, friction factors, and flow transition Reynolds numbers for microchannels, with some noted discrepancies, in particular for two-phase flows (Joshi and Wei 2005, Mehendale *et al.* 2000). Further studies have focused on understanding these discrepancies (Baviere *et al.* 2004, Lee and Garimella 2003, Liu and Garimella 2002, Sharp *et al.* 2000, Xu *et al.* 2000, Zeighami *et al.* 2000), but the conclusive explanations are not yet available (Garimella and Sobhan 2003). In this section, we will review the recent literature published since 2003 for liquid cooling using μHEX.

3.16.3.2.3.(i) Single-phase liquid cooling

High heat flux rates in the range of 1–100 kW cm^{-2} have been achieved with various μHEX using active pumping schemes. Micropumps related to the liquid microchannel cooling will be briefly discussed in a later section. A great deal of recent research on the optimization of μHEX (a single-layered microchannel or a microchannel stack) has been reported to increase the overall thermal performance by minimizing thermal resistances for a given pumping power. Wei and Joshi (2003) reported the optimization study of stacked microchannel heat sinks based on a simple thermal resistance network model. Other optimization studies for the microchannels in single-phase liquid cooling systems have been reported (Khan *et al.* 2006, Koo *et al.* 2005, Park *et al.* 2006, Zhang *et al.* 2005a). The latest review articles for single-phase μHEXs are available elsewhere (Hassan *et al.* 2004, Morini 2004, Schmidt 2003).

The μHEXs developed by companies have also been reported on or made commercially available. Colgan at IBM demonstrated a practical implementation of a single-phase silicon μHEX directly bonded to the back of a chip with low-resistance TIMs (Colgan *et al.* 2005). **Figure 21** illustrates the multilayer stacks of

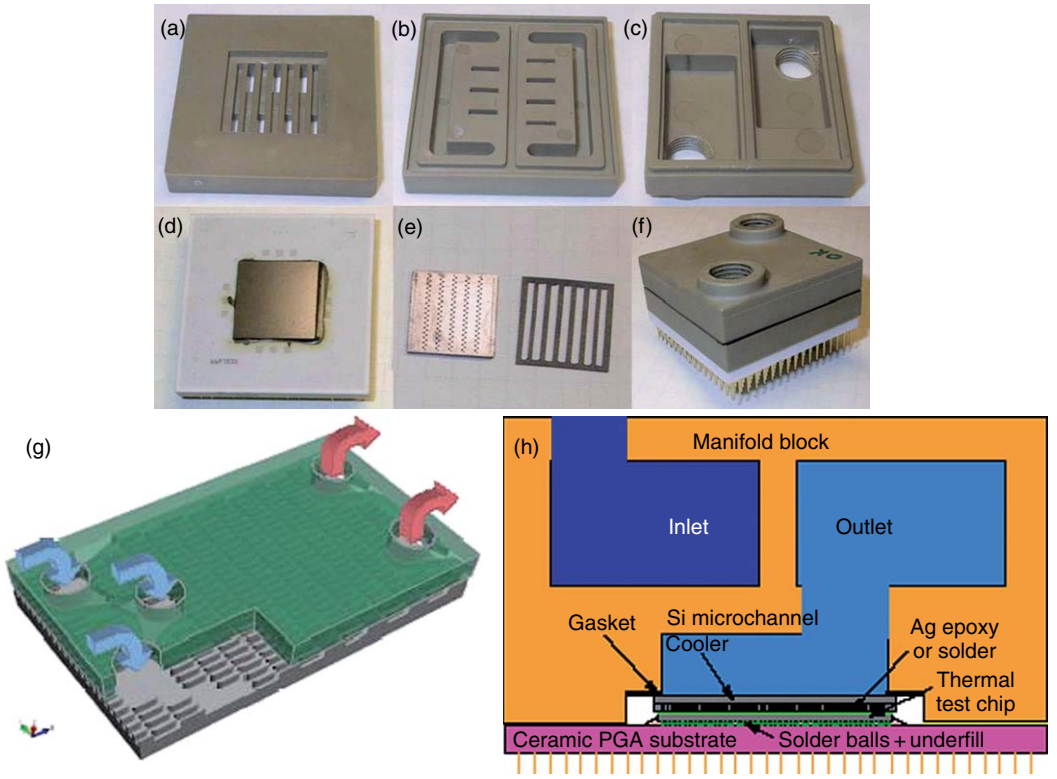


Figure 21 Schematics of microchannel single-chip module cooler. (a–c) Manifold blocks made of a high-temperature plastic, (d) test chip mounting stage, (e) microchannel cooler, and (f–h) a complete microchannel single-chip module. (Source: Colgan E G, Furman B, Gaynes M, Graham W, LaBianca N, Magerlein J H, Polastre R J, Rothwell M B, Bezama R J, Choudhary R, Marston K, Toy J, Wakil J, Zitz J, Schmidt R 2005 A practical implementation of silicon microchannel coolers for high power chips. *Proc. IEEE 21st SEMI-THERM Symp.*, pp. 1–7, (© 2005 IEEE)).

a microchannel single chip module cooler with a 3D and cross-sectional schematic view. Using water, the authors reported cooling power of 300 W cm^{-2} with a reasonable flow rate of 1.2 l min^{-1} and a pressure drop in the μHEX of <35 k Pa. Chang *et al.* (2005) at the Intel Corporation reported the direct fabrication of μHEXs into the back of silicon chips and integrated front-side heaters and temperature sensors to estimate reduction in thermal resistance and pressure drop. The lowest thermal resistance of 0.09 K cm^2 W^{-1} was obtained for a 61-μm-wide and 272-μm-deep microchannel. Some μHEX-based cooling technologies using both single-phase and two-phase flow are commercially available. Several examples of these companies are Mikros Technologies (http://www.mikrostechnologies.com/), Nanolytics (http://www.nanolytics.com/), Advanced Cooling Technologies (http://www.1-act.com/), iCurie (http://www.celsiatechnologies.com/), Lytron (http://www.lytron.com/), Koolance (http://www.koolance.com/), Novel Concepts (http://www.novelconceptsinc.com/), and Microcooling Concepts (MC2, http://www.microcoolingconcepts.com/) to name a few.

3.16.3.2.3.(ii) Two-phase cooling Flow boiling heat transfer and two-phase hydrodynamics have been heavily studied and are currently investigated by many researchers. The Mudawar group experimentally determined hydrodynamic and heat transfer correlations in two-phase μHEX and predicted pressure drops and CHFs for various geometries and applications (Lee and Mudawar 2005a, b, Qu and Mudawar 2002, 2003a, b, c, 2004). Experiments in micromachined silicon channels with a range of 27–171 μm in hydraulic diameter and with varying surface roughness were conducted to investigate the boiling process and two-phase flow behavior such as bubble nucleation, flow patterns, and transient pressure fluctuations (Zhang *et al.* 2005a). Faulkner *et al.* at MicroEnergy Technologies reported a prototype of a 1000 W cm^{-2} cooling system for microwave electronics, employing flow boiling heat transfer in a parallel μHEX. Heat fluxes up to 125 W cm^{-2} for saturated boiling and 280 W cm^{-2} for subcooled (25°C) boiling were removed with water as a coolant and a flow rate of 0.75 l min^{-1} (Faulkner *et al.* 2003). Researchers at Intel studied the effects of nonuniform heating conditions on the thermal performance of the two-phase μHEX to address cooling of local hot spots and observed temperature fluctuation as high as 30°C (Prasher *et al.* 2005).

Thin-film and flexible microchannel heat exchangers were developed by the Shannon group (Pourmohamadian *et al.* 2004, Selby *et al.* 2001). These heat exchangers, shown in **Figure 22**, are made from heat-sealable Kapton (special polyimide films from Dupont) and range from a thickness of 200 to 250 μm. The outer panels, or caps, in the figure are 50 μm thick. Even with a relatively low thermal conductivity (~1 W m^{-1}°C^{-1}) of the Kapton, the thermal resistance is still small (order 5×10^{-5}°C W^{-1}). When the four panels shown in **Figure 22(a)**, which have channels cut for the heat exchanger, manifold, and caps are bonded and assembled into a single unit, the result is a flexible microchannel heat exchanger that can withstand pressures over 18 atm. The maximum pressure is determined by delamination and/or rupture at the critical point shown in **Figure 22(b)**. Under pressure, the polymer channels inflate, increasing the effective cross-sectional area. These flexible heat exchangers were developed for two-phase heat transfer, specifically for R-134a (~13 atm absolute pressure at room temperature). Applications for these types of heat

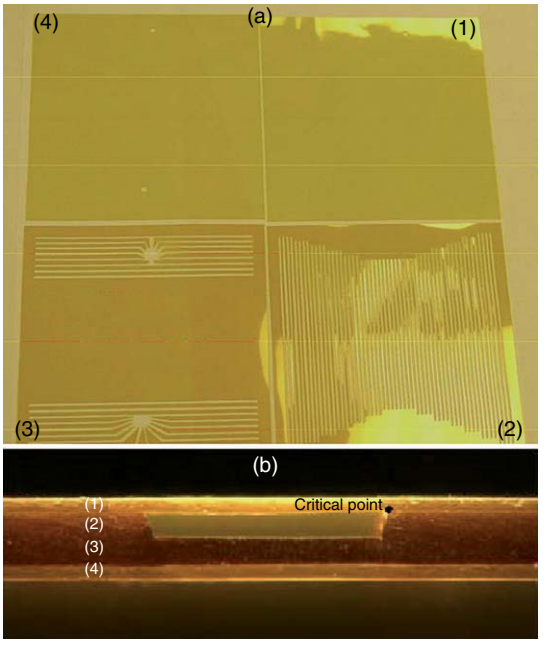


Figure 22 (a) Four polyimide panels comprising the flexible microheat exchangers. (b) Cross-sectional view of the four bonded layers, showing 75 by 750 μm channel. (Permission of the Society of Manufacturing Engineers, Copyright 2001, from the NAMRC Conference Proceedings) (Source: Selby J C, Philpott M L, Shannon M A 2001 Fabrication of mesoscopic, flexible, high pressure, microchannel heat exchangers (MHEx). *Trans. NAMRI/SME, XXIX*, pp. 469–76).

exchangers include insertion into cooling garments for people working in hot environments, and for providing large area, light weight, thin-film heat exchangers for electronics.

3.16.3.2.3.(iii) Liquid metal cooling Liquid metal as a coolant can often be found in high-temperature heat pipes, thanks to metal's high vapor pressure on boiling and high thermal conductivity as a liquid (Faghri 1995), while the density-weighted heat capacity (ρc_p) is smaller, but of the same order, as water, which allows significant heat to be carried per unit volume of evaporated liquid. The thermophysical properties of several liquid metals are summarized in **Table 4**. Recently, the liquid metal-based cooling using magnetofluid dynamic (MFD) pumping has been proposed and commercialized by Nanocoolers (Ghoshal 2005a, Miner and Ghoshal 2004). Due to its high thermal conductivity, $Ga^{68}In^{20}Sn^{12}$ can provide a heat transfer coefficient from the case to the evaporating liquid that is an order of magnitude higher than that of water for laminar flows (Miner and Ghoshal 2004). An MFD direct current (DC) conduction pump utilizing $Ga^{61}In^{25}Sn^{13}Zn^{1}$ (melting point ~8°C) as the working fluid generated a maximum pressure head and a maximum efficiency of approximately 25 kPa and 10%. **Figure 23** shows the typical configuration of a liquid metal cooling loop. Cooling power over 200 W cm^{-2} with high heat transfer coefficients of 20 W cm^{-2} K has been reported with this type of cooler (Ghoshal 2005a). Despite their advantages of being power efficient compared to the electrokinetic-based pumps, no-moving parts, and independence on orientation to gravity/orientation, MFD-based liquid metal coolers also face challenges such as dealing with potential material incompatibilities and the need for nontraditional power supplies because the MFD DC conduction pumps operate at low voltages and high currents.

3.16.3.2.3.(iv) Electrowetting and thermocapillary pumping and cooling Electrowetting refers to the phenomenon that an electric field can modify the wetting behavior of a droplet in contact with an insulated electrode and allows us to control surface tension between the liquid and the solid surface, thus changing the contact angle from hydrophobic to hydrophilic (Oprins and Baelmans 2006). Individual droplet manipulations (merging, cutting, and transporting) in microfluidic applications have already been demonstrated using electrowetting and/or electrothermocapillary actuations (Cho *et al.* 2003, Pollack *et al.* 2000, 2002, Prins *et al.* 2001). The combination of microchannel cooling with electrowetting actuation was first proposed by Pamula *et al.* as a real-time active and smart cooling method where nanoliter-sized discrete liquid droplets are spread or transported to hot spots without external pumps (Paik *et al.* 2004, Pamula and Chakrabarty 2003). Their calculations showed that a heat flux of more than 90 W cm^{-2} can be removed with a flow rate of only 0.4 ml min^{-1}. More in-depth theoretical and experimental studies were followed to understand the oscillating droplet behavior on the microchannel surfaces (Oprins *et al.* 2004, 2005). Taking advantage of the high thermal conductivity of liquid metals, Mohseni (2005) recently proposed digitized electrowetting of liquid metals for active thermal management of high-power microsystems. No experimental results are yet available for heat transfer or cooling from any of these electrowetting coolers.

Table 4 Summary of the thermophysical properties (at 20°C) of liquid metals used in heat transfer applications and wafer for comparison.

	Hg	*$Ga^{68}In^{20}Sn^{12}$*	*$Na^{27}K^{78}$*	*SnPbInBi*	*Water*
Density (kg m^{-3})	13560	6363	868	9230[b]	998
Melting point (°C)	−38.9	10.5	−11	58	0
Heat capacity (J kg^{-1} K^{-1})	139.1	365.8	982.1	209[b]	4181
(ρc_p) liquid (J ml^{-1} K^{-1})	1.89	2.33	0.852	1.93	4.17
Kinematic viscosity (10^{-6} m^2 s^{-1})	0.115	0.348	1.05	4.04[b]	0.96
Electrical conductivity (S μm^{-1})	1.045	3.307	2.878	1.28[b]	5.5×10^{-12}
Thermal conductivity (W m^{-1} K^{-1})	8.717	39[a]	21.8	10[a]	0.606
Prandtl number	0.0248	0.0208	0.0411	0.779	6.62

[a]100°C.
[b]75°C.
Source: Reprinted with permission from Miner A, Ghoshal U 2004 Cooling of high-power-density microdevices using liquid metal coolants. *Appl. Phys. Lett.* 85, 506–8, Copyright 2004, American Institute of Physics.

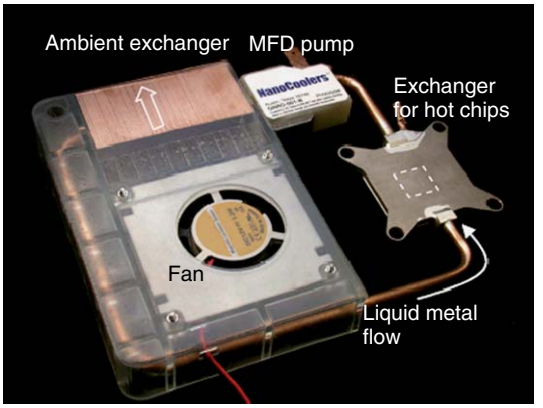


Figure 23 Typical configuration of liquid metal cooling loop for portable mobile applications. (Source: Ghoshal U 2005 Advanced thermoelectric cooling systems. *Proc. Next-Generation Thermal Mngt. Mater. Syst.*, Phoenix, AZ, June 2005 (also see http://www.nanocoolers.com).).

3.16.3.2.3.(v) EHD augmented MHPs MHPs generally suffer from the capillary limit in the performance, just as the macroscale counterpart does. In order to enhance the effectiveness of fluid transport that maximizes heat transfer within heat pipes, Jones (1973) first proposed to use electrostatic fields on a dielectric liquid coolant inside heat pipes to create forces to move the fluid. When the electrical field interacts with the flow field in a dielectric fluid medium, three important forces arise: (1) coulombic forces acting on free charges, (2) dielectrophoretic forces, created by a local change of the permittivity in an electrical field, in this case occurring at the interface between a liquid and a vapor, and (3) electrostriction forces due to the nonuniform electrical field imposed on a dielectric fluid (Suman 2006). Jones and Perry (1974) later conducted the experiments by replacing the capillary wick structure with an electrohydrodynamic (EHD) pump that utilized polarization forces (the dielectrophoretic and electrostriction forces acting on dipoles) to collect, guide, and pump the condensate liquid to the evaporator. Since then, several researchers reported developing EHD enhanced heat pipes at the macroscale (Bologa and Majboroda 1993, Bologa and Savin 1990, Loehrke and Debs 1975); however, none of these experiments showed significant enhancement in heat pipe performance compared to capillary heat pipes.

As heat pipes are reduced in size, mainly to the microscale, EHD pumping becomes more viable. Bryan and Seyed-Yagoobi (1997) reported over 100% enhancement for a monogrooved heat pipe using EHD pumping operating at 20 kV. Hallinan and his coworkers successfully demonstrated a six-fold increase in the heat transport capability of EHD augmented MHP arrays and developed an analytical model to predict the maximum heat transport capability for various electrical field intensities and MHP geometries (Hallinan *et al.* 1998, Yu *et al.* 2002). Mo *et al.* (2000) investigated the application of the EHD technique to the CPL and showed that enhancement, up to three times, could be obtained in heat transfer coefficients and that the depriming conditions of a CPL can be prevented. Suman (2006) recently presented a comprehensive steady-state model for the fluid flow and heat transfer in an EHD augmented MHP and argued that the EHD technique could increase the critical heat input by 100 times.

A different type of the EHD pumping was also investigated to enhance the heat transport capacity of MHPs (Babin *et al.* 1993, Seyed-Yagoobi *et al.* 1995, Yu *et al.* 2000). Instead of employing polarization forces, an ion-drag pump-assisted MHP harnesses the coulombic force exerted on free charges, which are created within a dielectric fluid by emission of the emitter as a corona source. As free ions travel toward the collector via an applied electric field, the working fluid is dragged along because of the friction developed. Both experiments and analytical models show the feasibility of achieving active temperature control and increasing the capillary limit of the heat pipes (Babin *et al.* 1993, Yu *et al.* 2000).

3.16.3.3 Solid-State Cooling

One of the earliest active cooling systems that were developed for miniature and smaller systems is based on solid-state cooling via the thermoelectric and thermionic effect, which are reviewed here, touching on efforts to miniaturize to the microscale.

3.16.3.3.1 Thermoelectric microcoolers

Thermoelectric coolers (TECs), also named Peltier coolers, are small electronic heat pumps capitalizing on the Peltier effect, so that when a temperature change occurs in the vicinity of the junction between dissimilar conductors, a current passes through the junction, resulting in the transfer of heat across the junction, since the heat is carried by the charge carriers. A typical TEC module consists of a number of alternating ingot-shaped n- and p-type semiconductor thermoelements, which are connected electrically in a series with metal connecting strips, sandwiched between two electrically insulating but

thermally conducting ceramic plates, as depicted in **Figure 24** (Rowe 1995). When a positive DC voltage is applied to the n-type thermoelement, electrons flow from the p-type element to the top metal connector, then to the n-type element. As heat is absorbed by electrons at the top junction and that heat energy is carried away from the junction via the Peltier effect, the cold-side (chip) temperature decreases. The electrons carrying heat energy travel to the bottom connector, where excess heat is released through an external heat sink. If the bias is reversed, the device will operate as a heat pump. The advantages of TECs are in general (1) an ability to reduce chip operating temperature locally, (2) the potential for system-level integration with on-board electronics, (3) high reliability (>250 000 h), (4) no moving part (silent), (5) small and lightweight, and (6) no gas or chemicals involved.

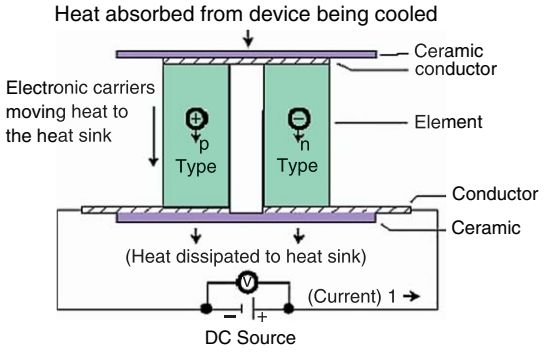


Figure 24 Schematic of a simple Peltier cooler. (Source: Lasance C J M, Simons R E 2005 Advances in high-performance cooling for electronics, *Electronics Cooling* **11**, 22–39, http://www.electronics-cooling.com/articles/).

The thermoelectric performance of a material at a given absolute temperature T is characterized by the nondimensional figure of merit, $ZT = \sigma S^2 T/\lambda$, where S, σ, and λ refer to the Seebeck coefficient and the electrical and the thermal conductivity of the material, respectively. The most challenging task in any thermoelectric application is to find materials that provide low thermal conductivity, to reduce heat losses due to heat conduction between the hot and the cold side, and a high electrical conductivity, to minimize Joule heating and allow for more heat conduction for the applied electric field. For metal and metal alloys, the ratio of the thermal conductivity to electrical conductivity is a constant (Wiedemann–Franz–Lorenz law). Thus, metals with the highest possible Seebeck coefficients ($\sim$10 μV K^{-1}) only produce efficiencies of a fraction of 1%. Through an exhaustive and long journey to find the right material, bismuth telluride (Bi_2Te_3) and its alloys of synthetic highly doped semiconductors were found to possess the highest ZT value (around 1 at room temperature). With these materials, heat pumping rates have been achieved ranging from a few milliwatts to a few tens of watts, and the maximum temperature difference is around 70°C for rejection of heat to a room-temperature reservoir. In **Figure 25**, the theoretical COP and efficiency of thermoelectric coolers and power generators for different ZT values are plotted and compared with other cooling and power generation technologies (Chen and Shakouri 2002). However, TECs made with conventional bismuth telluride compounds ($ZT \sim 1$) cannot compete against the mechanical refrigeration cooling in a large-scale cooling

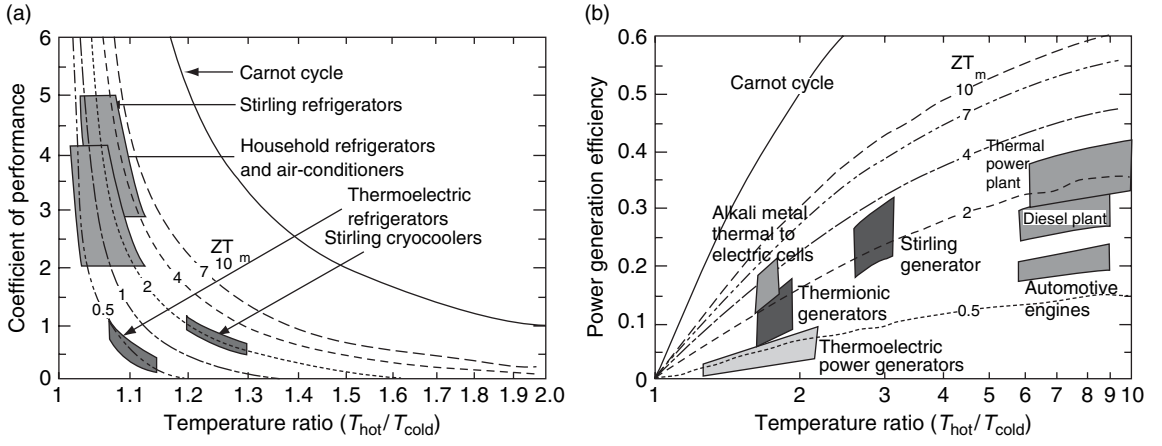


Figure 25 Comparison of thermoelectric technology with other energy conversion methods for (a) cooling and (b) power generation. (Source: Chen G, Shakouri A 2002 Heat transfer in nanostructures for solid-state energy conversion. *ASME J. Heat Transf.* **124**, 242–52).

technology. Despite their low efficiency, TECs are used for applications in areas such as (1) cooling of semiconductor lasers, infrared detectors, charge-coupled devices, blood analyzers, and microprocessors, where accurate control of cooling is required below ambient temperature and (2) small portable refrigerators and picnic coolers (Lasance and Simons 2005).

Along with the low efficiency, the relatively high manufacturing cost of thermoelectric materials has also limited TECs from being widely applied in electronics cooling or as the refrigeration source for consumer product coolers (Phelan *et al.* 2001), although as volumes are increasing and new TEC elements are being created, prices are dropping and more TEC applications are being found. Conventional crystal growth techniques, such as the Bridgeman method, to produce bismuth telluride impose significant limitations on thermoelectric element dimensions due to poor manufacturing yields. The weak bonds holding adjacent crystals together fall apart during wafer processing. A great deal of research has been directed to develop a smaller size of Peltier elements because the cooling capacity of the TEC is inversely proportional to its leg length. A review of the recent industrial efforts can be found elsewhere (Chu and Simons 1999, Lasance and Simons 2005). Bierschenk and Johnson (2005) at Marlow Industries reported new fine-grain microalloyed bismuth telluride materials that can offer high heat capacity over $40\,\mathrm{W\,cm^{-2}}$. Nanocoolers Inc. claims to develop wafer-scale thin-film TEC technology by which the small Peltier coolers are monolithically fabricated (Ghoshal 2005b). A small, flexible thermoelectric module with overall dimension of $16\,\mathrm{mm} \times 20\,\mathrm{mm} \times 0.05\,\mathrm{mm}$ has been made using copper foil as a template (Qu *et al.* 2001). Multiple micro Sb–Bi thermocouple strips are electroplated on a thin Cu foil previously coated with a patterned epoxy layer. As another viable technique, a pulsed layer deposition was used to grow high-quality thermoelectric $Ca_3CO_4O_9$ thin film formed on top of an SiO_x amorphous layer in a self-assembled manner (Hu *et al.* 2005). Bulk thermoelectric material (cubic $AgPb_mSbTe_{2+m}$) with $ZT \sim 2.2$ at 800 K was synthesized using a standard ingot crystallization process, but at room temperature, the ZT is less than 1 (Hsu *et al.* 2004). To develop a microscale TEC device, Snyder *et al.* (2003) developed a novel MEMS-like electrochemical process to fabricate a thermoelectric microcooler device containing 126 n-type and p-type $(Bi, Sb)_2Te_3$ thermoelements, 20 μm in length and 60 μm in diameter with bridging metal interconnects as shown in **Figure 26**. Both cooling and power generation using this device were demonstrated. However, the performance has not yet been optimized partly because the electrodeposited thermoelectric materials have defect structures that effectively reduce their Seebeck coefficients.

In parallel with searching for high-ZT materials, extensive research on the thermoelectric properties of low-dimensional structures has been performed and shows promise in future microcoolers. Pioneering work by Hicks and Dresselhaus (1993) on nanostructured superlattice materials for enhancing the thermoelectric figure of merit started a renewed interest and inspired much of the recent research on the subject. Superlattices are made of alternating thin layers of different thermoelectric materials stacked periodically (Bottner *et al.* 2006). Many bulk materials with relatively good thermoelectric properties have been explored with superlattice TECs: V–VI semiconductors such as Bi_2Te_3/Sb_2Te_3 (Beyer *et al.* 2002, Venkatasubramanian *et al.* 2001), IV–VI semiconductors such as PbTe/PbSe (Beyer *et al.* 2002, Harman *et al.* 2000), IV–IV semiconductor such as Si/Ge (Zeng *et al.* 1999), and V–V semiconductor such as Bi/Sb (Cho *et al.* 2001). Compared to the research in bulk materials that focuses on decreasing the thermal conductivity, nanostructures provide a means to alter both the electron and the phonon transport through the use of quantum and classical size and interface effect (Chen and Shakouri 2002). Outstanding examples of nanostructured materials with high ZT (up to 2.4) were reported with the use of thin-film superlattices (Venkatasubramanian *et al.* 2001) and quantum dot superlattices (Harman *et al.* 2002). Venkatasubramanian *et al.* (1999) used low-temperature organometallic epitaxy to form a Bi_2Te_3/Sb_2Te_3 heterogeneous superlattice structure, with one of the individual layers as small as 10 Å. The anticipated quantum-well-type, heterostructure band diagram with various short periods (10–50 Å) is shown in **Figure 27(a)**, and this ultra-short period superlattice offers significantly higher in-plane mobility while at the same time more phonon backscattering at the interface, thus reducing the thermal conductivity. With a $ZT \sim 2.4$ for their p-type device, the COP shown in **Figure 27(b)** should be comparable to typical mechanical refrigeration systems ($\mathrm{COP} = 2 \sim 4$) when combined with a similar ZT n-type structure, with an estimate of the cooling power density as high as $700\,\mathrm{W\,cm^{-2}}$ at 353 K, which is over 300 times that of the bulk material

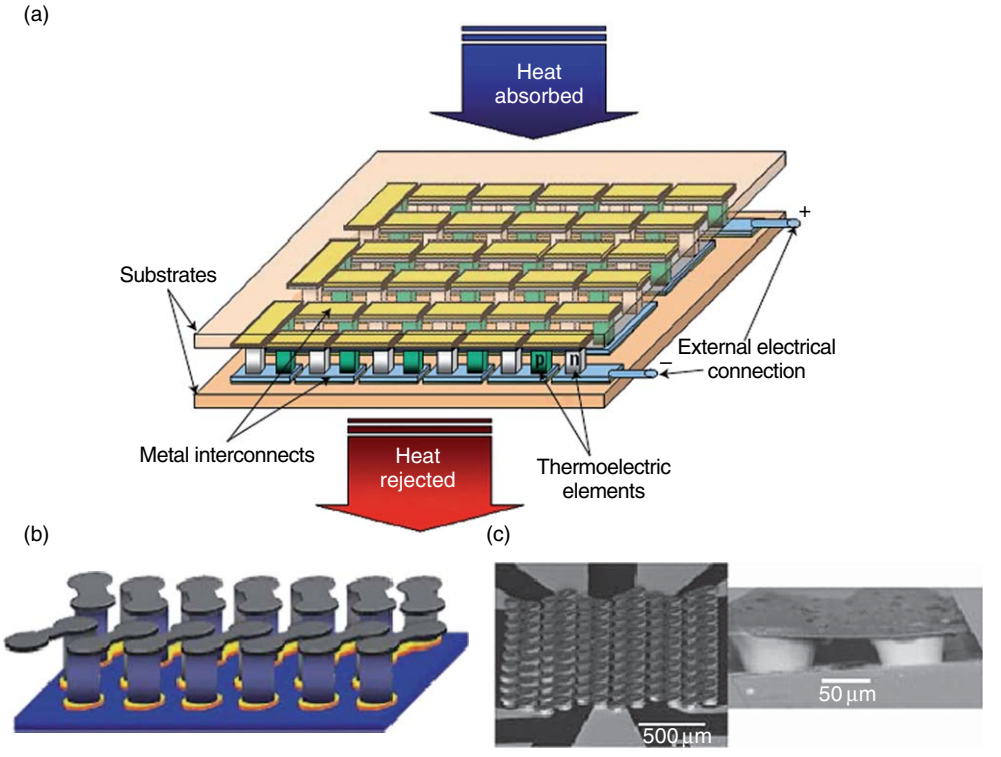


Figure 26 (a) A typical thermoelectric device in which over a hundred n–p couples are connected electrically in series, but thermally in parallel between the hot and the cold side. A schematic (b) and scanning electron micrographs (c) of the electrochemical MEMS-fabricated thermoelectric microdevice. (Source: Reprinted by permission from Macmillan Publishers Ltd: Snyder G J, Lim J R, Huang C-K, Fleurial J-P 2003 Thermoelectric microdevice fabricated by a MEMS-like electrochemical process. *Nat. Mater.* **2**, 528–31. copyright 2003).

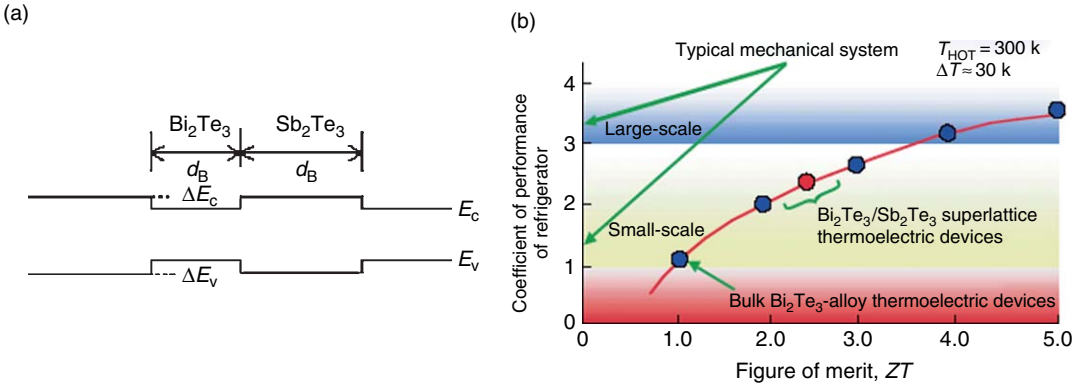


Figure 27 (a) Anticipated band diagram of Bi_2Te_3/Sb_2Te_3 superlattice interface and (b) the potential COP as a function of ZT with other cooling technologies. (Source: Reprinted by permission from Macmillan Publishers Ltd: Venkatasubramanian R, Siivola E, Colpitts T, O'Quinn B 2001 Thin film thermoelectric devices with high room-temperature figures of merit. *Nature* **413**, 597–602. copyright 2001).

(Venkatasubramanian *et al.* 2001). Harman's group (2000) utilized molecular beam epitaxy to grow Bi-doped (n-type) $PbSe_{0.98}Te_{0.02}/PbTe$ self-assembled quantum dot superlattices and showed a significantly higher ZT (~2) than their corresponding bulk materials. It is believed that the enhancement in ZT values may result from a delta-function-like feature in the electron density states, increased phonon scattering (Harman *et al.* 2002), and possibly electron energy filtering (Shakouri 2004). Very recently, Zhang *et al.*

(2006a) demonstrated a 3D silicon microrefrigerator that could cool at a maximum of 1.2°C at room temperature and an easy integration of a microcooler into a silicon chip to remove hot spots.

A direct profiling of the Seebeck coefficient, *S*, across a semiconductor p–n junction with nanometer resolution was investigated using scanning thermoelectric microscopy in order to better understand the effects of low dimensions and nanoscale structures on *S* (Lyeo *et al.* 2004). Thorough knowledge of nanostructures' dependence on Seebeck coefficients and thermal and electrical conductivity will help to design and optimize these superlattice thermoelectric coolers. Theoretical treatments with rigorous reviews on the progress of low-dimensional thermoelectric materials research can be found elsewhere (Böttner *et al.* 2006, Chen 2006, Chen and Shakouri 2002, Chen *et al.* 2003, DiSalvo 1999, Shakouri 2004, Tritt 2001). As an example of the progress, the Majumdar group at the University of California, Berkeley, have recently investigated the thermoelectric properties of metal–molecular junctions with scanning tunnelling microscopy and introduced the possibility of developing inexpensive and efficient molecular TECs (Reddy *et al.* 2007).

3.16.3.3.2 Thermionic microcoolers

The concept of thermionic energy conversion is relatively new and is based on the concept that a hot, high work function cathode in contact with a heat source will emit electrons, which are absorbed by a cold, low work function anode separated by a vacuum gap, resulting in a current flow across a voltage potential caused by the work function difference (Hatsopoulos and Kaye 1958). When the anode and the cathode are connected through an external load, electrons flow and electric power can be measured to calculate the efficiency. A thermal efficiency of about 13% was reported with a cathode held at higher than 1500 K. Mahan (1994) proposed a vacuum diode-type refrigerator based on the same principle by imposing a current in the opposite spontaneous current direction. Mahan pointed out that thermionic refrigerators would not work at room temperature because of a high work function of a typical cathode. The minimum work function currently available for a cathode is over 0.7 eV, for which the refrigerator would work only at temperatures above 500 K with the theoretical efficiency of over 80% of the Carnot value. Another practical limitation is that space charge can also create an additional potential barrier imposed by free charges in the space between the anode and the cathode, thereby diminishing the thermionic current. Since this seminal work, many researchers have attempted to create efficient thermionic refrigerators running at room temperature.

Since no material has a work function less than 0.3 eV for a viable room-temperature thermionic refrigerator in a metal–vacuum–metal system, metal–semiconductor or semiconductor–semiconductor layers have been considered to make lower-temperature thermionic systems (Mahan and Woods 1998, Shakouri and Bowers 1997). Shakouri and his collaborators at the University of California, Santa Cruz, and Bower at the University of California, Santa Barbara, proposed heterostructure integrated thermionic (HIT) coolers that use semiconductor heterolayers with a smaller barrier height (0.1–0.4 eV) determined by the band edge discontinuity (Shakouri and Bowers 1997). The calculation based on nonlinear transport regime of $Hg_{1-x}Cd_xTe$ heterolayers with a barrier thickness of about 1 μm can achieve room-temperature single-stage cooling of 20 to 40°C and cooling power densities over 100 W cm^{-2} (Shakouri and Bowers 1997, Shakouri *et al.* 1998). Experimental investigation of HIT cooling was performed by fabricating a single InGaAsP barrier (~1.3 μm thick and 0.1 eV) surrounded by an n^+-InGaAs and lattice-matched to InP substrate (Shakouri *et al.* 1999a). Cooling of 1° at 80°C with cooling power densities exceeding 50 mW cm^{-2} were observed. The monolithic integration of this cooler with optoelectronic devices was demonstrated, and it was shown that the performance of the single-barrier HIT device with monolithic structures was limited by the nonideal effects, such as Joule heating and the metal–semiconductor contact resistance (LaBounty *et al.* 2000, 2001). Mahan and Woods (1998) proposed to utilize thermionic emission in the multilayer structures, but focusing on the use of the Schottky barriers between some metals and semiconductors. A barrier height less than 0.1 eV can be made for the Schottky barriers such as InGaAs/Au (Kajiyama *et al.* 1973), InSb/Au, and EuPbTe/PbTe (Hicks *et al.* 1996). The authors argued, based on their analyses, that a multilayer is necessary to reduce the thermal conduction loss by maintaining a small bias and thus a small temperature difference (<1 K) for each layer and that their ZT values increased by a factor of 2 from the bulk counterpart.

In thin-film heterostructure devices, thermionic cooling differs from thermoelectric cooling in that the selective emission of hot electrons, equivalent to evaporation cooling, brings about strong electron–electron and electron–phonon interactions to restore

the quasiequilibrium distribution in the cathode by absorbing energy from the lattice, subsequently cooling the emitter junction (Shakouri and Bowers 1997). A larger thermoelectric power factor (σS^2) than that of bulk materials can be obtained, thanks to an increase in the effective Seebeck coefficient by selective emission while maintaining high electrical conductivity. Further investigations showed that in the superlattice structure, the thermoelectric power factor of the TEC devices is higher than that of thermionic devices and that its ZT enhancement is associated with reduced lattice (or phonon) thermal conductivity rather than improved electronic transport (Radtke *et al.* 1999, Vining and Mahan 1999). Ulrich *et al.* (2000) also concluded that for all known materials, single-barrier thermionic refrigeration is less efficient than TEC refrigeration.

However, although high electrical conductivity and low thermal conductivity are still necessary for an HIT cooler to work, the requirement of high bulk Seebeck coefficients is relaxed in thermionic cooling, which broadens the selection of materials and corresponding fabrication processes, making HIT still potentially useful for microcoolers. It is also important to note that the HIT coolers with SiGe or III–V semiconductors are not intended to provide highly efficient and macroscopic chip cooling but to selectively cool local hot spots in the chip, where the efficiency of the microcooler is no longer important and the small-size refrigerator with minimal additional thermal resistance is required (Shakouri and Zhang 2005). Many multilayer HIT coolers for high–heat flux optoelectronic devices and hot spots in microprocessors have been demonstrated using standard IC fabrication techniques with conventional semiconductor materials such as SiGe/Si (Fan *et al.* 2001a) and SiGeC/Si (Fan *et al.* 2001b). Zeng *et al.* experimentally investigated the cooling power density of SiGe/Si HIT microcooler consisting of a 3-μm thick 200 × (5 nm Si0.7Ge0.3/10 nm Si), MBE-deposited superlattice films by integrating thin-film resistor heaters (also serving as resistive temperature detectors (RTDs)) on the cooling surface, as shown in **Figure 28(a)** (Shakouri and Zhang 2005, Zeng *et al.* 2004, Zhang *et al.* 2004). Net cooling of over 4.1 K with a cooling power density of 600 W cm^{-2} for 40 μm × 40 μm devices were reported at room temperature. A more recent study on a similar HIT microrefrigeration system (**Figure 28(b)**) for cooling localized hot spots demonstrated a maximum cooling of 7 K with a maximum cooling power density of over 500 W cm^{-2} at 100°C (Fukutani *et al.* 2006). A noncontact thermoreflectance technique was also used to map the temperature distribution on top of the microcoolers as shown in **Figure 28(c)**. Fast transient response less than 40 μs was recently demonstrated with these microcoolers for on-chip high speed localized cooling (Zhang *et al.* 2006a).

A recent breakthrough in HIT microcooler experiments from Shakouri's group is related to the use of 3D structured thermoelectric elements. When HIT elements are combined with a tall barrier, highly degenerate superlattice structures such as multiple quantum wells and quantum dots, thermoelectric power factors an order of magnitude higher than the bulk values can be potentially achieved (Shakouri *et al.* 1999b). A detailed theoretical analysis of electron and thermoelectric transport perpendicular to heterostructure superlattices revealed that the high performance of metal-based superlattices with tall barriers occurs from the nonconservation of lateral momentum during the thermionic emission process (Vashaee and Shakouri 2004a, b). Analysis based on a Monte Carlo transport model also supported that a nonplanar potential barrier structure in HIT can achieve a much larger emission current, increasing the thermoelectric figure of merit (Bian and Shakouri 2006).

Along these lines, modifications to existing thermionic microcoolers have been discussed in the literature to improve cooling. Xuan (2002) suggested replacing the passive metal conductor in a vacuum thermionic refrigerator with an active p-type thermoelectric element to form a combined thermionic–thermoelectric refrigerator. More efficient and higher cooling power than either thermionic cooler or TEC was predicted for ideal materials. Mal'shukov and Chao (2001) suggested an optothermionic refrigerator, combining the ideas of the laser cooling and thermionic cooling, which consists of a narrow quantum well embedded in the middle of a p–n junction. Electrons in the n-doped region and holes in the p-doped region are thermally excited into the quantum well and recombined to emit light, extracting heat from the system. A quantitative analysis of competing events of cooling by radiative recombination and heating by nonradiative recombination at the junction showed that a cooling rate of a GaAs/AlGaAs optothermionic refrigerator was calculated to be several watts per square centimeter. Further analysis suggested that a cooling rate as high as 10 W cm^{-2} is possible with the same cooler system (Han *et al.* 2006). Experiments are still needed to validate this idea.

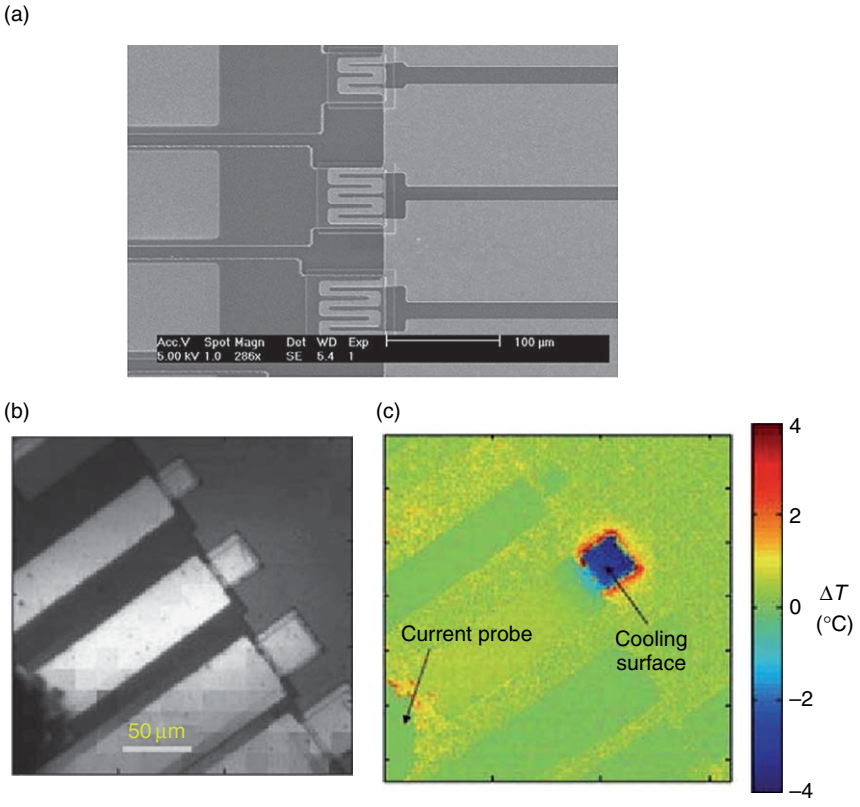


Figure 28 (a) Scanning electron micrographs of SiGe/Si HIT microcoolers integrated with thin-film resistive temperature detectors (RTDs). (Source: Zeng G, Fan X, LaBounty C, Croke E, Zhang Y, Christofferson J, Vashaee D, Shakouri A, Bowers J E 2004 Cooling power density of SiGe/Si superlattice micro refrigerators. *Mater. Res. Soc. Symp. Proc.* **793**, S2.2.1–2.2.7), (b) scanning electron micrographs of SiGe/Si HIT microcoolers, and (c) temperature distribution of (b) showing uniform local cooling. (Source: Fukutani K, Zhang Y, Shakouri A 2006 Solid-state microrefrigerator on a chip. *Electronics Cooling* **12(3)**, http://www.electronics-cooling.com/articles/).

Finally, a brief description of thermionic tunneling (or so-called thermotunneling) is useful for this discussion of solid-state microcoolers. The idea is to use electron tunneling at nanometer-thick vacuum gaps for thermionic microcoolers to operate at room temperature. To ensure only hot electrons (originating above the Fermi level) participate in tunneling, the barrier height can be adjusted by size and bias (Hishinuma *et al.* 2001) or by using heterogeneous structures (Hishinuma *et al.* 2002). Preliminary experiments, however, did not show substantial cooling (<1 mK) and posed significant fabrication challenges, such as maintaining uniform nanometer gap sizes and atomically clean electrode surfaces (Hishinuma *et al.* 2003). An increase in the emission current by 10 orders of magnitude must be accomplished for practical cooling capabilities of 10–1000 W. Other approaches were discussed to increase the thermionic emission currents using resonant tunneling or the inverse Nottingham effect (Korotkov and Likharev 1999, 2000, Tsu and Greene 1999), as well as sharp nanostructured tips like nanotubes (Fisher and Walker 2002, Fisher *et al.* 2003). No experimental results on cooling have yet been reported at near room temperature. In addition, significant improvements in developing lower work function materials should be made to capitalize the nanogap vacuum thermionic emission.

3.16.3.4 Mechanical Refrigeration Methods

While a majority of work in microcoolers has focused on solid-state refrigeration, the pursuit of building the mesoscale/microscale mechanical refrigerators has continued in both academia and industry due to the potential for mechanical systems to have high COP and to move relatively large heat fluxes. Unlike the solid-state cooling systems just covered, which have relatively low COP at the normal scale, but offer potentially higher COP at the

microscale, mechanical cooling systems regularly achieve high COP at the normal scale but have not yet succeeded to do so at the microscale. In terms of development effort and maturation, microscale mechanical coolers are still a nascent technology. Miniaturization efforts using micromechanical machining and MEMS processes have been reviewed for energy systems in general (Ameel *et al.* 1997) and for refrigerators or cryocoolers (Burger 2001, Phelan *et al.* 2002). The performance and feasibility of various miniaturized mechanical refrigeration systems are compared using a simple thermodynamic model (Phelan *et al.* 2003). No complete mechanical refrigeration system has been demonstrated in the length scale of interest here, but we will discuss the basic working principles and preliminary attempts to build such systems. In addition, we will cover outstanding research issues that need to be resolved before mechanical microcoolers find wide acceptance.

3.16.3.4.1 Active two-phase cooling systems

Even though heat flux spontaneously moves from hotter to colder reservoirs, higher heat fluxes can be achieved using active pumping schemes to increase fluid flow rates. Goodson *et al.* developed a closed-loop, two-phase microchannel cooling system to remove high heat fluxes in VLSI circuits (Jiang *et al.* 2002, Tripp *et al.* 2004). Utilizing an electroosmotic micropump made of a porous glass frit, which produces a maximum pressure of 160 kPa and flow rate of 7 ml min^{-1}, and microchannel heat exchangers, the cooling system can remove 38 W with a pump power of 2 W. Cooligy Inc. has developed a closed-loop liquid cooling system based on this electroosmotic pumping principle that can handle heat fluxes greater than 500 W cm^{-2} for desktop computers and servers (Upadhya *et al.* 2003, 2004, Zhou *et al.* 2004). **Figure 29** shows the schematic of their recent prototype cooling system presented recently (Upadhya *et al.* 2006).

3.16.3.4.2 Vapor compression refrigerators

A detailed description of the vapor compression refrigeration is available in any basic thermodynamics textbook. All vapor compression refrigerators are made of four main components: a compressor, a condenser, an expansion orifice, and an evaporator. Briefly, a refrigerant (e.g., R-134a) enters the evaporator as a two-phase (but mostly liquid) liquid–vapor mixture, and as the refrigerant absorbs heat from the cold region, it changes phase from liquid to vapor, leaving the evaporator as a saturated vapor. The refrigerant is then compressed adiabatically and sent to the condenser, where the refrigerant rejects heat to the warm region and changes phase from a saturated vapor to a saturated liquid. Finally, the refrigerant passes through an orifice (or expansion valve), expands to a two-phase mixture, and enters the evaporator, thus completing a cycle. The advantages of vapor compression refrigerator are its (1) capability of removing large quantities of heat with a small mass flow of refrigerant, (2) high efficiency, arguably one of the most efficient refrigeration systems at the macroscale, producing high COP, and (3) the capability of achieving subambient temperature without injecting additional pumping energy at the cold junction that too must be removed.

Unfortunately, vapor compression refrigerators typically perform less efficiently as their sizes become smaller. Since the cooling capacity scales with the volume of gas while many loss mechanisms, such as friction and heat loss, scale with the surface area, losses tend to begin to dominate as the surface-to-volume ratio increases and the size shrinks, making efficient microscale mechanical-based refrigerators difficult to achieve (Shakouri and Zhang 2005). A few companies attempted to develop

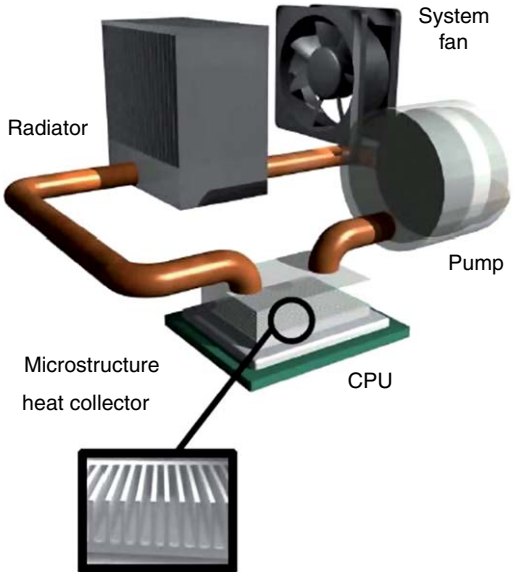


Figure 29 Schematics of the microscale liquid cooling system from Cooligy, Inc. (Source: Upadhya G, Munch M, Zhou P, Horn J, Werner D, McMaster M 2006 Micro-scale liquid cooling system for high heat flux processor cooling applications. *IEEE 22nd SEMI-THERM Symp.*, San Jose, CA, USA, pp. 116–19. (© 2006 IEEE)).

relatively small vapor compression refrigerators for electronics cooling, but their sizes are rather big and considered more suitable for larger, less portable applications. IBM employed refrigeration cooling in the large server (IBM S/390 G4 CMOS system, first shipped in 1997) to improve the system performance. The size of the cooling unit is 26 cm × 26 cm × 70 cm with a weight of 27 kg (Schmidt 2000, Schmidt and Notohardjono 2002). KryoTech was one of the first companies to have introduced to the commercial market a vapor compression refrigerator that can be incorporated to cool an AMD CPU chip down to −40°C or 25°C (Peeples 2001). As shown in **Figure 30**, the refrigeration system is made to be installed below the desktop tower, resembling a typical refrigeration unit in a dorm room. The evaporator of the refrigerator is placed on the high-speed microprocessor to maintain it at the desired temperature and allow overclocking far beyond what is possible with conventional air/liquid cooling. Fujitsu Ltd. in Japan also presented a prototype dual-loop vapor compression refrigerator mounted on CPU-MCM package for a heat load ranging from 250 to 2500 W (Wei *et al.* 2003a, b). Despite the superior cooling performance, these vapor compression-type refrigeration systems for commercial electronics cooling are not yet popular in the market because of the issues of packaging, cost, noise, and reliability of mechanical systems.

More analyses and experimental investigations continued to improve the miniature vapor compression refrigeration systems for electronics cooling application (Bash 2001, Heydari 2002, Trutassanawin *et al.* 2006). Chiriac proposed constructing a miniaturized mechanical refrigeration system onto the printed circuit board by integrating commercially available, miniature components and analyzed each component (evaporator, condenser, compressor, and valve) to optimize the COP of the overall system (Chiriac and Chiriac 2005). Using R-134a refrigerant, the refrigeration system was optimized for cooling power ranging from 20 to 100 W, with the COP up to 4.5. However, the size of these refrigeration systems for electronics cooling is too large, putting them outside of the micro-cooler realm.

The main challenges in miniaturizing vapor compression refrigeration systems for designs based on normal-scale systems have to do with component miniaturization and integration. Especially the development of a small, yet efficient, gas compressor is the most crucial factor in all the mechanical refrigeration systems. Based on the qualitative calculations of the entropy generation rates for each component, Jeong (2004) has argued that the mechanical compressor due to frictional and heat transfer across gradients is the most difficult component to be miniaturized in a vapor compression refrigerator. Different alternatives of actuation in a mesoscale compressor were discussed by Phelan *et al.* (2004). Our group and others have attempted to build a micropump that can pump a refrigerant (gas phase) at high pressure ratio (Laveau *et al.* 2000, Shannon *et al.* 1999). A group at the University of Central Florida employed electrostatic actuation integrated with centrifugal

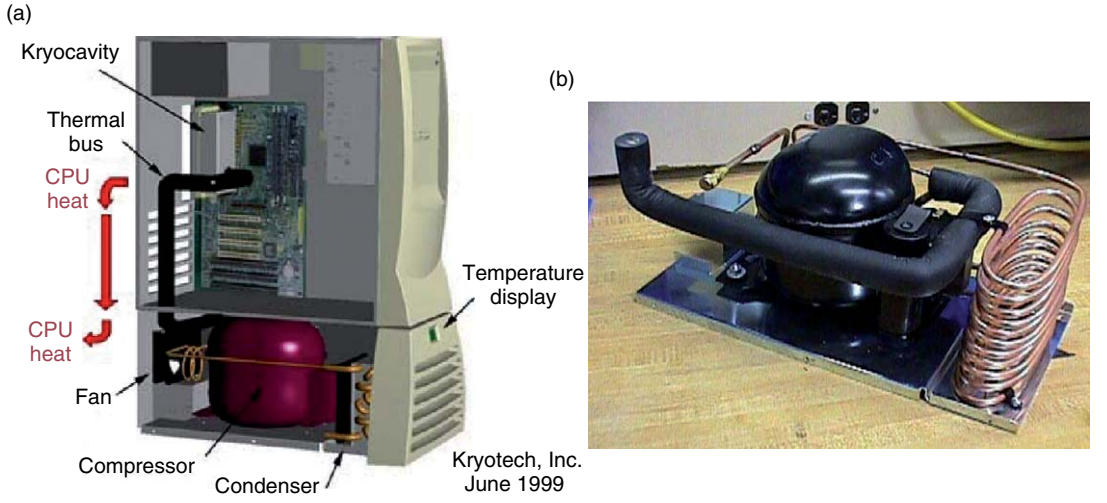


Figure 30 (a) KryoTech's Cool Athlon 800 MHz and (b) KryoTech's Renegade ATX-SE. (Source: Peeples J W 2001 Vapor compression cooling for high performance applications. *Electronics Cooling* **7(3)**, http://www.electronics-cooling.com/articles).

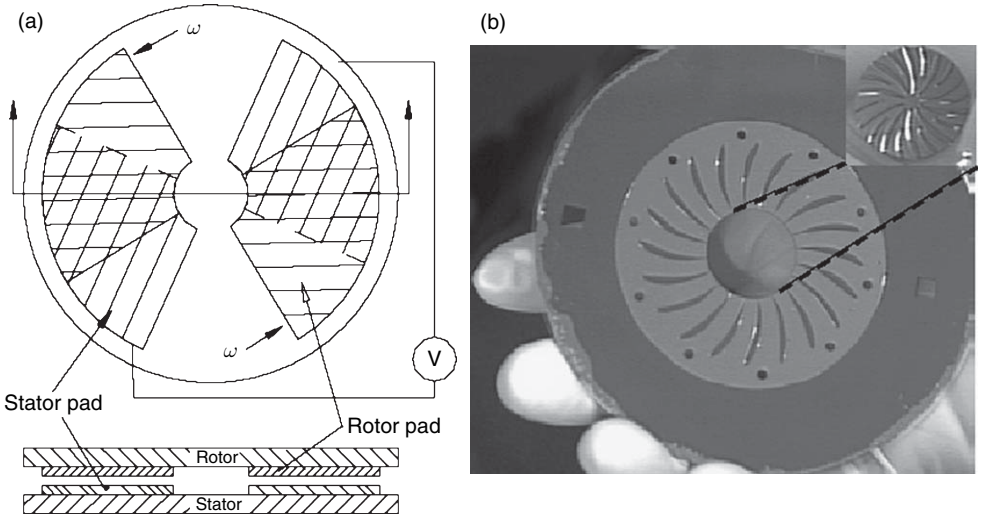


Figure 31 (a) A schematic of a simple variable capacitance motor and (b) an impeller micromachined in a silicon wafer. (Source: Agrawal R, Hasan Q, Ashraf N, Sundaram K B, Chow L C, Kapat J S, Vaidya J 2003 Design and fabrication of meso-scale variable capacitance motor for miniature heat pumps. *J. Micromech. Microeng.* **13**, 1–7).

compression and built a three-phase variable capacitance motor (VCM) by means of the standard Si MEMS technologies (Agrawal *et al.* 2003, Carter *et al.* 1999, Laveau *et al.* 2000). The intended application of their technology is to integrate a heat removal system for microelectronics with an actively cooled jacket. As shown in a schematic of a simple VCM in **Figure 31(a)**, the top rotor pads freely rotate with low friction slightly above the bottom stator pads and experience fluctuation of electrostatic force parallel to the surface of the plates as two conducting plates are displaced with respect to each other and change overlap areas. By switching on and off the excitation to the stator pads, the rotor continues to rotate either by electrostatic attraction or by inertia. The impeller and the diffuser of a compressor unit were fabricated on silicon and are shown in **Figure 31(b)**. The size of the compressor is about 7.5 cm in diameter. Although some electrical (capacitance) measurements showed a feasibility of the method, no pumping/compression data have yet been reported.

The development of a mesoscale flexible vapor compression refrigeration system based on a micro-scale compressor was attempted at the University of Illinois by Shannon *et al.* (1999). As seen in **Figure 32**, the entire system fits within a 100 mm × 100 mm by a 2.5-mm thick structure. An electrostatic compressor shown in **Figure 33** operating on touch capacitance was developed in an attempt to produce a microscale pump that could compress R-134a to 8.5 atm. The underlying principle of touch capacitance is shown in

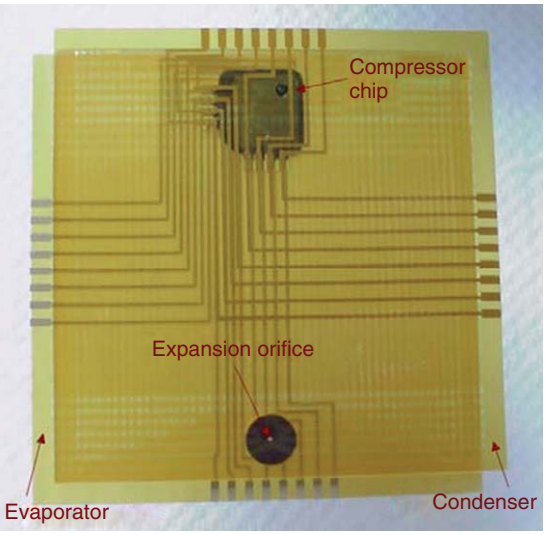


Figure 32 A picture of an integrated mesoscale vapor compression cycle cooler. The heat exchangers are 100 × 100 mm wide and 250 μm thick. The expansion orifice is a drilled alumina pore encased in steel, and the compressor chip is electrostatically actuated.

Figure 34. For very thin dielectric layers (t_1 and t_3 of order 0.2 μm), the electrostatic force per unit area (F_E/A) that compresses the refrigerant at a pressure, P, can be high ($F_E/A > 10$ atm) for relatively modest voltages (~100 V). The ultimate maximum pressure is determined by the electrostatic breakdown strength and the strain energy that needs to be overcome in the membrane. This type of electrostatic

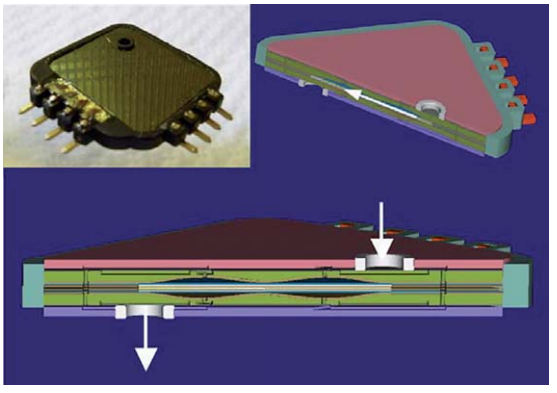


Figure 33 A picture of compressor chip (*upper left*), with cross-sectional views showing the fluid flow. The cross-sectional views are enlarged in the thickness to see internal features. The overall packaged size is 22 mm × 22 mm × 2 mm thick.

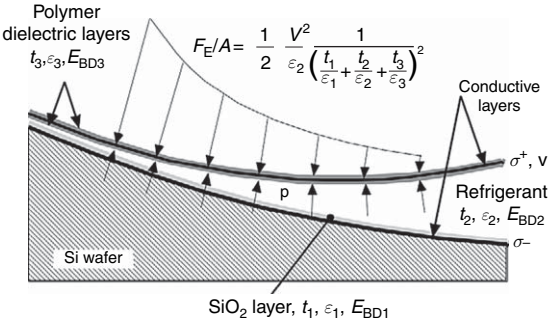


Figure 34 The operating principle underlying touch-capacitance compression. Refrigerant vapor is compressed between a movable polymer membrane and the silicon wafer as an electrostatic force on application of a voltage, V, pulls the two conductors together, zipping the membrane into contact with the wafer. The electrostatic force per unit area F_E/A is at a maximum when the gap, t_2, between the membrane and wafer is zero.

actuation improves with scaling to the microscale, since the amount of PV work is delivered from the total contact area, and thus improves with higher surface area to volume ratios. Moreover, the touch-mode actuation does not generate any sliding or frictional losses. The primary loss mechanism is via electrical resistance and heat transfer across the membrane. While a promising technique for compression, this method can be used to achieve a maximum pressure of less than 2 atm, due primarily to bubble modes that occur from nonaxisymmetric mode compression and charge transfer through the membrane that causes charge shielding, which can dramatically reduce the effective electric fields doing the compression work. Research has been ongoing to solve both the compression modes and charge shielding (Wu and Shannon 2004).

3.16.3.4.3 Stirling refrigerators

A Stirling refrigeration cycle is the most general case of the regenerative cooling cycle in which reversible work is performed on the warm surroundings when heat is absorbed at the cold side of the system. Depending on the methods of compression, expansion, and movement of the regenerator, variations from a Stirling cycle were invented, such as the Gifford–McMahon, Vuillemier, and pulse tube. In a Stirling refrigerator, compression, displacement, and expansion of a gas are performed using two reciprocating pistons and one displacer (regenerator), which can pose significant challenges in miniaturizing the system. These types of the regenerative refrigeration cycles are often found in the spacecraft, for example, cooling detectors of thermal imaging systems. Peterson investigated the lower size limit of a Stirling refrigerator and suggested that a refrigerator made of a low thermal conductivity material can scale down to under 1 cm (Peterson and Al-Hazmy 1997). Organ (1999) developed a comprehensive model to simulate the performance of the Stirling refrigerator. Sunpower Inc. patented a silicon-based micromachined free-piston cryocooler. As with other mechanical refrigerators, a complete Stirling refrigerator with the length scale of interest here is not available to the authors' knowledge.

3.16.3.4.4 Joule–Thompson refrigerators

Joule–Thompson (J–T) refrigerators are a broad category of the refrigerators capitalizing on cooling that occurs with the expansion of the high-pressure gas, utilizing the J–T effect. Typically, high-pressure gas (either from a cylinder or from a compressor) passes down a heat exchanger at cold side through a porous plug or expansion capillary, and this process can be combined with other thermodynamic cycles. One of the simplest modes of J–T refrigerators is to use Linde–Hampson (L–H) regenerative cycle, in which a cooled gas from a J–T valve is rerouted to a counterflow heat exchanger to precool the incoming high-pressure gas. The J–T refrigerator with its variations is one of the most popular methods to provide low-temperature cooling of electronic components such as IR detectors, SQUID, and IC chips (Ng *et al.* 2002). In order to obtain acceptable efficiencies, J–T coolers must be operated with very high pressures (200–500 atm). Therefore, many J–T

coolers employ an open-cycle operation using a high-pressure gas cylinder in spite of cost and inconvenience. Closed-loop operation of the J–T devices is conceivable but requires the high-pressure compressors that can produce a pressure ratio of 100–1000. This type of compressor is extremely difficult to make and expensive even at the macroscale, and therefore, miniaturizing these compressors to build a completely miniaturized J–T refrigerator has proved to be extremely difficult. Recently, electroosmotic microscale pumps have been developed with reported pressures of 0.35 MPa that have the potential to go much higher in pressure (Tripp *et al.* 2004). However, this is pumping of liquid and there would need to be a liquid-to-gas converter. The pumping rates are also very small (on the order of 0.01 ml min^{-1} at ~0.3 MPa) and would be even smaller at higher pressure. Efforts have also been made to reduce the size of the J–T devices in open-cycle operation because the small gas consumption would greatly reduce the cost. Little (1984) reviewed the early development of the miniature J–T cryogenic refrigerator. His group utilized photolithographic techniques to fabricate a small J–T cold stage (15 mm × 2 mm × 0.5 mm) out of glass plates and demonstrated a cooling power of 25 mW at 101 K with a N_2 flow rate of 3.1×10^{-6} kg s^{-1} at a pressure of 16.5 MPa (Garvey *et al.* 1983). Holland *et al.* (1998) presented a small L–H cooler with glass tube heat exchanger with a length of 270 mm and an outer diameter of 0.67 mm that has a cooling power of 60 mW at 88 K with a N_2 flow rate of 7.3×10^{-6} kg s^{-1} at a pressure of 10 MPa. More recently, a J–T microrefrigerator was incorporated into high-resolution magnetic resonance imaging for cryogenically cooling an rf detector coil to enhance image signal-to-noise ratio (Wright *et al.* 2005). A two to three times enhancement in performance was observed. Numerical studies on a miniature J–T refrigerator were also followed to reveal steady-state behaviors and provide the design guidelines (Chua *et al.* 2006, Ng *et al.* 2002). As long as the compressed gas cylinder is used as the high-pressure source, a true miniaturized, portable J–T cooler is not likely to be achieved.

3.16.3.4.5 Brayton gas refrigerators

The Brayton refrigeration cycle, better known as the gas refrigeration cycle, is the reverse of the closed Brayton power cycle, in which a working refrigerant remains a gas throughout a cycle without going through phase change. The cycle is similar to J–T refrigerators but differs by the fact that a J–T valve is replaced by a turbine enabling the expansion work to be recovered. Gas refrigeration systems in general require high pressure and volumetric flow rates and suffer from higher cost and lower COP compared to vapor compression. However, in cryogenic applications (below −100°C), temperatures well below the liquid–vapor saturation temperature of normal refrigerants are needed, and gas refrigeration system can achieve temperatures of below −150°C. The miniaturization of Brayton refrigerators involves the manufacturability of small, efficient turbines as well as compressors. The same group at the University of Central Florida extended their related work on a miniature centrifugal compressor (Laveau *et al.* 2000) to build a single-stage, lightweight, and compact Brayton refrigerator for space applications (Zhou *et al.* 2000). Implementing a microscale recuperative heat exchanger in their preliminary design, they anticipated a COP of about 0.2 at 80 K, the performance comparable to a corresponding Stirling refrigerator. A complete system has yet to be built and tested to show cooling performance.

3.16.3.4.6 Sorption refrigerators

Sorption refrigeration cycles have some features in common with the vapor compression cycle but differ in two main aspects. Instead of an ordinary mechanical compressor (compressing vapor between the evaporator and the condenser), a sorption compressor has a material with a large surface area to absorb an extra refrigerant to maintain a constant pressure in the chamber. After desorbing from sorption materials via heating, the sorption compressor can generate high pressure that pushes the refrigerant to a J–T valve and the evaporator. A heat sink is often required to return the temperature of the sorption material to ambient, completing a cycle. Since there is no moving part necessary in sorption refrigerators, they are expected to be more reliable in operation and more miniaturizable. However, sorption refrigerators face their own challenges such as periodic (not continuous) operation limiting heat capacity and required high heat input (Phelan *et al.* 2002). Macroscale sorption refrigerators find applications in power plants to take advantage of the waste heat for desorption or for space missions where the reliability and long-term operation are crucial (Duband and Collaudin 1999).

Despite continuous research efforts in macroscale sorption refrigerators, there have been only a few attempts to miniaturize this cooling technique. A

group at the University of Twente in the Netherlands developed a miniature sorption refrigerator for cooling of small low-temperature electronic devices with a cooling power in the range of 10 mW at 80 K (Burger *et al.* 1998, Ter Brake *et al.* 1997). A prototype sorption compressor cell is approximately 5 cm long and 5 mm in diameter. Various configurations of sorption compressors combined with J–T cold stages were analytically investigated to provide active cooling of cryogenic electronics in the space mission, thanks to its vibration-free operation (Burger *et al.* 2002). A thermodynamic analysis based on nitrogen–charcoal adsorbate–adsorbent pair was presented to design a sorption compressor for cryogenic cooling of electronics and showed a relatively small COP (Suman *et al.* 2005). This study suggested two important research directions to improve the COP of a miniature sorption refrigerator: (1) the development of absorbent materials with stronger affinity to refrigerants and (2) the design of a multistage compressor with multiple thin flat absorbent beds with fast heating/cooling response.

3.16.4 Integration of other MEMS Devices into Microcoolers

3.16.4.1 Micropumps

As discussed in the previous section, the success of active microcooler technologies such as active liquid cooling and mechanical refrigerators depends on the development of small, efficient, and reliable micropumps. Pumping a liquid for microchannel cooling systems is often the most critical factor for widespread commercial use because of the large pressure drops encountered in microchannels (Singhal *et al.* 2003). A comprehensive review of existing micropumps can be found in Laser and Santiago (2004). Micropumping technologies for driving fluid through microchannels are also reviewed (Joshi and Wei 2005, Singhal *et al.* 2004) with an emphasis on small-scale electronics cooling applications.

3.16.4.2 Microvalves

Microvalves are another critical component in all microfluidic applications, and the operation of micropumps often relies on how well passive/check microvalves can perform. A recent review of various actuation schemes and applications of microvalves can be found elsewhere (Hesketh *et al.* 2004). The review articles on micropumps in previous sections also describe the functions and challenges of the microvalves. Here we briefly discuss examples of microvalves used in microcooler applications. In fabricating the EDIFICE liquid jet impingement cooler, Amon *et al.* (2001) addressed a need to develop a microvalve that can regulate a flow rate with high back pressure. As a part of efforts to develop a miniature sorption refrigerator, Burger *et al.* (1999) reported a check valve that can hold gas pressure up to 10 MPa in the closed direction with a low cracking pressure in the forward direction. We have developed electrostatically actuated microvalves to control high-pressure flows through a microcompressor for a mesoscale flexible vapor compression refrigeration system (Philpott *et al.* 2000). Employing pneumatic assistance to balance the pressure between inlet and outlet, Bae *et al.* (2006) built a bidirectional electrostatic microvalve and demonstrated a valve switching times less than 50 μs. **Figure 35** shows a schematic and photographs of the fabricated microvalve with the three electrode layers. This type of microvalve with a fast switching response can be exploited in many microfluidic applications including microcoolers.

3.16.4.3 Integrated MEMS Sensors

To optimize the operating conditions of the microfabricated cooler devices, a macroscale testing setup is often used to take measurements such as temperature, pressure, and flow rate. However, due to a large thermal mass with macroscale measurements, collecting accurate information may be challenging. To this end, many researchers attempted to integrate MEMS sensors into microcooler devices. A few examples of simple RTD-type sensors to accurately measure temperatures and flow conditions in the microcooler applications are introduced here. Wang *et al.* (2004) integrated heaters and temperature sensors into their liquid microjet cooling devices using the same fabrication steps. LeBerre *et al.* (2006) processed the temperature sensor arrays together with silicon MHP arrays to accurately measure the axial temperature variations along the length of the heat pipe arrays. In addition to the temperature measurements, we have developed a thin-film RTD sensor that can detect liquid mass fraction at the evaporator exit of the vapor refrigeration system (Hrnjak *et al.* 2001, Shannon *et al.* 2001).

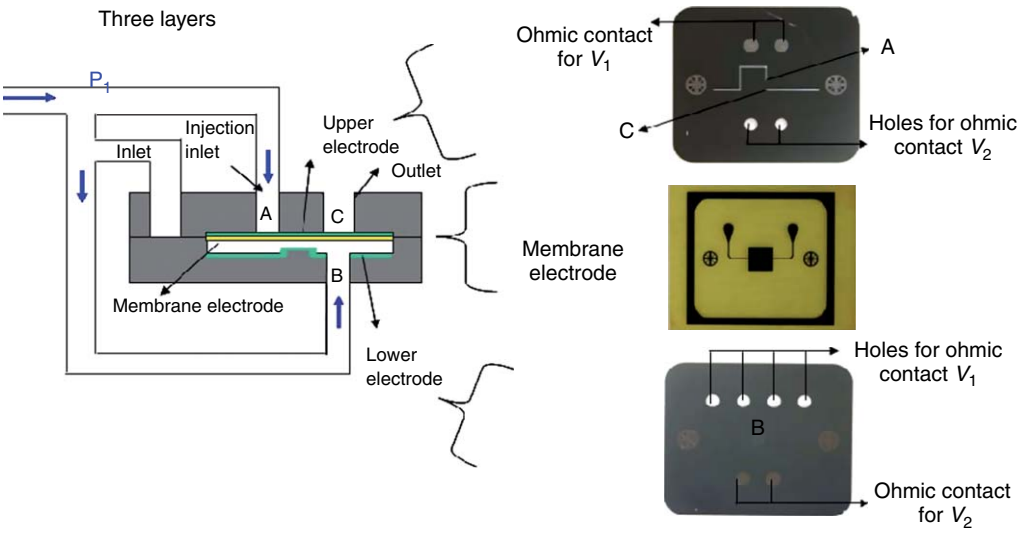


Figure 35 Schematic of an electrostatically actuated microvalve, with pictures of the three microfabricated layers that make up the valve. (Source: Bae B, Masel R I, Shannon M A 2006 A touch-mode capacitance microvalve equipped with high speed, pressure microsecond switching performance. *Proc. IEEE MEMS06*, Istanbul, Turkey. pp. 766–9. (© 2006 IEEE)).

3.16.5 Nanotechnology in Microcoolers

Recently, the exponential increase of interest in nanotechnology has encouraged researchers to incorporate nanoscale materials into microcooling technologies. CNTs, the number one driver in nanotechnology, have gained significant attention in the microcooler community because of their unique mechanical and electrical properties. CNT-based TIMs were discussed in Section 3.16.3.2. Utilizing CNT's excellent vacuum field emission properties (low turn-on voltage and high current density), a group at Purdue proposed a microscale air pump in which the ionized air molecules under the influence of an electric field induce a secondary flow (Peterson *et al.* 2005, Schlitz *et al.* 2004). The numerical simulation of this microscale ionic wind reveals an enhancement in local heat transfer coefficients (Go *et al.* 2006), but experimental verification has not yet been presented. Another group at Purdue utilizes CNT arrays on the chip surface to delay the CHF and enhance boiling heat transfer (Ujereh *et al.* 2005). An increase of 50% in the CHF and a reduction of 60% in the wall superheat were observed with CNT-coated surfaces, which resulted in a 400% enhancement of the effective heat transfer coefficients. Instead of silicon fins, Mo *et al.* (2004, 2005) proposed to use aligned CNTs grown in the surface of microchannel heat exchangers, which can dissipate heat more efficiently due to CNT's high thermal conductivity. Many features discussed in thermoelectric/thermionic coolers are intrinsically related to nanotechnology in terms of (1) the related materials such as thin-film superlattice structures, quantum well, and quantum dots and (2) the underlying physics governing the nanoscale heat transport of electrons and phonons. The development of scanning probe microscopes also helps to understand thermoelectric materials at the nanoscale (Lyeo *et al.* 2004, Miner *et al.* 2001, Reddy *et al.* 2007).

3.16.6 Summary and Conclusions

Clearly, the converging trends of ever-higher-powered and ever-smaller electronic systems as well as new types of portable lightweight devices have led to a rapid growth in microcooler technologies. Microcooling technologies for the most part are not simply scaled down from normal-scale systems, and the often-exploited physics where high surface area to volume ratios enhance heat transfer. Moreover, for single-phase convective heat transfer, the heat transfer per unit area can dramatically increase using microjet technology over normal-scale systems, and two-phase heat transfer coefficients in microchannel heat exchangers also see a dramatic increase in heat transfer coefficients. MHPs and looped thermocapillary systems see some enhancement in heat transfer,

although not uniformly in similar devices over normal scale. Nonetheless, due to the high effective heat conductivity of such devices over even the highest thermal conductivity materials, they are finding extensive use in practical systems. Active microcoolers, however, do not share the same extensive adoption. Solid-state coolers have found applications in microscale coolers and are used to reduce temperatures below ambient or to increase the heat rejection temperature of the heat sink. Solid-state devices also potentially scale well with size reduction. However, they still suffer from high cost, low efficiency, and nonlinear response with applied load, which further research, materials advancements, and overall adoption should all improve. Solid-state cooler technologies are mature compared with mechanically active coolers. Although many microscale mechanical coolers offer the promise of higher performance, efficiency, and stable response with load, a fair amount of research and development is still needed to move the technology forward. For all microcooling technologies, more work is needed to increase reliability due to important mechanical issues such as fatigue and fracture due to thermal cycling and coefficient of thermal expansion mismatch, materials incompatibility with refrigerants and solvents, and wear. In addition, manufacturing costs for microcooling technologies still need to be reduced for greater adoption. With this being said, the compelling need to create new, robust, low-cost microcooling technologies is not going to abate any time soon, and better integrated systems, particularly those with active heat pumping, are needed. Hopefully in time we will see this progress come to pass.

References

Açikalin T, Wait S M, Garimella S V, Raman A 2004 Experimental investigation of the thermal performance of piezoelectric fans. *Heat Transf. Eng.* **25(1)**, 4–14

Agrawal R, Hasan Q, Ashraf N, Sundaram K B, Chow L C, Kapat J S, Vaidya J 2003 Design and fabrication of meso-scale variable capacitance motor for miniature heat pumps. *J. Micromech. Microeng.* **13**, 1–7

Allan R 1999 MEMS micro heat exchanger employ impinging jets to boost cooling efficiency. *Electron. Des.* **47(7)**, 29–30

Ameel T A, Warrington R O, Wegeng R S, Drost M K 1997 Miniaturization technologies applied to energy systems. *Energy Convers. Mgmt.* **38**, 969–82

Amon C H, Murthy J, Yao S C, Narumanchi S, Wu C F, Hsieh C C 2001 MEMS-enabled thermal management of high-heat-flux devices EDIFICE: Embedded droplet impingement for integrated cooling of electronics. *Exp. Therm. Fluid Sci.* **25**, 231–42

Amon C H, Yao S C, Wu C F, Hsieh C C 2005 Microelectromechanical system-based evaporative thermal management of high heat flux electronics. *ASME J. Heat Transf.* **127**, 66–75

Babin B R, Peterson G P, Seyed-Yagoobi J 1993 Experimental investigation of an ion-drag pump assisted heat pipe. *J. Thermophys. Heat Transf.* **7**, 340–5

Bae B, Masel R I, Shannon M A 2006 A touch-mode capacitance microvalve equipped with high speed, pressure microsecond switching performance. *Proc. IEEE MEMS06,* Istanbul, Turkey, pp. 766–9

Bash C E 2001 Analysis of refrigerated loops for electronics cooling. *Proc. Pacific Rim/ASME Int. Electron. Packag. Tech. Conf. Exhib. (IPACK'01),* Kauai, HI, USA, pp. 811–19

Baviere R, Ayela F, Person S L, Favre-Marinet M 2004 An experimental study of water flow in smooth and rough rectangular micro-channels. *Proc. ASME 2nd ICMM,* Rochester, NY, pp. 221–8

Bergles A E 1997 Enhancement of pool boiling. *Int. J. Refrig.* **20(8)**, 545–51

Bergles A E, Leinhard J H, Kendall G E, Griffth P 2003 Boiling and evaporation in small diameter channels. *Heat Transf. Eng.* **24**, 18–40

Beyer H, Nurnus J, Bottner H, Lambrecht A, Wagner E, Bauer G 2002 High thermoelectric figure of merit ZT in PbTe and Bi2Te3-based superlattices by a reduction of the thermal conductivity. *Physica E* **13**, 965–8

Bian Z, Shakouri A 2006 Enhanced solid-state thermionic emission in nonplanar heterostructure. *Appl. Phys. Lett.* **88**, 012102

Bierschenk J, Johnson D A 2005 Latest developments in thermoelectrically enhanced heat sinks. *Electronics Cooling* **11(3)**, 24–32

Boeschoten P A, Pence D V, Liburdy J A 2000 Simulation of micro-scale jet impingement heat transfer. *Proc. ASME Heat Transf. Div.,* 2000, pp. 375–81

Bologa M K, Majboroda A N 1993 Heat exchange in electrohydrodynamic heat pipes with a slit evaporation zone. *Elektronnaya Obrabotka Materialov* **5**, 56–60

Bologa M K, Savin I K 1990 Electrohydrodynamic heat pipes. *Proc. 7th Int. Heat Pipe Conf.,* Begell House, NY, pp. 549–62

Böttner H, Chen G, Venkatasubramanian R 2006 Aspects of thin-film superlattice thermoelectric materials, devices, and applications. *MRS Bull.* **31**, 211–17

Bowers M, Mudawar I 1994 High-flux boiling in low-flow rate, low-pressure drop mini-channel and microchannel heat sinks. *Int. J. Heat Mass Transf.* **37**, 321–32

Brignoni L A, Garimella S V 1999 Experimental optimization of confined air jet impingement on a pin fin heat sink. *IEEE Trans. Comp. Packag. Technol.* **22(3)**, 399–404

Bryan J E, Seyed-Yagoobi J 1997 Heat transport enhancement of monogroove heat pipe with electrohydrodynamic pumping. *J. Thermophys. Heat Transf.* **11**, 454–60

Burger J 2001 Cryogenic microcooling. Ph.D. thesis, University of Twente, the Netherlands

Burger J F, Holland H J, Van Der Wekken M C, Ter Brake H J M, Wade L A, Elwenspoek M C, Rogalla H 1998 Development of a microminiature sorption cooler. *Proc. 7th Int. Cryogenic Eng. Conf.,* Bournemouth, UK, 123

Burger J F, van der Wekken M C, Berenschot E, Holland H J, ter Brake H J M, Rogalla H, Gardeniers J G E, Elwenspoek M 1999 High pressure check valve for application in a miniature cryogenic sorption cooler. *Tech. Dig. IEEE 12th Int. Conf. MEMS*, pp. 183–8

Burger J F, ter Brake H J M, Rogalla H, Linder M 2002 Vibration-free 5 K sorption cooler for ESA's Darwin mission. *Cryogenics* **42**, 97–108

Bürmann P, Raman A, Garimella S V 2003 Dynamics and topology optimization of piezoelectric fans. *IEEE Trans. Comp. Packag. Technol.* **25(4)**, 592–600

Campbell J S, Black W Z, Glezer A, Hartley J G 1998 Thermal management of a laptop computer with synthetic air microjets. *Proc. IEEE Intersoc. Conf. Therm. Thermomech. Phenom. Electron. Syst.,* May 27–30, pp. 43–50

Cao Y, Faghri A 1994 Micro/miniature heat pipes and operating limitations. *J. Enhanced Heat Transf.* **1**, 265–74

Cao Y, Gao M 1996 Experiments and analyses of flat miniature heat pipes. *IEEE IECEC96*, Washington, DC, USA, pp. 1402–9

Carter lll H C, Chow L C, Kapat J S, Laveau A, Sundaram K B 1999 Component fabrication and testing for a meso-scale refrigerator. *AIAA 99–4514,* Albuquerque, NM, pp. 1–11

Chang J Y, You S M 1996 Heater orientation effects on pool boiling of micro-porous-enhanced surfaces in saturated FC-72. *ASME J. Heat Transf.* **118**, 937–43

Chang J Y, You S M 1997a Boiling heat transfer phenomena from microporous and porous surfaces in saturated FC-72. *Int. J. Heat Mass Transf.* **40(18)**, 4437–47

Chang J Y, You S M 1997b Enhanced boiling heat transfer from micro-porous surfaces: A coating composition and method. *Int. J. Heat Mass Transf.* **40(18)**, 4449–60

Chang J-Y, Prasher R, Chau D, Myers A, Dirner J, Prstic S, He D 2005 Convective performance of package based single phase microchannel heat exchanger. *Proc. ASME InterPACK'05,* San Francisco, CA, pp. 183–8

Chen G 2006 Nanoscale heat transfer and nanostructured thermoelectrics. *IEEE Trans. Comp. Packag. Technol* **29**, 238–46

Chen G, Shakouri A 2002 Heat transfer in nanostructures for solid-state energy conversion. *ASME J. Heat Transf.* **124**, 242–52

Chen G, Dresselhaus M S, Dresselhaus G, Fleurial J-P, Caillat T 2003 Recent developments in thermoelectric materials. *Int. Mater. Rev.* **48**, 45–66

Chien L-H, Chen C-W 2005 Boiling of enhanced surfaces at high heat fluxes in a small boiler. *Proc. ASME IPACK'05*, pp. 655–62

Chiriac V, Chiriac F 2005 An alternative method for the cooling of power microelectronics using classical refrigeration. *Proc. ASME InterPACK'05,* San Francisco, CA, pp. 425–30

Chiriac V A, Rosales J L 2003 The unsteady characteristics of a laminar flow and heat transfer for a pair of confined impinging air jets. *Proc. ASME IMECE,* November 15–21, 2003, Washington, DC, USA, pp. 219–25

Chiriac V A, Lee T T, Rosales J L 2001 A novel cooling enhancement in microelectronic devices, systems using oscillatory impinging air jets. *Proc. ASME IMECE,* November 11–16, 2001, New York, NY, USA, pp. 81–89

Cho S, Kim S J, Youn A, DiVenere A, Wong G K L, Freeman A J, Meyer J R, Hoffman C A 2001 Artificially ordered Bi/Sb superlattice alloys: Fabrication and transport properties. *Phys. Rev. B* **64**, 235330

Cho S K, Moon H, Kim C-J 2003 Creating, transporting, cutting, and merging liquid droplets by electrowetting-based actuation for digital microfluidic circuits. *J. Microelectromech. Syst.* **12**, 70–80

Chu R C, Simons R E 1999 Application of thermoelectrics to cooling electronics: review and prospects. *IEEE 18th Int. Conf. Thermoelectrics,* Baltimore, MD, pp. 270–9

Chu R C, Simons R E, Chrysler G M 1999 Experimental investigation of an enhanced thermosyphon heat loop for cooling of a high performance electronics modules. *Proc. 15th IEEE SEMI-THERM Symp.*, pp. 1–9

Chua H T, Wang X, Teo H Y 2006 A numerical study of the Hampson-type miniature Joule–Thompson cryocooler. *Int. J. Heat Mass Transf.* **49**, 582–93

Chung J, Grigoropoulos C P, Greif R 2001 Capillary pumped loop heat spreader for electronics cooling. *Proc. ASME IMECE Heat Transf. Div.,* November 11–16, 2001, New York, NY, pp. 283–90

Colgan E G, Furman B, Gaynes M, Graham W, LaBianca N, Magerlein J H, Polastre R J, Rothwell M B, Bezama R J, Choudhary R, Marston K, Toy J, Wakil J, Zitz J, Schmidt R 2005 A practical implementation of silicon microchannel coolers for high power chips. *Proc. IEEE 21st SEMI-THERM Symp.*, pp. 1–7

Cotter T P 1984 Principal and prospect of micro heat pipes. *Proc. 5th Int. Heat Pipe Conf.,* Tsububa, Japan, Part I, pp. 328–35

Cytrynowicz D, Hamdan M, Medis P, Shuja A, Henderson H T, Gerner F M, Golliher E 2002 MEMS loop heat pipe based on coherent porous silicon technology. *Proc. AIP Space Technol. Applications Int. Forum*, pp. 220–32

Dani A, Matayabas J C, Koning P 2005 Thermal interface material technology advancements and challenges – An overview. *Proc. ASME/Pacific Rim Technical Conf. Exhib. Integr. Packag. MEMS, NEMS, Electron. Syst.*, pp. 511–16

DiSalvo F J 1999 Thermoelectric cooling and power generation. *Science* **285**, 703–6

Duband L, Collaudin B 1999 Sorption coolers development at CEA-SBT. *Cryogenics* **39**, 659–63

El-Genk M S, Guo A 1993 Transient boiling from inclined and downward-facing surfaces in a saturated pool. *Int. J. Refrig.* **6**, 414–22

El-Genk M S, Saber H H, Parker J 2005 Thermal analyses of composite copper/porous graphite spreaders for immersion cooling applications. *Proc. ASME IPACK'05*, pp. 305–14

Estes K A, Mudawar I 1995 Comparison of two-phase electronic cooling using free jets and sprays. *ASME J. Electron. Packag.* **117**, 323–32

Faghri A 1995 *Heat Pipe Science and Technology*. Taylor and Francis, Washington, DC

Fan X, Zeng G, Croke E, LaBounty C, Ahn C C, Vashaee D, Shakouri A, Bowers J E 2001a High cooling power density SiGe/Si microcooler. *Electron. Lett.* **37**, 2

Fan X, Zeng G, LaBounty C, Bowers J E, Croke E, Ahn C C, Huxable S, Majumdar A, Shakouri A 2001b SiGeC/Si superlattice microcoolers. *Appl. Phys. Lett.* **78**, 1580–2

Faulkner D, Khotan M, Shekarriz R 2003 Practical design of a 1000 W/cm2 cooling system. *IEEE 19th SEMI-THERM Symp.*, pp. 223–30

Fisher T S, Walker D G 2002 Thermal and electrical energy transport and conversion in nanoscale electron field emission processes. *ASME Trans. J. Heat Transf.* **124**, 954–62

Fisher T S, Walker D G, Weller R A 2003 Analysis and simulation of anode heating due to electrical field emission. *IEEE Trans. Comp. Packag. Technol.* **26**, 317–23

Fukutani K 2006 Solid-state microrefrigerator on a chip. *Electronics Cooling* **12(3)**, http://www.electronics-cooling.com/articles

Gambill W R, Lienhard J H 1989 An upper bound for the critical boiling heat flux. *ASME J. Heat Transf.* **111**, 815–81

Garimella S V, Singhal V 2004 Single-phase flow and heat transport and pumping considerations in microchannel heat sinks. *Heat Transf. Eng.* **25**, 15–25

Garimella S, Sobhan C 2003 Transport in microchannels – A critical review. *Annu. Rev. Heat Transf.* **13**, 1–50

Garimella S V, Joshi Y K, Bar-Cohen A, Mahajan R, Toh K C, Carey V P, Baelmans M, Lohan J, Sammakia B, Andros F 2002 Thermal challenges in next generation electronic systems – Summary of panel presentations and discussions. *IEEE. Trans. Comp. Packag. Technol.* **25**, 569–75

Garvey S, Logan S, Rowe R, Little W A 1983 Performance characteristics of a low-flow rate 25 mW, LN2

Joule-Thompson refrigerator fabricated by photolithographic means. *Appl. Phys. Lett.* **42**, 1048–50
Ghiaasiaan S M, Abdel-Khalik S I 2001 *Advances in Heat Transfer, Vol. 34*., Academic Press, New York, pp. 145–254
Ghoshal U, Grimm D, Ibrani S, Johnston C, Miner A 2005a High performance liquid metal cooling loops. *IEEE 21st SEMI-THERM Symp.,* pp. 16–19
Ghoshal U 2005b Advanced thermoelectric cooling systems. *Proc. Next-Generation Therm. Mngt. Mater. Syst.,* Phoenix, AZ, June 2005 (also see www.nanocoolers.com)
Go D B, Garimella S V, Fisher T S 2006 Numerical simulation of microscale ionic wind for local cooling enhancement. *Proc. 10th Intersoc. Conf. Therm. Thermomech. Phenom. Electron. Syst.*, p. 9
Grimes R, Walsh E, Kunz S, Davies M, Quin D 2003 Scaling the performance of micro-fans. *Proc. ASME 1st Int. Conf. Microchannels Minichannels,* Rochester, New York, USA, pp. 895–902
Groll M, Schneider M, Sartre V, Zaghdoudi M C, Lallemand M 1998 Thermal control of electronic equipment by heat pipes. *Rev. Gen. Therm.* **37**, 323–52
Gwinn J P, Webb R L 2003 Performance and testing of thermal interface materials. *Microelectron. J.* **34(3)**, 215–22
Hallinan K P, Bhagat W, Kashaboina B, Kashani A R 1998 Electro-hydrodynamic augmentation of heat transport in micro heat pipe arrays. *ASME HTD* 361–3, Anaheim, CA, USA, pp. 165–71
Hamdan M, Gerner F M, Henderson H T 2003 Steady state model of a loop heat pipe with coherent porous silicon wick in the evaporator. *IEEE 19th SEMI-THERM Symp.,* San Jose, CA, USA, pp. 88–96
Han P, Jin K-J, Zhou Y, Wang X, Ma Z, Ren S-F, Mal'shukov A G, Chao K A 2006 Analysis of optothermionic refrigeration based on semiconductor heterojunction. *J. Appl. Phys.* **99**, 074504
Hanly K, Grimes R, Walsh E, Rodgers B, Punch J 2005 The effect of Reynolds number on the aerodynamic performance of micro radial flow fans. *Proc. ASME Summer Heat Transf. Conf.,* San Francisco, CA, USA, July 17–22, 2005, pp. 549–52
Harman T C, Taylor P J, Spears D L, Walsh M P 2000 Thermoelectric quantum-dot superlattices with high ZT. *J. Electron. Mater. Lett.* **29**, L1–L4
Harman T C, Taylor P J, Walsh M P, LaForge B E 2002 Quantum dot superlattice thermoelectric materials and devices. *Science* **297**, 2229–32
Hassan I, Phutthavong P, Abdelgawad M 2004 Microchannel heat sinks: An overview of the state-of-the-art. *Microscale Thermophys. Eng.* **8**, 183–205
Hatsopoulos G N, Kaye J 1958 Measured thermal electron engine. *J Appl. Phys.* **29**, 1124–5
Heffington S N, Glezer A 2004 Two-phase thermal management using a small-scale, heat transfer cell based on vibration-induced droplet atomization. *Proc. Intersoc. Conf. Therm. Phenom., ITHERM2004*, pp. 90–4
Heffington S N, Black W Z, Glezer A 2000 Vibration-induced droplet cooling of microelectronic components. *Proc. Intersoc. Conf. Therm. Phenom., ITHERM2000,* Kauai, HI, USA, pp. 328–32
Heffington S N, Black W Z, Glezer A 2002 Vibration-induced droplet atomization heat transfer cell for high-heat flux applications. *Proc. Intersoc. Conf. Therm. Phenom., ITHERM2002*, pp. 408–12
Hesketh P J, Bintoro J S, Luharuka R 2004 Microvalve for fuel cells and miniature gas chromatographic system. *Sensors Update* **13**, 233–302
Heydari A 2002 Miniature vapor compression refrigeration systems for active cooling of high performance computers. *Proc. 8th Intersoc. Conf. Therm. Thermomech. Phenom. Electron. Syst. (I-THERM)*, pp. 371–8
Hicks L D, Dresselhaus M S 1993 Effects of quantum-well structures on the thermoelectric figure of merit. *Phys. Rev. B* **47**, 12727–31
Hicks L D, Harman T C, Sun X, Dresselhaus M S 1996 Experimental study of the effect of quantum-well structures on the thermoelectric figure of merit. *Phys. Rev. B* **53**, R10493–6
Hishinuma Y, Geballe T H, Moyzhes B Y, Kenny T W 2001 Refrigeration by combined tunneling and thermionic emission in vacuum: Use of nanometer scale design. *Appl. Phys. Lett.* **78**, 2572–4
Hishinuma Y, Geballe T H, Moyzhes B Y, Kenny T W 2002 Vacuum thermionic refrigeration with semiconductor heterojunction structure. *Appl. Phys. Lett.* **81**, 4242–4
Hishinuma Y, Geballe T H, Moyzhes B Y, Kenny T W 2003 Measurements of cooling by room-temperature thermionic emission across a nanometer gap. *J. Appl. Phys.* **94**, 4690–6
Ho C-M, Tai Y-C 1998 Micro-electro-mechanical-systems (MEMS) and fluid flows. *Annu. Rev. Fluid Mech.* **30**, 579–612
Holland H J, Burger J F, Boersme N, ter Brake H J M, Rogalla H 1998 Miniature 10–150 mW Linde–Hampson cooler with glass-tube heat exchanger operating with nitrogen. *Cryogenics* **38**, 407–10
Honda H, Takamastu H, Wei J J 2002 Enhanced boiling of FC-72 on silicon chips with micro-pin-fins and submicron-scale roughness. *J. Heat Transf.* **124**, 383–90
Hopkins R, Faghri A, Khrustalev D 1999 Flat miniature heat pipes with micro capillary grooves. *Trans. ASME* **121**, 102–9
Howard A H, Mudawar I 1999 Orientation effects on pool boiling CHF and modeling of CHF for near-vertical surfaces. *Int. J. Heat Mass Transf.* **42**, 1665–88
Hrnjak P S, Shannon M A, Leicht T M, Miller N R 2001 Detection of liquid mass fraction at the evaporator exit of refrigeration systems. *Int. J. Therm. Sci.* **40**, 773–86
Hsu K F, Loo S, Guo F, Chen W, Dyck J S, Uher C, Hogan T, Polychroniadis E K, Kanatzikis M G 2004 Cubic AgPbmSbTe2+m: Bulk thermoelectric materials with high figure of merit. *Science* **303**, 818–21
Hu Y F, Si W D, Sutter E, Qiang 2005 In situ growth of C-axis-oriented Ca3Co4O9 thin films on Si (100). *Appl. Phys. Lett.* **86**, 082103
iNEMI Report 2005, www.inemi.org
James A J, Vukasinovic B, Smith M K, Glezer A 2003a Vibration-induced drop atomization and bursting. *J. Fluid Mech.* **476**, 1–28
James A J, Smith M K, Glezer A 2003b Vibration-induced drop atomization and the numerical simulation of low-frequency single-droplet ejection. *J. Fluid Mech.* **476**, 29–62
Jeong S 2004 How difficult is it to make a micro refrigerator. *Int. J. Refrig.* **27**, 309–13
Jiang L, Mikkelsen J, Koo J-M, Huber D, Yao S, Zhang L, Zhou P, Maveety J G, Prasher R, Santiago J G, Kenny T W, Goodson K E 2002 Closed-loop electroosmotic microchannel cooling system for VLSI circuits. *IEEE Trans. Comp. Packag. Technol.* **25**, 347–55
Jones T B 1973 Electrohydrodynamic heat pipes. *Int. J. Heat Mass Transf.* **16**, 1045–8
Jones T B, Perry M P 1974 Electrohydrodynamic heat pipe experiments. *J. Appl. Phys.* **45**, 2129–32
Joshi Y, Wei X 2005 Micro and meso scale compact heat exchangers in electronics thermal management – A review. *Proc. 5th Int. Conf. Enhanced Compact Ultra-Compact Heat Exchangers*, pp. 162–79
Kajiyama K, Mizushima M, Sakata S 1973 Schottky barrier height of n-InxGal-xAs diodes. *Appl. Phys. Lett.* **23**, 458–9
Kandlikar S G 2005 High flux heat removal with microchannels – A roadmap of challenges and opportunities. *Proc. 3rd Int. Conf. Microchannels Minichannels,* Toronto, Ontario, pp. 59–68

Kang S-W, Huang D 2002 Fabrication of star grooves and rhombus grooves micro heat pipes. *J. Micromech. Microeng.* **12**, 525–31

Kang S-W, Tsai S-H, Chen H-C 2002 Fabrication and test of radial grooved micro heat pipes. *Appl. Therm. Eng.* **22**, 1559–68

Kang S-W, Tsai S-H, Ko M-H 2004 Metallic micro heat pipe heat spreader fabrication. *Appl. Therm. Eng.* **24**, 299–309

Kercher D S, Lee J B, Brand O, Allen M G, Glezer A 2003 Microjet cooling devices for thermal management of electronics. *IEEE Trans. Comp. Packag. Technol.* **26(2)**, 359–66

Khan W A, Yovanovich M M, Culham J R 2006 Optimization of microchannel heat sinks using entropy generation minimization method. *IEEE 22nd SEMI-THERM Symp.,* Ontario, Canada, pp. 78–86

Khandekar S, Groll M, Luckchoura V, Findl W, Zhuang J 2003 Micro heat pipes for stacked 3D microelectronic modules. *Proc. ASME InterPACK,* July 6–11, 2003, Maui, HI, USA, pp. 245–51

Khrustalev D 2002 Loop thermosypons for cooling of electronics. *Proc. 18th IEEE SEMI-THERM Symp.,* San Jose, CA, USA, pp. 145–50

Khrustalev D, Faghri A 1994 Thermal analysis of a micro heat pipe. *J. Heat Transf.* **116**, 189–97

Kim J, Golliher E 2002 *Steady state model of a micro heat pipe. IEEE 18th SEMI-THERM Symp.,* San Jose, CA, USA, pp. 137–44

Kirshberg J, Yerkes K L, Liepmann D 2000 Demonstration of a micro-CPL based on MEMS fabrication technologies. *Proc. Intersoc. Energy Conv. Eng. Conf. 2,* Las Vegas, NV, USA, pp. 1198–204

Klett J W, Trammell M 2004 Parametric investigation of a graphite foam evaporator in a thermosyphon with FluorinertTM and a silicon CMOS chip. *IEEE Trans. Device Mater. Reliabil.* **4(3)**, 626–37

Koo J-M, Im S, Jiang L, Goodson K E 2005 Integrated microchannel cooling for three-dimensional electronic circuit architectures. *ASME J. Heat Transf.* **127**, 49–58

Korotkov A N, Likharev K K 1999 Possible cooling by resonant Fowler-Nordheim emission. *Appl. Phys. Lett.* **75**, 2491–3

Korotkov A N, Likharev K K 2000 Cooling by resonant Fowler-Nordheim emission. *Physica B* **284–288**, 2030–1

LaBounty C, Shakouri A, Abraham P, Bowers J E 2000 Monolithic integration of thin-film coolers with optoelectronic devices. *Opt. Eng.* **39**, 2847–52

LaBounty C, Shakouri A, Bowers J E 2001 Design and characterization of thin film microcoolers. *J. Appl. Phys.* **89**, 4059–64

Lai A, Gillot G, Ivanova M, Avenas Y, Louis C, Schaeffer C, Founier E 2004 Thermal characterization of flat silicon heat pipes. *IEEE 20th SEMI-THERM Symp.,* San Jose, CA, USA, pp. 21–5

Lasance C 1997 Technical data column. *Electronics Cooling* **3(1),** http://www.electronics-cooling.com/articles/

Lasance C J M, Simons R E 2005 Advances in high-performance cooling for electronics. *Electronics Cooling* **11**, 22–39 http://www.electronics-cooling.com/articles/

Laser D J, Santiago J G 2004 A review of micropumps. *J. Micromech. Microeng.* **14**, R35–64

Launay S, Sartre V, Lallemand M 2004 Experimental study on silicon micro-heat pipe arrays. *Appl. Therm. Eng.* **24**, 233–43

Laveau A, Kapat J, Chow J, Enikov E, Sundaram K B 2000 Design analysis and fabrication of a meso-scale centrifugal compressor. *Proc. ASME Adv. Energy Div.* **40**, Orlando, FL, USA, pp. 129–37

Le Berre M, Launay S, Sartre V, Lallemand M 2003 Fabrication and experimental investigation of silicon micro heat pipes for cooling electronics. *J. Micromech. Microeng.* **13**, 436–41

Le Berre M, Pandraud G, Morfouli P, Lallemand M 2006 The performance of micro heat pipes measured by integrated sensors. *J. Micromech. Microeng.* **16**, 1047–50

Lee J, Mudawar I 2005a Two phase flow in high-heat-flux micro-channel heat sink for refrigeration cooling applications: Part I – pressure drop characteristics. *Int. J. Heat Mass Transf.* **48**, 928–40

Lee J, Mudawar I 2005b Two phase flow in high-heat-flux micro-channel heat sink for refrigeration cooling applications: Part II – heat transfer characteristics. *Int. J. Heat Mass Transf.* **48**, 941–55

Lee P, Garimella S 2003 Experimental investigation of heat transfer in microchannels. *ASME Proc.* HT2003-47293, Las Vegas, NV, USA

Lefèvre F, Lallemand M 2006 Coupled thermal and hydrodynamic models of flat micro heat pipes for the cooling of multiple electronics components. *Int. J. Heat Mass Transf.* **49**, 1375–83

Liter S G, Kaviany M 2001 Pool-boiling CHF enhancement by modulated porous-layer coating: theory and experiment. *Int. J. Heat Mass Transf.* **44**, 4287–311

Little W A 1984 Microminiature refrigeration. *Rev. Sci. Instrum.* **55**, 661–80

Liu D, Garimella S V 2002 Investigation of liquid flow in microchannels. *8th AIAA/ASME Joint Thermophys. Heat Transf. Conf.,* St. Louis, MO, pp. 2002–776

Liu W Y, Mohammadi S, Katehi L P B 2003 Polymer micro-heat-pipe for InP/InGaAs integrated circuit. *IEEE 19th SEMI-THERM Symp.,* San Jose, CA, USA, pp. 82–7

Loehrke R I, Debs R J 1975 Measurements of the performance of an electrohydrodynamic heat pump. *AIAA Paper 75–659*

Loh B G, Hyun S, Ro P I, Kleinsteuer C 2002 Acoustic streaming induced by ultrasonic flexural vibrations and associated enhancement of convective heat transfer. *J. Acoust. Soc. Am.* **111**, 875–83

Lyeo H-K, Khajetoorians A A, Shi L, Pipe K P, Ram R J, Shakouri A, Shih C K 2004 Profiling the thermoelectric power of semiconductor junctions with nanometer resolution. *Science* **303**, 816–18

Ma C F, Bergles A E 1986 Jet impingement nucleate boiling. *Int. J. Heat Mass Transf.* **29**, 1095–101

Mahalingam R, Glezer A 2003 Low-profile synthetic jet cooling for portable computers. *Proc. ASME IPACK03,* Maui, HI, USA, pp. 147–51

Mahalingam R, Glezer A 2005 Forced air cooling with synthetic jet ejectors. *Proc. ASME IPACK05,* San Francisco, CA, pp. 59–64

Mahalingam R, Rumigny N, Glezer A 2004 Thermal management using synthetic jet ejectors. *IEEE Trans. Comp. Packag. Technol.* **27(3)**, 359–66

Mahan G D 1994 Thermionic refrigeration. *J. Appl. Phys* **76**, 4362–6

Mahan G D, Woods L M 1998 Multilayer thermionic refrigeration. *Phys. Rev. Lett.* **80**, 4016–19

Mal'shukov A G, Chao K A 2001 Opto-thermionic refrigeration in semiconductor heterostructures. *Phys. Rev. Lett.* **86**, 5570–3

Mallik A K, Peterson G P, Weichold M H 1992 Use of micro heat pipes as an integral part of semiconductor devices. *ASME J. Electron. Packag.* **114(4)**, 436–42

Martin H 1977 Heat and mass transfer between impinging gas jets and solid surfaces. *Adv. Heat Transf.* **13**, 1–60

Massey B S 1983 *Mechanics of Fluids*. Van Nostrand Reinhold Co. Ltd., New York

McGillis W R, Fitch J S, Hamburgen W R, Carey V P 1990 Pool boiling enhancement techniques for water at low pressure. *WRL Research Report 90/9*, Western Digital, CA, USA

McGlen R J, Jachuck R, Lin S 2004 Integrated thermal management techniques for high power electronic devices. *Appl. Therm. Eng.* **24**, 1143–56

Mehendale S S, Jacobi A M, Shah R K 2000 Fluid flow and heat transfer at micro and meso-scales with application to heat exchanger design. *Appl. Mech. Rev.* **53**, 174–93

Miner A, Ghoshal U 2004 Cooling of high-power-density microdevices using liquid metal coolants. *Appl. Phys. Lett.* **85**, 506–8

Miner A C, Shi L, Li D, Tan K-L, Majumdar A 2001 Nanoscale thermal, electrical, and thermoelectric measurements using batch-fabricated scanning probes. *Proc. ASME Heat Transf. Div.* **369–1**, 355–8

Mo B, Ohadi M M, Dessiatoun S V, Wrenn K R 2000 Capillary pumped-loop thermal performance improvement with electrohydrodynamic technique. *J. Thermophys. Heat Transf.* **14**, 103–8

Mo Z, Anderson J, Liu J 2004 Integrating nano carbontubes with microchannel cooler. *Proc. 6th IEEE Conf. High Density Microsyst. Design Packag. Comp. Failure Analysis,* Shanghai, China, pp. 373–6

Mo Z, Morjan R, Anderson J, Campbell E E B, Liu J 2005 Integrated nanotube microcooler for microelectronics applications. *Proc. 55th Electron. Comp. Technol.,* Orlando, FL, USA, pp. 51–4

Mohapatra S C 2006 An overview of liquid coolants for electronics cooling. *Electronics Cooling* **12(2)**, http://www.electronics-cooling.com/articles/

Mohapatra S C, Loikitis D 2005 Advances in liquid coolant technologies for electronics cooling. *IEEE 21st SEMI-THERM Symp.,* San Jose, CA, USA, pp. 354–60

Mohseni K 2005 Effective cooling of integrated circuits using liquid alloy electrowetting. *IEEE 21st SEMI-THERM Symp.,* San Jose, CA, pp. 20–5

Moon S W, Hwang G, Ko S C, Kim Y T 2004 Experimental study on the thermal performance of micro-heat pipe with cross-section of polygon. *Microelectron. Reliabil.* **44**, 315–21

Morini G L 2004 Singl-phase convective heat transfer in microchannels: A review of experimental results. *Int. J. Therm. Sci.* **43**, 631–51

Mudawar I 2001 Assessment of high-heat-flux thermal management schemes. *IEEE Trans. Comp. Packag. Technol.* **24(2)**, 122–41

Mudawar I, Anderson T M 1993 Optimization of enhanced surfaces for high flux chip cooling by pool boiling. *ASME J. Electron. Packag.* **115**, 89–100

Mudawar I, Estes K A 1996 Optimizing and predicting CHF in spray cooling of a square surface. *ASME J. Heat Transf.* **118**, 672–9

Mudawar I, Wadsworth D C 1991 Critical heat flux from a stimulated electronic chip to a confined rectangular impinging jet of dielectric liquid. *Int. J. Heat Mass Transf.* **34**, 1465–79

Mudawar I, Howard H, Gersey C O 1997 An analytical model for near-saturated pool boiling CHF on vertical surfaces. *Int. J. Heat Mass Transf.* **40**, 2327–39

Mukherjee S, Mudawar I 2003 Smart pumpless loop for micro-channel electronic cooling using flat and enhanced surfaces. *IEEE Trans. Comp. Packag. Technol.* **26(1)**, 99–109

Munkejord S T, Mæhlum H S, Zakeri G R, Nekså P, Pettersen J 2002 Micro technology in heat pumping systems. *Int. J. Refrig.* **25**, 471–8

Murthy S S, Joshi Y K, Nakayama W 2000 Single chamber compact thermosyphons with micro-fabricated components. *IEEE Intersoc. Conf. Therm. Phenom.,* Las Vegas, NV, USA, pp. 321–7

Murthy S S, Joshi Y K, Nakayama W 2002 Single chamber compact two-phase heat spreaders with microfabricated boiling enhancement structures. *IEEE Trans. Comp. Packag. Technol* **25(1)**, 156–63

Murthy S, Joshi Y, Nakayama W 2003 Orientation independent two-phase heat spreader for spaced constrained applications. *Microelectron. J.* **34**, 1187–93

Murthy S, Joshi Y, Nakayama W 2004 Two-phase heat spreaders utilizing microfabricated boiling enhancement structures. *Heat Transf. Eng.* **25(1)**, 26–36

Nakayama W, Daikoku T, Kuwahara H, Nakajima T 1980a Dynamic model of enhanced boiling heat transfer on porous surfaces, Part I: Experimental investigation. *ASME J. Heat Transf.* **102**, 445–50

Nakayama W, Daikoku T, Kuwahara H, Nakajima T 1980b Dynamic model of enhanced boiling heat transfer on porous surfaces, Part II: Analytical modeling. *ASME J. Heat Transf.* **102**, 451–6

Nakayama W, Daikoku T, Nakajima T 1982 Effects of pore diameters and system pressure on saturated pool boiling heat transfer from porous surfaces. *ASME J. Heat Transf.* **104**, 286–91

Nakayama W, Nakajima T, Hirasawa S 1984 Heat sink studs having enhanced boiling surfaces for cooling microelectronic components. *ASME Paper No. 84-WA/HT-89*

Ng K C, Xue H, Wang J B 2002 Experimental and numerical study on a miniature Joule-Thompson cooler for steady-state characteristics. *Int. J. Heat Mass Transf.* **45**, 609–18

Oprins H, Baelmans M 2006 On-chip electrowetting cooling. *Electronics Cooling* **12(2)**, http://www.electronics-cooling.com/articles/

Oprins H, Vandevelde B, Beyne E, Borghs G, Baelmans M 2004 Selective cooling of microelectronics using electrostatic actuated liquid droplets. *Proc. THERMINIC,* Sophia Antipolis, Côte d'Azur, France, pp. 207–12

Oprins H, Nicole C, Baret J C, Van der Veken G, Lasance C, Baelmans M 2005 On-chip liquid cooling with integrated pump technology. *Proc. 21st SEMI-THERM Symp.,* San Jose, CA, pp. 347–53

Organ A J 1999 The miniature, reversed Stirling cycle cryo-cooler: Integrated simulation of performance. *Cryogenics* **39**, 253–66

Paik P, Pamula V K, Chakrabarty K 2004 Thermal effects on droplet transport in digital microfluidics with applications to chip cooling. *IEEE Intersoc. Conf. Therm. Phenom.,* San Jose, CA, USA, pp. 649–54

Pal A, Joshi Y K, Beitelmal M H, Patel C D, Wenger T M 2002 Design and performance evaluation of a compact thermosyphon. *IEEE Trans. Comp. Packag. Technol.* **25(4)**, 601–7

Palm B, Tengblad N 1996 Cooling of electronics by heat pipes and thermosyphons – A review of methods and possibilities. *Proc. 31st National Heat Transf. Conf. ASME-HTD* **329(7)**, 97–108

Pamula V K, Chakrabarty K 2003 Cooling of integrated circuits using droplet-based microfluidics. *Proc. ACM Great Lakes Symp. on VLSI*, 84–7

Park K A, Bergles A E 1988 Ultrasonic enhancement of saturated and subcooled pool boiling. *Int. J. Heat Mass Transf.* **31(3)**, 664–7

Park J S, Choi J H, Cho H C, Yang S S, Yoo J S 2001 Flat micro heat pipe arrays for cooling and thermal management at the package level. *Proc. SPIE, Design, Test, Integration, and Packaging of MEMS/MOEMS,* 4408, Cannes, France, pp. 424–9

Park H, Jo J, Chang J, Kim S 2006 Methodology of optimization for microchannel heat exchanger. *IEEE 22nd SEMI-THERM Symp.,* San Jose, CA, USA, pp. 65–8

Pastukhov V G, Maidanik Y F, Vershinin C V, Korukov M A 2003 Miniature loop heat pipes for electronics cooling. *Appl. Therm. Eng.* **23**, 1125–35

Pautsch G, Bar-Cohen A 2001 An overview on the system packaging of the Cray SV2 supercomputer, *Proc. ASME IPACK01 Conf.,* Kauai, HI, USA

Pavlova A A, Amitay M 2005 Electronic cooling using synthetic jets. *Proc. ASME Summer Heat Transf. Conf.,* San Francisco, CA, pp. 493–501

Peeples J W 2001 Vapor compression cooling for high performance applications. *Electronics Cooling* **7(3)**, http://www.electronics-cooling.com/articles/

Perret C, Avenas Y, Gillot C, Boussey J, Schaeffer C 2002 Integrated cooling devices in silicon technology. *Eur. Phys. J.* **123**, 115–23

Peterson G P 1992 Overview of micro heat pipe research and development. *Appl. Mech. Rev.* **45(5)**, 175–89

Peterson G P 1996 Modeling, fabrication, and testing of micro heat pipes: An update. *Appl. Mech. Rev.* **49(10)**, S175–83

Peterson R B, Al-Hazmy M 1997 Size limits for Stirling cycle refrigerators and cryocoolers. *Proc. Intersoc. Energy Conv. Eng. Conf.,* Honolulu, HI, USA, pp. 997–1002

Peterson G P, Duncan A B, Weichold M H 1993 Experimental investigation of micro heat pipes fabricated in silicon wafers. *ASME J. Heat Transf.* **115**, 751–6

Peterson M S, Zhang W, Fisher T S, Garimella S V 2005 Low-voltage ionization of air with carbon-based materials. *Plasma Sources Sci. Technol.* **14**, 654–60

Pettigrew K, Kirshberg J, Yerkes K, Trebotich D, Liepmann D 2001 Performance of a MEMS based micro capillary pumped loop for chip-level temperature control. *Tech. Dig. MEMS2001 14th IEEE Int. Conf. MEMS,* Interlaken, Switzerland, pp. 427–30

Phelan P E, Chiriac V, Lee T W 2001 Current and future miniature refrigeration cooling technologies for high power microelectronics. *Proc. 17th IEEE SEMI-THERM Symp.,* San Jose, CA, USA, pp. 158–67

Phelan E P, Chiriac V A, Lee T-Y 2002 Current and future miniature refrigeration cooling technologies for high power microelectronics. *IEEE Trans. Comp. Packag. Technol.* **25**, 356–65

Phelan P E, Chiriac V, Lee T T 2003 Performance comparison of mesoscale refrigeration technologies for electronics packaging. *Proc. ASME InterPACK'03,* Maui, HI, USA, pp. 647–54

Phelan P E, Swanson J, Chiriac F, Chiriac V 2004 Designing a mesoscale vapor-compression refrigerator for cooling high-power microelectronics. *Proc. IEEE Intersoc. Conf. Therm. Phenom.*, pp. 218–23

Philpott M L, Beebe D J, Fischer A, Flachsbart B, Marshall M, Miller N R, Selby J C, Shannon M A, Wu Y 2000 Switchable electrostatic micro-valves with high hold-off pressure. *Proc. Solid-State Sensor and Actuator Workshop,* Hilton Head Island, SC, pp. 227–9

Pollack M, Fair R B, Shenderov A D 2000 Electrowetting-based actuation of liquid droplets for microfluidic applications. *Appl. Phys. Lett.* **77**, 1725–6

Pollack M, Shenderov A D, Fair R B 2002 Electrowetting-based actuation of droplets for integrated microfluidics. *Lab Chip* **2**, 96–101

Ponnappan R 2002 Novel groove-shaped screen-wick miniature heat pipe. *J. Thermophys. Heat Transf.* **16(1)**, 17–21

Pourmohamadian N, Philpott M L, Shannon M A 2004 Novel connections for non-metallic, flexible, thin, microchannel heat exchangers. *Proc. ASME 2nd Int. Conf. Microchannels Minichannels,* Rochester, NY, June 17–19, 2004

Prasher R S, Koning P, Shipley J, Devpura A 2003 Dependence of thermal conductivity and mechanical rigidity of particle-laden polymeric thermal interface material on particle volume fraction. *ASME Trans. J. Electron. Packag* **125(3)**, 386–91

Prasher R S, Dirner J, Chang J-Y, Myers A, Chau D, Prstie S, He D 2005 Effect of localized hotspot on the thermal performance of two-phase microchannel heat exchanger. *Proc. ASME InterPACK'05,* San Francisco, CA, pp. 99–103

Prins M W J, Welters W J J, Weekamp J W 2001 Fluid control in multichannel structures by electrocapillary pressure. *Science* **291**, 277–80

Punch J 2005 Point-of-source thermal management. *Proc. IEEE 6th Int. Conf. Therm. Mech. Multiphys. Simul. Exp. Micro-Electron. Micro-Syst.*, pp. 688–94

Qu W, Mudawar I 2002 Transport phenomena in two-phase micro-channel heat sinks. *Proc. ASME IMECE,* New Orleans, LA, pp. 135–47

Qu W, Mudawar I 2003a Measurement of prediction of pressure drop in two-phase micro-channel heat sinks. *Int. J. Heat Mass Transf.* **46**, 2737–53

Qu W, Mudawar I 2003b Flow boiling heat transfer in two-phase micro-channel heat sinks? I. Experimental investigation and assessment of correlation methods. *Int. J. Heat Mass Transf.* **46**, 2755–71

Qu W, Mudawar I 2003c Flow boiling heat transfer in two-phase micro-channel heat sinks? II. Annular two-phase flow model. *Int. J. Heat Mass Transf.* **46**, 2773–84

Qu W, Mudawar I 2004 Measurement of correlation of critical heat flux in two-phase micro-channel heat sinks. *Int. J. Heat Mass Transf.* **47**, 2755–71

Qu W, Plotner M, Fisher W-J 2001 Microfabrication of thermoelectric generators on flexible foil substrates as a power source for autonomous microsystems. *J. Micromech. Microeng.* **11**, 146–52

Radtke J R, Ehrenreich H, Grein C H 1999 Multilayer thermoelectric refrigeration in Hg1-xCdxTe superlattices. *J. Appl. Phys.* **86**, 3195–8

Rainey K N, You S M 2000 Pool boiling heat transfer from plain and microporous square pin, finned surfaces in saturated FC-72. *ASME J. Heat Transf.* **122(3)**, 509–16

Rainey K N, You S M, Lee S 2003 Effect of pressure, subcooling, and dissolved gas on pool boiling heat transfer from microporous, square pin-finned surfaces in FC-72. *Int. J. Heat Mass Transf.* **46(1)**, 23–35

Ramaswamy C, Joshi Y K, Nakayama W, Johnson W B 2000 Combined effects of sub-cooling and operating pressure on the performance of a two-chamber thermosyphon. *IEEE Trans. Comp. Packag. Technol.* **23(1)**, 61–9

Ramaswamy C, Joshi Y, Nakayama W, Johnson W B 2003 Effects of varying geometrical parameters on boiling from microfabricated enhanced structures. *ASME J. Heat Transf.* **125**, 103–9

Reddy P, Jang S-Y, Segalman R A, Majumdar A 2007 Thermoelectricity in molecular junctions. *Science* **315**, 1568–71

Rodgers P, Eveloy V, Pecht M G 2005 Limits of air-cooling: status and challenges. *Proc. 21st IEEE SEMI-THERM Symp.,* San Jose, CA, USA, pp. 116–24

Roh H, Kim J, Boudreaux P J 2003 Thermal characteristics of graphite foam thermosyphon for electronic cooling. *Proc. ASME IPACK'03,* Maui, HI, USA, pp. 41–6

Rowe D M 1995 *Handbook of Thermoelectrics*. CRC Press, Boca Raton, New York

Schmidt R R 1994 Local and average transfer coefficients on a vertical surface due to convection from a piezoelectric fan. *Proc. IEEE Intersoc. Conf. Therm. Phenom.,* Washington, DC, May 4–7, 1994, pp. 41–9

Schmidt R R 2000 Low temperature electronic cooling. *Electronics Cooling* **6(3)**, 18–24

Schmidt R 2003 Challenges in electronic cooling – Opportunities for enhanced thermal management techniques – Microprocessor liquid cooled minichannel heat sink. *Proc. 1st Int. Conf. Microchannels Minichannels,* Rochester, NY, pp. 951–9

Schmidt R R, Notohardjono B D 2002 High-end server low-temperature cooling. *IBM J. Res. Dev.* **46**, 739–51

Schlitz D J, Garimella S V, Fisher T S 2004 Microscale ion-driven air flow over a flat plate. *Proc. ASME Heat Transf./Fluids Eng. Summer Conf.,* Charlotte, NC, pp. 463–8

Selby J C, Philpott M L, Shannon M A 2001 Fabrication of mesoscopic, flexible, high pressure, microchannel heat

exchangers (MHEx). *Trans. NAMRI/SME, XXIX*, pp. 469–76
Seyed-Yagoobi J, Bryan J E, Castaneda J A 1995 Theoretical analysis of ion-drag pumping. *IEEE Trans. Ind. Appl.* **31**, 469–76
Shakouri A 2004 Nanoscale devices for solid state refrigeration and power generation. *Proc. 20th IEEE SEMI-THERM Symp.,* San Jose, CA, USA, pp. 1–9
Shakouri A, Bowers J E 1997 Heterostructure integrated thermionic coolers. *Appl. Phys. Lett.* **71**, 1234–6
Shakouri A, Zhang Y 2005 On-chip solid-state cooling for integrated circuits using thin-film microrefrigerators. *IEEE Trans. Comp. Packag. Technol.* **28**, 65–9
Shakouri A, Lee E Y, Smith D L, Narayanamurti V, Bowers J E 1998 Thermoelectric effects in submicron heterostructure barriers. *Microscale Thermophys. Eng.* **2**, 37–47
Shakouri A, LaBounty C, Piprek J, Abraham P, Bowers J E 1999a Thermionic emission cooling in single barrier heterostructures. *Appl. Phys. Lett.* **74**, 88–9
Shakouri A, LaBounty C, Abraham P, Piprek J, Bowers J E 1999b Enhanced thermionic emission cooling in high barrier superlattice heterostructures. *Mater. Res. Soc. Symp. Proc.* **545**, pp. 449–58
Shannon M A, Philpott M L, Miller N R, Bullard C W, Beebe D J, Jacobi A M, Hrnjak P S, Saif T, Aluru N, Sehitoglu H, Rockett A, Economy J 1999 Integrated mesoscopic cooler circuits (IMCCs). *Proc. IMECE Adv. Energy Syst. Div.*, **39**, Nashville, TN, pp. 75–82
Shannon M A, Leicht T M, Hrnjak P S, Miller N R, Khan F A 2001 Thin-film resistance sensor for measuring liquid mass fraction in super-heated refrigerant. *Sens. Actuators A* **88**, 164–77
Sharp K V, Adrian R J, Beebe D J 2000 Anomalous transition to turbulence in microtube. *Proc. ASME IMECE MEMS,* Orlando, FL, USA, pp. 461–6
Singh R, Akbarzadeh A, Mochizuki M, Saito Y, Nguyen T, Kao B, Sataphan T, Takenaka E, Wuttijumnong V 2005 Flat miniature heat pipe with composite fiber wick structure for cooling of mobile handheld devices. *Proc. ASME InterPACK05,* July 17–22, 2005, San Francisco, CA, USA, pp. 409–14
Singhal V, Liu D, Garimella S V 2003 Analysis of pumping requirements for microchannel cooling systems. *Proc. ASME Adv. Electron. Packag. 2,* Maui, HI, pp. 473–9
Singhal V, Garimella S V, Raman A 2004 Microscale pumping technologies for microchannel cooling systems. *Appl. Mech. Rev.* **57**, 191–221
Snyder G J, Lim J R, Huang C-K, Fleurial J-P 2003 Thermoelectric microdevice fabricated by a MEMS-like electrochemical process. *Nat. Mater.* **2**, 528–31
Stefanescu S, Mehregany M, Leland J, Yerkes K 1999 Micro jet array heat sink for power electronics. *IEEE 12th Int. Conf. MEMS,* Orlando, FL, USA, January 17–21, 1999, pp. 165–70
Sugumar D, Tio K-K 2006 Thermal analysis of inclined micro heat pipes. *ASME J. Heat Transf.* **128**, 198–202
Suman B 2006 A steady state model, maximum heat transport capacity of an electrohydrodynamically augmented micro-grooved heat pipe. *Int. J. Heat Mass Transf.* **49**, 3957–67
Suman B, De S 2005 A model of the capillary limit of a micro heat pipe, prediction of the dry-out length. *Int. J. Heat Fluid Flow* **26**, 495–505
Suman S K, Fedorov A G, Joshi Y K 2005 Thermodynamic design of compact thermal compressor for sorption assisted cryogenic cooling of electronics. *Proc. ASME InterPACK'05,* San Francisco, CA, pp. 647–54
Ter Brake H J M, Burger J F, Rogalla H 1997 Microcooling of low temperature electronics. *Proc. 16th Int. Cryogenic Eng. Conf.,* Kitakyushu, Japan, pp. 391–4
Thome J R 1990 *Enhanced Boiling Heat Transfer*. Hemisphere, New York
Tien C L 1975 Fluid mechanics of heat pipes. *Annu. Rev. Fluid Mech.* **7**, 167–85
Tio K-K, Liu C Y, Toh K C 2000 Thermal analysis of micro heat pipes using a porous-medium model. *Heat Mass Transf.* **36**, 21–8
Toda M 1979 Theory of air flow generation by a resonant type PVF2 bimorph cantilever vibrator. *Ferroelectrics* **22(3–4)**, 911–18
Tripp J A, Svec F, Fréchet J M J, Zeng S, Mikkelsen J C, Santiago J G 2004 High-pressure electroosmotic pumps based on porous polymer monoliths. *Sens. Actuators B* **99**, 66–73
Tritt T M 2001 Overview of various strategies and promising new bulk materials for potential thermoelectric applications. *Mater. Res. Soc. Symp. Proc.* **691**, 3–14
Trutassanawin S, Groll E A, Garimella S V, Cremaschi L 2006 Experimental investigation of a miniature-scale refrigeration system for electronics cooling. *IEEE Trans. Comp. Packag. Technol.* **29**, 678–87
Tsai M-C, Yu C-S, Kang S-W 2005 Flat plate pipe with a novel evaporator structure. *IEEE 21st SEMI-THERM Symp.,* San Jose, CA, USA, pp. 187–90
Tsu R, Greene R F 1999 Inverse Nottingham effect cooling in semiconductors. *Electrochem. Solid-State Lett.* **2**, 645–7
Tuckerman D B, Pease R F 1981 High performance heat sinking for VLSI. *IEEE Electron. Dev. Lett.* **5**, 126–9 EDL-2
Tuma P E 2006 Evaporator/boiler design for thermosyphons utilizing segregated hydrofluoroether working fluids. *Proc. 22nd IEEE SEMI-THERM Symp.,* San Jose, CA, USA, pp. 69–77
Tuma P E, Mortazavi H R 2006 Indirect thermosyphons for the cooling electronic devices. *Electronics Cooling* **12(1)**, 26–33
Ujereh S O Jr., Mudawar I, Amama P B, Fisher T S, Qu W 2005 Enhanced pool boiling using carbon nanotube arrays on a silicon surface. *Proc. ASME Heat Transf. Div.,* 376–1, Orlando, FL, USA, pp. 691–6
Ulrich M D, Barnes P A, Vining C B 2001 Comparison of solid-state thermionic refrigeration with thermoelectric refrigeration. *J. Appl. Phys.* **90**, 1625–31
Upadhya G, Zhou P, Goodson K, Munch M, Kenny T 2003 Closed-loop cooling technologies for microprocessors. *IEEE Tech. Dig. Int. Electron Dev. Meet.,* Washington, DC, USA, pp. 775–8
Upadhya G, Zhou P, Horn J, Goodson K, Munch M 2004 Electro-kinetic microchannel cooling system for servers. *IEEE Intersoc. Conf. Therm. Phenom. 1,* Las Vegas, NV, USA, pp. 367–71
Upadhya G, Munch M, Zhou P, Horn J, Werner D, McMaster M 2006 Micro-scale liquid cooling system for high heat flux processor cooling applications. *IEEE 22nd SEMI-THERM Symp.,* San Jose, CA, USA, pp. 116–19
Vader D T, Incropera F P, Viskanta R 1992 Convective nucleate boiling on a heat surface cooled by an impingement planar jet of water. *ASME J. Heat Transf.* **114**, 152–60
Vandervort C L, Bergles A E, Jensen M K 1994 An experimental study of critical heat flux in very high heat flux subcooled boiling. *Int. J. Heat Mass Transf.* **37(1)**, 161–73
Vashaee D, Shakouri A 2004a Improved thermoelectric power factor in metal-based superlattices. *Phys. Rev. Lett.* **92**, 106103
Vashaee D, Shakouri A 2004b Electronic and thermoelectric transport in semiconductor and metallic superlattices. *J. Appl. Phys.* **95**, 1233–45
Vasiliev L L 2005 Heat pipes in modern heat exchangers. *Appl. Therm. Eng.* **25**, 1–19

Vasiliev L L 2006 Micro and miniature heat pipes – Electronic component coolers. *Appl. Therm. Eng.*, in press

Venkatasubramanian R, Colpitts T, O'Quinn B, Liu Sandra, El-Masry N, Lamvik M 1999 Low-temperature organometallic epitaxy and its application to superlattice structures in thermoelectrics. *Appl. Phys. Lett.* **75**, 1104–6

Venkatasubramanian R, Siivola E, Colpitts T, O'Quinn B 2001 Thin film thermoelectric devices with high room-temperature figures of merit. *Nature* **413**, 597–602

Vesligaj M, Amon C H 1999 Transient thermal management of temperature fluctuation during time varying workloads on portable electronics. *IEEE Trans. Comp. Packag. Technol.* **22**, 541–50

Vining C B, Mahan G D 1999 The B factor in multilayer thermionic refrigeration. *J. Appl. Phys.* **86**, 6852–3

Vukasinovic B, Smith M K, Glezer A 2003 Spray characterization during vibration-induced drop atomization. *Phys. Fluid* **16**, 306–16

Wadsworth D C, Mudawar I 1992 Enhancement of single-phase heat transfer and critical heat flux from an ultra-high-flux simulated microelectronic heat source to a rectangular impinging jet of dielectric liquid. *ASME J. Heat Transf.* **114**, 764–8

Walsh E, Grimes R 2006 The development of an integrated fan and heat sink solution for thermal management in low profile applications. *Proc. ASME ICNMM2006 4th Int. Conf. Nanochannels, Microchannels, Minichannels,* Limerick, Ireland, June 19–21, 2006

Wang E N, Zhang L, Jiang L, Koo J-M, Maveety J G, Sanchez E A, Goodson K E, Kenny T W 2004 Micromachined jets for liquid impingement cooling of VLSI chips. *J. Microelectromech. Syst.* **13**, 833–41

Webb R L, Kim N H 2005 *Principles of Enhanced Heat Transfer*. Taylor and Francis Group, New York

Webb R L, Yamauchi S 2002 Test results on a thermosyphon concept to high-power cool desktop computers and servers. *IEEE 18th SEMI-THERM Symp.,* San Jose, CA, USA, pp. 151–8

Webb R L, Gilley M D, Zamescu V 1998 Advanced heat exchange technology for thermoelectric cooling devices. *ASME J. Electron. Packag.* **120(3)**, 98–105

Wei X, Joshi Y 2003 Optimization study of stacked micro-channel heat sinks for micro-electronic cooling. *IEEE Trans. Comp. Packag. Technol.* **26**, 55–61

Wei J, Chan A, Copeland D 2003a Measurement of vapor chamber performance. *Proc. 19th IEEE SEMI-THERM Symp.,* San Jose, CA, USA, pp. 191–4

Wei J, Kubo H, Ishimine J 2003b Thermal design and packaging of a prototype refrigeration cooling system for CMOS-based MCMs. *Proc. ASME InterPACK'03,* Maui, HI, pp. 351–6

Wolf D H, Incropera F P, Viskanta R 1996 Local jet impingement boiling heat transfer. *Int. J. Heat Mass Transf.* **39**, 1395–406

Wright A C, Song H K, Elliott D M, Wehrli F W 2005 Use of a Joule-Thompson micro-refrigerator to cool a radio-frequency coil for magnetic resonance microscopy. *Rev. Sci. Instrum.* **76**, 014301

Wu Y, Shannon M A 2004 Theoretical analysis of the effect of static charges in silicon-based dielectric thin films on micro- to nanoscale electrostatic actuation. *J. Micromech. Microeng.* **14**, 989–98

Wu S, Mai J, Tai Y C, Ho C M 1999 Micro heat exchanger by using MEMS impinging jets. *IEEE 12th Int. Conf. MEMS,* Orlando, FL, USA, January 17–21, 1999, pp. 171–6

Wu Y, Liu C H, Huang H, Fan S S 2005 Effects of surface metal layer on the thermal contact resistance of carbon nanotube arrays. *Appl. Phys. Lett.* **87**, 213108

Wuttijumnong V, Nguyen T, Mochizuki M, Mashiko K, Saito Y, Nguyen T 2004 Overview latest technologies using heat pipe and vapor chamber for cooling of high heat generation notebook computer. *IEEE 20th SEMI-THERM Symp.,* San Jose, CA, USA, pp. 221–4

Xuan X C 2002 Combined thermionic-thermoelectric refrigerator. *J. Appl. Phys.* **92**, 4746–50

Xu J, Fisher T S 2006 Enhancement of thermal interface materials with carbon nanotube arrays. *Int. J. Heat Mass Transf.* **49(9–10)**, 1658–66

Xu B, Ooti K T, Wong N T, Choi W K 2000 Experimental investigation of flow friction for liquid flow in microchannels. *Int. Comm. Heat Mass Transf.* **27**, 1165–76

Yang J, Chow L C, Pais M R 1996 Nucleate boiling heat transfer in spray cooling. *ASME J. Heat Transf.* **118**, 668–71

Yao S C, Amon C H, Gabriel K, Kumta P, Murthy J Y, Wu C F, Hsieh C C, Boyalakuntia D, Narumanchi S V J, Rebello K 2001 MEMS enabled micro spray cooling system for thermal control of electronic chips. *Proc. ASME IMECE HTD 369(7),* New York, NY, USA, pp. 181–92

Yao S, Hertzog D, Zeng S, Mikkelsen J, Santiago J 2003 Porous glass electroosmotic pumps: Design and experiments. *J. Coll. Interface Sci.* **268**, 143–53

Yoo J H, Hong J L, Cao W 2000 Piezoelectric ceramic bimorph coupled to thin metal plate as cooling fan for electronic devices. *Sens. Actuators A* **79**, 8–12

Yu Z, Pohlman N A, Hallinan K P, Kashani R 2000 Active thermal control of an ion-drag pump assisted micro heat pipe. *Proc. ASME HTD* 366–4 pp. 435–41

Yu Z, Hallinani K, Bhagat W, Kashani R 2002 Electrohydrodynamically augmented micro heat pipes. *J. Thermophys. Heat Transf.* **16**, 180–6

Zeighami R, Laser D, Zhou P, Asheghi M, Devasenathipathy S, Kenny T, Santiago J, Goodson K 2000 Experimental investigation of flow transition in microchannel using micro-resolution particle image velocimetry. *Proc. ITHERM 2,* Las Vegas, NV, USA, pp. 148–53

Zeng G H, Shakouri A, LaBounty C, Robinson G, Croke E, Abraham P, Fan X F, Reese H, Bowers J E 1999 SiGe micro-cooler. *IEEE Electron Lett.* **35**, 2146–7

Zeng G, Fan X, LaBounty C, Croke E, Zhang Y, Christofferson J, Vashaee D, Shakouri A, Bowers J E 2004 Cooling power density of SiGe/Si superlattice micro refrigerators. *Mater. Res. Soc. Symp. Proc.* **793**, S2.2.1–2.2.7

Zhang Y, Shakouri A, Zeng G 2004 High-power-density spot cooling using bulk thermoelectrics. *Appl. Phys. Lett.* **85**, 2977–9

Zhang H Y, Pinjala D, Wong T N, Toh K C, Joshi Y K 2005a Single-phase liquid cooled microchannel heat sink for electronic packages. *Appl. Therm. Eng.* **25**, 1472–87

Zhang L, Wang E N, Goodson K E, Kenny T W 2005b Phase change phenomena in silicon microchannels. *Int. J. Heat Mass Transf.* **48**, 1572–82

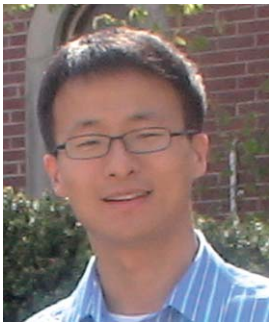
Zhang Y, Christofferson J, Shakouri A, Zeng G, Bowers J E, Croke E T 2006a On-chip high speed localized cooling using superlattice microrefrigerators. *IEEE. Trans. Comp. Packag. Technol.* **29**, 395–401

Zhang Y, Zeng G, Shakouri A 2006b Silicon microrefrigerator. *IEEE. Trans. Comp. Packag. Technol.* **29**, 570–6

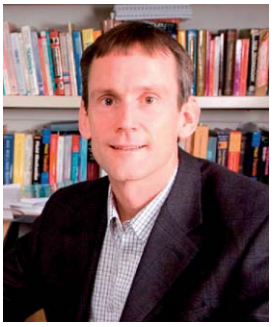
Zhou L, Kapat J S, Chow L C, Lei S Y 2000 Design of a high performance cryocooler for propellant liquefaction and storage on mars. *Proc. ASME Heat Transf. Div.* **366**, 331–8

Zhou P, Horn J, Upadhya G, Goodson K, Munch M 2004 Electro-kinetic microchannel cooling system for desktop computers. *IEEE 20th SEMI-THERM Symp.,* San Jose, CA, USA, pp. 26–9

Biographies



Junghoon Yeom is a PhD candidate of Mechanical Science Engineering at University of Illinois and received his B.S. (2000) from Yonsei University, Seoul, Korea and M.S. (2003) from University of Illinois. During his graduate study, he has been involved with several DARPA projects including micro-formic acid fuel cell and micro gas chromatography system and developed an expertise in designing and fabricating MEMS devices and microfluidic systems. Mr. Yeom has 5 journal publications, 13 peer-reviewed conference proceedings, 3 conference presentations, and 1 patent application.



Mark A. Shannon is the Director of a NSF STC the WaterCAMPWS, which is a multiple university and government laboratory center for advancing the science and engineering of materials and systems for revolutionary improvements in water purification for human use. He is also the Director of the Micro-Nano-Mechanical Systems (MNMS) Laboratory at the University of Illinois at Urbana-Champaign, a 2000 sq. ft class 10 and 100 cleanroom laboratory devoted to research and education in the design and fabrication of micro- and nanoelectromechanical systems (MEMS and NEMS), microscale fuel cells and gas sensors, high-temperature microchemical reactors, micro-nano-fluidic sensors for biological fluids, and has published over 100 papers on these subjects. He chairs the Instrument Systems Development Study Session for the National Institutes of Health. He is the James W. Bayne Professor of Mechanical Engineering at UIUC, and received his B.S. (1989) M.S. (1991) and Ph.D. (1993) degrees in Mechanical Engineering from the University of California at Berkeley. He received the NSF Career Award in 1997 to advance microfabrication technologies, the Xerox Award for Excellence in Research (2004), the Kritzer Scholar (2003–2006), the Willet Faculty Scholar (2004–2007), the BP Innovation in Education Award in 2006, and the CERL 2006 Research, Development, or Operational Team Award.

3.17 Resonators, Oscillators, and Frequency References

Wan-Thai Hsu, Discera Inc., Ann Arbor, MI, USA

Glossary

CC Beam Clamped-Clamped Beam

FF Beam Free–Free Beam

VCO Voltage Controlled Oscillator

3.17.1 Standards for Frequency References

Frequency reference unit usually consists of a resonator that defines the frequency and an electronic circuitry that sustains the oscillation. Over the past 30 years, people have been using vibrating mechanical resonators, such as ceramic resonators, quartz crystal, and surface acoustic wave (SAW) resonators, for frequency references. These mechanical resonators can be excited only at some specific frequencies defined by their geometry and mode shape. The vibration energy is confined within this geometry so well that the energy loss is very small. Therefore, these resonators exhibit good frequency selectivity and stability compared to electrical resonators such as inductor–capacitor (LC) tanks.

Frequency reference standard is well-defined based on the accuracy of the oscillator frequency. This accuracy is defined as the frequency variation across all environmental parameters (e.g., temperature) and one year of aging. As shown in **Table 1** (Vig 2005), the frequency accuracy ranges from 10^{-4} to 10^{-12}, while of course the cost per unit ranges from 50 cents to $50 000.

One of the earliest works that micromechanical systems (MEMS) resonators researchers reference to is Nathanson's article 'The resonant gate transistors' published in 1967 (Nathanson *et al.* 1967). In this paper, gate oxide of a transistor was removed to allow metal gate to mechanically vibrate. This vibration provided high quality factor (Q) frequency selection. A 12-kHz integrated oscillator was demonstrated together with 90–150 ppm per degree Celsius of thermal stability.

Table 1 Frequency standard of oscillators

Oscillators type[a]	*Accuracy*[b]	*Typical applications*
IC oscillators	10^{-2} to 10^{-3}	Toys
Crystal oscillator (XO)	10^{-5} to 10^{-4}	Computer, camcorder, HDD
Temperature -compensated crystal oscillator (TCXO)	10^{-6}	Cell phones, GPS,
Oven-controlled crystal oscillator (OCXO)	10^{-8} (with 10^{-10} per g option	Navigation system clock and frequency standard, MTI radar
Small atomic frequency standard (Rb, RbXO)	10^{-9}	Satellite terminals, bistatic, and multistatic radar
High-performance atomic standard (Cs)	10^{-12} to 10^{-11}	Strategic C^3, EW

[a] Sizes range from <5 mm^3 for clock oscillators to >30l for Cs standards. Costs range from <$0.5 for clock oscillators to >$0.5 000 for Cs standards.
[b] Including environmental effects (e.g., −40°C to +85°C) and 1 year aging.

To date (the end of year 2006), MEMS oscillator is able to fulfill all the accuracy requirements in crystal oscillator (XO) category (i.e., 10–100 ppm frequency variation from −40°C to 85°C) with the advantages of lower cost, higher mechanical reliability (Hsu 2006), and fully compatible system applications. The future development would place the MEMS oscillator into the temperature-compensated crystal oscillator (TCXO) category for communication applications.

3.17.2 Resonators

3.17.2.1 Resonator Materials

As the key component that determines the oscillator frequency, the resonator requires materials that exhibit excellent reliability and stability in all operation conditions. The major requirements of the materials are (1) low mechanical stress, (2) no or negligible hysteresis, and (3) low aging rate.

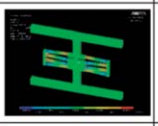
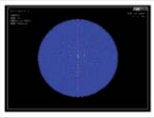
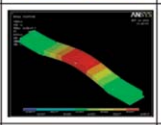
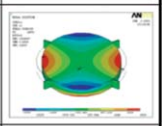
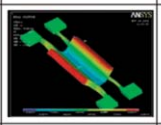
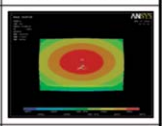
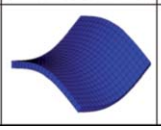
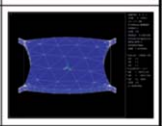
In the literature, materials such as single-crystal silicon, polycrystalline silicon (Tang *et al.* 1989), silicon germanium (Franke and Howe 1999), AlN (Piazza *et al.* 2005), tantalum nitride (Pacheco *et al.* 2004), titanium (Parker *et al.* 2005), poly-diamond (Wang *et al.* 2004), and plated nickel (Hsu and Nguyen 1998) have been recorded as resonator materials. Each material has its advantages and disadvantages. For example, polycrystalline silicon has the advantage of being a complementary metal oxide semiconductor (CMOS)- compatible process, so almost all the foundries have the process. But the deposition temperature of standard poly-Si prevents it from being deposited directly on finished CMOS wafers, thus in turn becoming an obstacle for integration. Another example is metal resonators can be fabricated directly on top of CMOS wafers with low-temperature processes; however, metal resonators often exhibit stronger hysteresis effects.

3.17.2.2 Resonator Designs

The MEMS resonator can be defined as a micromachined mechanical structure that vibrates at its natural resonance frequency due to external excitations. One mechanical structure could be excited to different vibration modes at different frequencies. The external excitations of micromechanical resonators could be electrostatic, piezoelectric, laser, mechanical vibrations, and magnetic field. But regardless of the resonators and the excitations mechanism, for oscillator applications, the resonator should be excited to a specific mode with highest quality factor, lowest motional impedance, and highest power before resonator peaks reach the nonlinear region.

Table 2 shows a variety of micromechanical resonators that can be used for oscillators ranging in frequency from low kilohertz to possibly gegahertz. As shown, folded-beam resonators with interdigitated comb-drive transducers that resonate at 32.768 kHz are used for real-time clock (RTC) in power management chips or in watches. The challenges to manufacture commercial grade kilohertz oscillators are making low bias voltage for electrostatic transducer and designing oscillator circuits with low operation power. An integrated clock oscillator was first demonstrated in 1993 (Nguyen and Howe 1993). The state-of-the-art RTC oscillator was presented in Cioffi and Hsu (2005), where the resonator can be operated at 1.5 V and the IC consumes only <1 μA of current.

Table 2 Resonators in difference shapes and modes

Flexural mode		*Bulk mode*	
Folder beam		Disk (contour mode)	
Clamped-clamped beam		Disk (wine-glass mode)	
Free-free beam		Square (extension)	
Square		Square (deform)	

The folded-beam comb-drive resonator structure may not be scaled to achieve tens of megahertz in frequency due to its large mass. The highest frequency that a comb-drive resonator achieved is 1 MHz (Roessig *et al.* 1998). Instead, as shown in **Table 2**, resonators with parallel capacitor transducer between resonator body and driving/sensing electrodes would enable resonators to be operated in 1–100 MHz frequency range. The examples are flexural clamped-clamped beams, free-free beams, and coupled square resonators. They exhibited good Q as high as 8 000–22 000 (Demirci *et al.* 2003, Hsu and Cioffi 2004), Hsu *et al.* 2001a, b, Wang *et al.* 2000) with polysilicon as the resonator material. Flexural resonators fabricated on silicon-on-insulator (SOI) wafers also exhibited good quality factor (Kim *et al.* 2005, Pourkamali *et al.* 2003).

For higher-frequency vibration, resonator stiffness needs to be higher. Therefore, other than flexural type of resonator, bulk acoustic mode resonators came into play after researchers figured out a process to form a lateral nanometer gap between electrodes and resonators for efficient electrostatic actuations (Hsu *et al.* 2001a, b). The resonators include disk resonator operating in radial mode (Clark *et al.* 2000, 2005), wine-glass mode (Abdelmoneum *et al.* 2003, Hao *et al.* 2004, Pourkamali *et al.* 2004), and the square resonator in extensional mode (Kaajakari *et al.* 2003). Bulk acoustic mode resonators have demonstrated quality factor as high as 98 000 in vacuum and 70 000 in air, which is very good for oscillator applications.

3.17.2.3 Resonator Mode of Operation

This section details the operation and mechanical design of MEMS resonators, using the simplest flexural modes as examples. **Figure 1** shows the simulated mode shapes of clamped-clamped beam, cantilever beam, and free-free beam, which will be discussed in this section.

The equation of motion of flexural beams can be expressed as

$$\frac{\partial^4 u(x)}{\partial x^4} = \left(\omega_0^2 \frac{\rho A}{EI}\right) u(x) = k^4 u(x) \qquad [1]$$

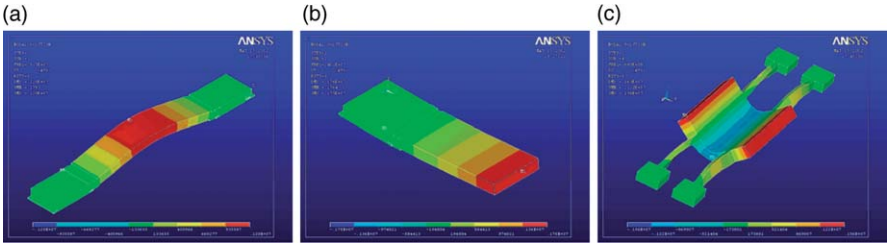


Figure 1 Mode shapes for flexural resonators: (a) clamped-clamped beam, (b) cantilever beam, and (c) free-free beam.

where ω_0 is the resonator frequency, ρ is the density of the material, E is the Young's modulus of the beam, A and I are the cross-section area and inertia of the resonator, respectively. Assume the solution is

$$u(x) = A\cosh kx + B\sinh kx + C\cos kx + D\sin kx \quad [2]$$

and apply various boundary conditions to the beam, frequency equations different mode shapes revealed the following:

(1) Clamped end: $u = 0$ and $u' = 0$
(2) Free end: $u'' = 0$ and $u''' = 0$
(3) Simple end (allow rotation but not movement): $u = 0$, $u'' = 0$

Clamped-clamped beam means both ends of the resonator beam are fixed or anchored to the substrate. Therefore, the boundary condition is $u(0) = 0$, $u(L) = 0$, $u'(0) = 0$, and $u'(L) = 0$. Therefore, the mode shape is

$$u_n(x) = \gamma_n(\cosh k_n x - \cos k_n x) + (\sinh k_n x - \sin k_n x) \quad [3]$$

where

$$\gamma_n = \frac{\sinh k_n L - \sin k_n L}{\cosh k_n L - \cos k_n L} \quad [4]$$

where k_nLs are the solutions of $\cosh kL \cos kL - 1 = 0$. Therefore, $k_1L = 4.73$, $k_2L = 7.853$, $k_3L = 10.966$, $k_nL \sim (n\pi + \pi/2)$ for $n > 3$.

The frequency equation becomes

$$f_n = \frac{(k_nL)^2}{2\pi L^2}\sqrt{\frac{EI}{\rho A}} \quad [5]$$

Cantilever beam means that one end is fixed while the other end is free. For a cantilever beam, applied clamped boundary condition is on one end and free boundary condition is on the other end, thus $u(0) = 0$, $u''(\mathrm{L}) = 0$, $u'(0) = 0$, and $u'''(L) = 0$. The mode shape is

$$u_n(x) = \gamma_n(\cosh k_n x - \cos k_n x) + (\sinh k_n x - \sin k_n x) \quad [6]$$

where

$$\gamma_n = \frac{\sinh k_n L + \sin k_n L}{\cosh k_n L + \cos k_n L} \quad [7]$$

and k_nLs are the solutions of $\cosh kL \cos kL + 1 = 0$. Therefore, $k_1L = 1.875$, $k_2L = 4.694$, $k_3L = 7.855$, $k_4L = 10.966$. Frequency of cantilever beam can be expressed as

$$f_n = \frac{(k_nL)^2}{2\pi L^2}\sqrt{\frac{EI}{\rho A}} \quad [8]$$

For a free-free beam, the boundary condition is $u''(0) = 0$, $u''(L) = 0$, $u'''(0) = 0$, and $u'''(L) = 0$ and the mode shape is

$$u_n(x) = \gamma_n(\cosh k_n x + \cos k_n x) + (\sinh k_n x + \sin k_n x) \quad [9]$$

where

$$\gamma_n = -\frac{\sinh k_n L - \sin k_n L}{\cosh k_n L - \cos k_n L} \quad [10]$$

where k_nLs are the solutions of $\cosh kL \cos kL - 1 = 0$. Therefore, $k_1L = 4.73$, $k_2L = 7.853$, $k_3L = 10.966$, $k_nL \sim (n\pi + \pi/2)$ for $n > 3$. The frequency equation becomes

$$f_n = \frac{(k_nL)^2}{2\pi L^2}\sqrt{\frac{EI}{\rho A}} \quad [11]$$

Note that the frequency equation of the free-free beam is the same as that of the clamped-clamped beam.

In theory, both ends of free-free beam are free. However, it is not possible to have a beam without any support. In reality, the free-free beam is supported on its nodal points by nonintrusive supporting beams that are strategically designed to isolate the resonator beam from its anchors via quarter wavelength impedance transformations to achieve a real free-free boundary condition (Wang *et al.* 2000). The mode shapes of the above beams are shown in **Figure 2**.

3.17.2.4 Quality Factor

The theory of the above section is for free vibration. It means the system does not need to work against resistive force, so its total energy remains constant at all times. In reality, the vibration system is always resisted by dissipative forces such as friction, acoustic transmission, air viscosity, and internal dissipations. This dissipative force is proportional to the speed of movement. Also assume that the vibration system is set to continuous vibration by a driving force

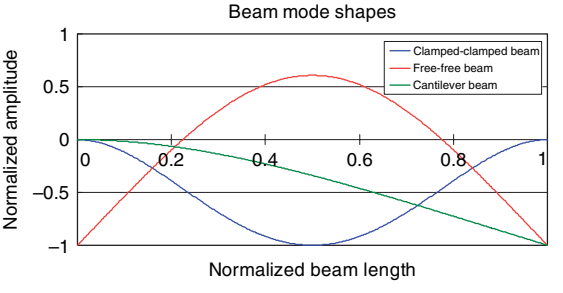


Figure 2 Normalized mode shapes of three flexural beams.

$F = F_0 \sin(\omega t)$; the mass–spring system can be modeled by a second-order mechanical system as

$$m\ddot{x} + kx + c\dot{x} = F_0 \sin(\omega t) \quad [12]$$

where m is the mass of the resonator beam, c is the friction coefficient, and k is the spring constant. Using $\omega_0^2 = k/m$ and $n = c/2m$, and assuming that the damping is small, which is the case for most resonators, the solution is

$$x_1 = A\mathrm{e}^{-nt}\left(\sin\left(\sqrt{(\omega_0^2 - n^2)}t + \alpha\right)\right) = A\mathrm{e}^{-nt}\sin(\omega_r t + \alpha) \quad [13]$$

$$x_2 = B\sin(\omega t - \varphi)$$

Substitute x_1 and x_2 into eqn [12]

$$B = \frac{\frac{F_0}{m}}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4n^2\omega^2}} = \frac{\frac{F_0}{k}}{\sqrt{\left(1 - (\omega/\omega_0^2)^2\right)^2 + 4(n/\omega_0)^2(\omega/\omega_0)^2}} \quad [14]$$

and the phase lag is

$$\varphi = \tan^{-1}\left(\frac{2n\omega}{\omega_0^2 - \omega^2}\right) \quad [15]$$

This is the phase lag between free vibration frequency ω_0 and the driving force frequency ω. Therefore, while $\omega \to \omega_0$, $\varphi = 90°$. It means that the phase of the mechanical system is 90° behind the driving force if the driving frequency is close to the natural resonance frequency. For oscillator application, this will be the case and the amplitude is close to

$$B(\omega_0) = \frac{F_0 m\omega_0}{kc} = \frac{F_0}{k}Q \quad [16]$$

where Q is defined as the quality factor. In physics, quality factor is defined by the total energy in the system divided by the energy dissipated by damping in one cycle of oscillation. The total energy of the system is

$$E = \frac{1}{2}m\dot{x}^2 = \frac{1}{2}mA^2\omega_r^2 \quad [17]$$

The energy dissipated in one cycle is

$$\Delta E = \frac{1}{2\pi}\int_0^T c\dot{x}\,dt = \frac{cA^2\omega_r}{2} \quad [18]$$

and

$$Q = \frac{E}{\Delta E} = \frac{m\omega_r}{c} \quad [19]$$

Let us consider a case where the resonator vibrates for half of the maximum energy. In this case the vibration amplitude is $1/\sqrt{2}$ of the $B(\omega_0)$ in eqn [16]. At this point, the vibration frequency can be the solutions ω_1 and ω_2 of

$$\frac{1}{\sqrt{\left(1 - (\omega/\omega_0)^2\right)^2 + 4(n/\omega_0)^2(\omega/\omega_0)^2}} = \frac{1}{\sqrt{2}}\frac{m\omega_0}{c} \quad [20]$$

$$\Delta\omega = \omega_1 - \omega_2 = \frac{2n}{\sqrt{1 - 2(n/\omega_0)^2}} \quad [21]$$

Therefore, in frequency domain, the Q is defined by the resonance frequency divided by 3 dB bandwidth, which is equal to $\Delta\omega$ in eqn [21]. We can see that this definition matches with the previous mechanical definition in eqn [19]

$$Q = \frac{\omega_r}{2n}\sqrt{1 - 2(n/\omega_0)^2} \approx \frac{\omega_r}{2n} = \frac{m\omega_r}{c} \quad [22]$$

As shown, resonator Q strongly depends on the friction coefficient, i.e., energy dissipation of the resonator beam. The energy dissipation is usually directly related to how well the resonator is anchored to the substrate. Here are some examples of how different anchors affect the Q.

3.17.2.5 Frequency Pulling

In Section 3.17.2.4, a general driving force was discussed. This section focuses on driving resonators with electrostatic force, which conserves the quality factor of the resonators better than other electrical actuation such as piezoelectric.

Electrostatic force generated by a parallel plate can be expressed as

$$F = \frac{1}{2}(V_P - v_i)^2\frac{\partial C}{\partial x} \quad [23]$$

Assume $v_i = |v_i|\cos(\omega t)$, according to eqn [15], the resultant displacement is $x = |x|\sin(\omega t)$. Thus

$$F = \frac{1}{2}(V_P - v_i)^2\frac{\partial C}{\partial x} = \frac{1}{2}\left(V_P^2 - 2V_P v_i + v_i^2\right)\frac{\partial}{\partial x}\left[C_0\left(1 + \frac{x}{d_0}\right)^{-1}\right] \quad [24]$$

Assuming x is small compared to d_0 and takes only the terms at ω_0, we have

$$F = V_P\frac{C_0}{d_0}|v_i|\cos\omega_0 t + \frac{C_0}{d_0^2}V_P^2|x|\sin\omega_0 t \quad [25]$$

Note that if $V_P = 0$, the driving force at ω_0 is zero. The first term is the driving force that is applied directly to the system. On the other hand, the second term represents a positive force that is dependent on the displacement. It means that this force is the same as the mechanical spring restoring force, which acts opposite to the input force. Therefore, we can define

$$k_e = \frac{C_0}{d_0^2} V_P^2 \quad [26]$$

as the electrical spring constant, which is subtracted from mechanical spring constant, changes the resonant frequency. Now the expression for the resonant frequency is

$$f = f_0 \left(1 - \frac{k_e}{k_m}\right)^{1/2} = f_0 \left(1 - \frac{V_P^2 C_0}{k_m d_0^2}\right)^{1/2} \quad [27]$$

The electrical spring constant pulled the frequency of the resonator.

3.17.2.6 Equivalent Circuit Model

As mentioned in Section 3.17.2.5, the resonator vibrates when the frequency of the AC excitation signal matches the resonant frequency of the resonator beam. To obtain a circuit model for the future oscillator design, an RLC equivalent circuit of the micromechanical resonator is derived in this section.

Since a DC bias is applied on the resonator, the capacitive gap between the electrode and the resonator now acts as a DC-biased time-varying capacitor yielding an output current

$$I_0 = \frac{d(C(t)V(t))}{dt} = C(t)\frac{dV(t)}{dt} + V(t)\frac{dC(t)}{dt} \quad [28]$$

where

$$C = C_0 + x\frac{\partial C}{\partial x} \quad [29]$$

$$V = -V_P + v_i \quad [30]$$

Substituting $v_i = |v_i|\cos(\omega t)$ and $x = |x|\sin(\omega t)$ into the equation, the output current becomes

$$I_0 = C_0\omega_0|v_i|\sin\omega_0 t - V_P\frac{\partial C}{\partial x}\omega_0|x|\cos(\omega_0 t) + \frac{\partial C}{\partial x}|v_i||x|\sin 2\omega_0 t \quad [31]$$

The first term is related to static parallel capacitance, which is not related to resonator's vibration motion. The second term is the output current related to the resonator vibration at ω_0, while the third term is the harmonic current at $2\omega_0$.

At resonance frequency ω_0, the displacement of the beam is

$$|x| = \frac{Q}{k} V_P|v_i|\frac{\partial C}{\partial x} \quad [32]$$

Then

$$I_0|_{at\omega_0} = -\frac{\omega_0 Q}{k}\left(V_P\frac{\partial C}{\partial x}\right)^2|v_i| \quad [33]$$

The motional impedance becomes

$$R_x = \left[\frac{\omega_0 Q}{k}\left(V_P\frac{\partial C}{\partial x}\right)^2\right]^{-1} = \frac{kd_0^4}{\omega_0 Q V_P^2(\varepsilon_0 A_{electrode})^2} \quad [34]$$

L_x and C_x can be calculated based on R_x, Q, and ω_0

$$L_x = \frac{R_x Q}{\omega_0} \quad [35]$$

$$C_x = \frac{1}{R_x Q \omega_0} \quad [36]$$

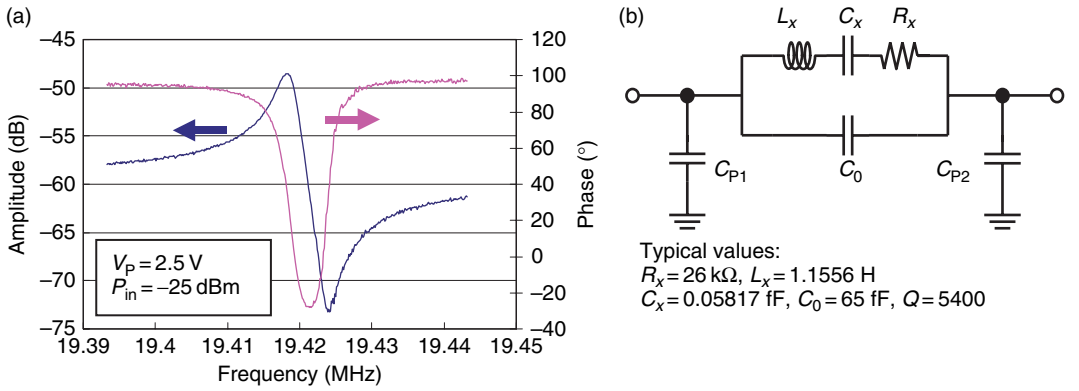


Figure 3 (a) Measured spectrum of a microelectromechanical systems (MEMS) resonator (b) its equivalent model and typical numbers for RLC model.

A typical measured resonator spectrum is shown in Figure 3(a). From the spectrum, Q, ω_0, C_0, and R_x can be extracted. Therefore, L_x and C_x can be calculated. Assuming Z_R is the resonator impedance

$$S_{21} = 10^{\text{level}/20}, \quad |Z_R| = 2.50\Omega\left(\frac{1-S_{21}}{S_{21}}\right) \qquad [37]$$

Based on the Z_R far from the resonance peak, C_0 can be calculated as

$$C_0 = \frac{1}{2\pi f_{\text{far from carrier}}|Z_R|} \qquad [38]$$

Knowing C_0, R_x can be calculated from the insertion loss at the resonance frequency

$$R_x = \frac{|Z_R|}{\sqrt{1-(2\pi f_0 C_0|Z_R|)^2}} \qquad [39]$$

3.17.3 MEMS Oscillators

Oscillator normally consists of a high-Q frequency reference tank, which determines the frequency and the stability of the oscillator, and a circuit that sustains the oscillation. As shown in Figure 4, at the stable oscillation, the loop gain of the oscillator must be 1 and the phase shift of the loop must be zero.

Figure 4 also illustrates how Q's of the resonator affect the frequency stability of the oscillators. As shown, in the case of $Q = 10$, if the phase of the oscillator circuits varies 40° due to external environment, the phase of the resonator needs to be adjusted −40° in order to meet oscillation criteria. In this case, the resonant frequency moves 400 kHz away from the original resonant frequency. In the case of $Q = 1000$, the deviation is only 4 kHz. Therefore, the larger the Q, the less the frequency deviation is.

Figure 5 presents the top-level schematic of the oscillator circuit (Lin *et al.* 2003), where the resonator is represented by its equivalent LCR circuit. As described in previous sections, MEMS resonator is driven by an AC voltage and outputs AC current. Therefore, a series resonant configuration is used, employing a transresistance amplifier in order to accommodate the resistance of the MEMS resonator. The conditions required for sustained oscillation can be stated as the following:

(1) The total phase shift around the closed positive feedback loop must be 0°. In this series resonant topology, both the resonator and the sustaining amplifier have 0° phase shift from input to output.
(2) The gain of the transresistance-sustaining amplifier R_{amp} must be larger than the sum of the motional resistance of the micromechanical resonator and the input and output resistances of the sustaining amplifier, i.e.

$$R_{\text{amp}} \geq R_x + R_i + R_o \qquad [40]$$

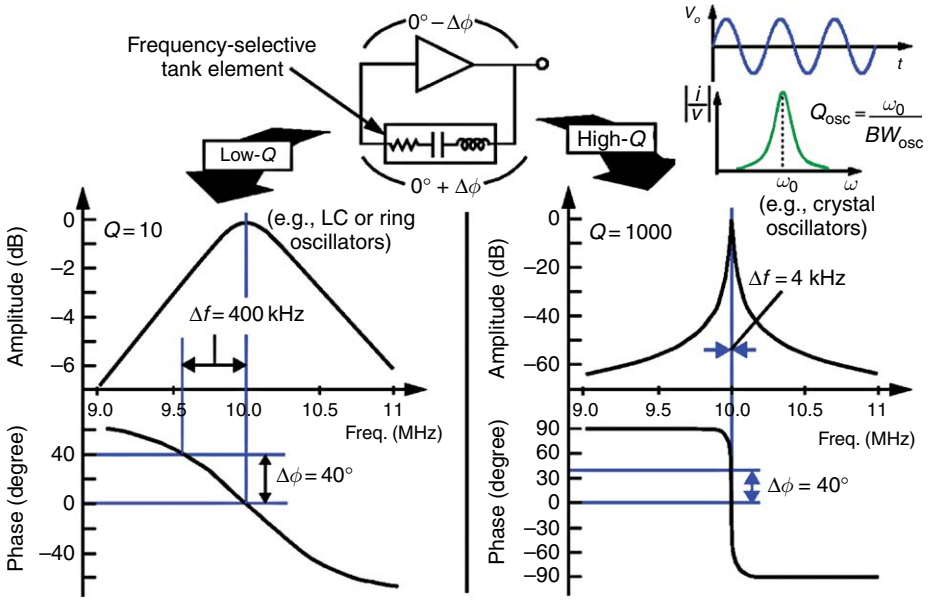


Figure 4 Oscillator diagram and the effect of resonator Q.

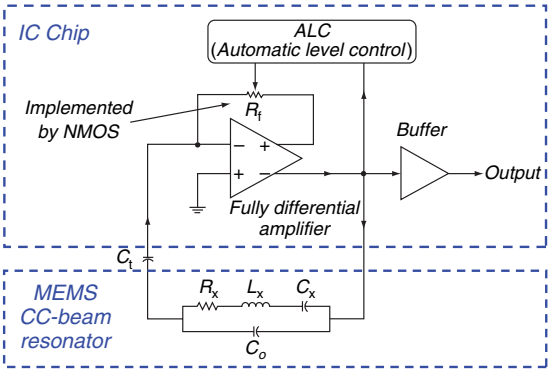


Figure 5 Top-level schematic of a microelectromechanical systems (MEMS) oscillator circuits. (Source: Lin Y-W, Lee S, Ren Z, Nguyen C T-C 2003 Series resonant micromechanical resonator oscillator. *Technical Digest, IEDM'03*, December 2003.)

R_{amp} is required for oscillation start-up to the point where a transresistance amplifier can achieve sufficient gain to instigate oscillation. The particular sustaining amplifier circuit of **Figure 7** achieves the needed 0° phase shift for oscillation in only a single stage instead of two stages, which improves both its noise and bandwidth performance. As shown in the schematic of the circuit, transistors M_1–M_5 comprise the basic differential op amp, while M_{11}–M_{18} constitute a common-mode feedback circuit that sets its bias point. The MOS resistors M_{Rf1} and M_{Rf2} serve as a shunt-shunt feedback elements that allow control of the transresistance gain via adjustment of their gate voltages.

The oscillation amplitude increases from noise till nonlinear effects make the oscillation level saturated. In conventional quartz XOs, the nonlinearity comes from the circuitry. However, for the MEMS oscillator, the resonator often becomes nonlinear before the circuitry. Duffing described the nonlinear effect of mechanical resonators and oscillators in 1918. MEMS researchers first noticed that a clamp-clamp bridge under tensile stress exhibits spring-hardening nonlinearity describable by Duffing's equation. IBM first reported both of the measured and theoretical jumps of MEMS oscillators in their amplitude and phase diagrams of the frequency response due to the spring-softening nonlinearity (Fan and Crawford 1993). As shown in **Figure 6(a)**, if the resonator is measured with lower input power, the resonator spectrum is not changed. However, as the input power increases, the resonance peak starts shifting toward the lower frequency – meaning the spring constant of the resonators is softened by the driving power. While the resonator is operated at the nonlinear region, frequency jumping could be observed at oscillation. **Figure 6(b)** shows the plot of frequency shift versus input power at various temperatures. If 10 ppm frequency shift is the maximum frequency variation we can tolerate due to the resonator nonlinearity, at 40°C, the maximum input power for the resonator is −17 dBm. As Q becomes higher at low temperature (discussed in Section 3.17.4), the resonator starts showing nonlinear effect at −32 dBm of input power. In fact, MEMS nonlinearity contributes the limiting mechanism of the oscillation when the automatic level control (ALC) is not engaged (**Figure 7**).

Therefore, in addition to the basic sustaining amplifier, the oscillator IC better includes an ALC circuit. As shown in **Figure 8**, this circuit consists of

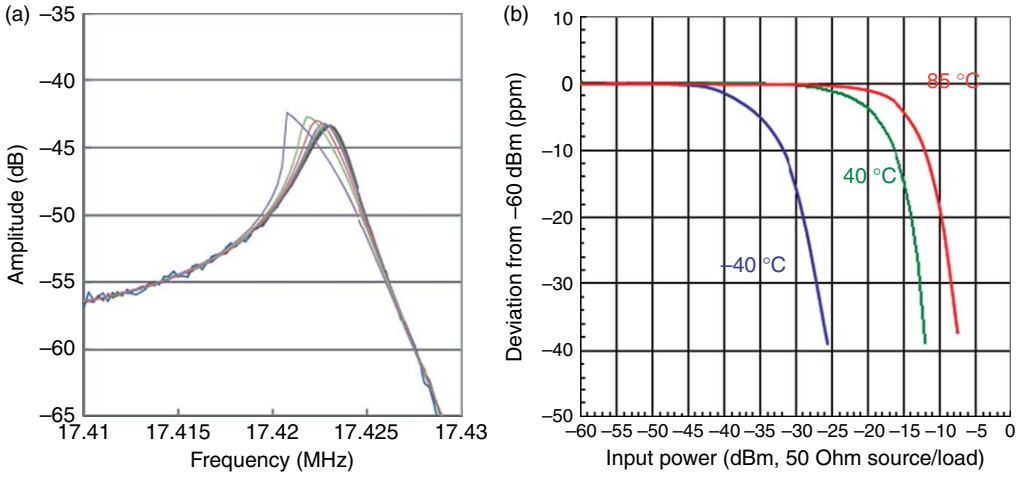


Figure 6 Duffing effect of resonators (a) shown in spectrum and (b) shown in power vs. frequency plot.

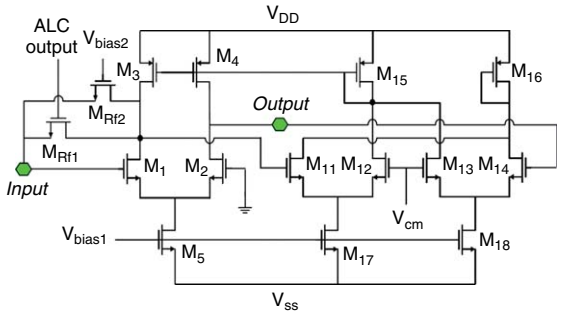


Figure 7 Example circuits for microelectromechanical systems (MEMS) oscillators. (Source: Lin Y-W, Lee S, Ren Z, Nguyen C T-C 2003 Series resonant micromechanical resonator oscillator. *Technical Digest, IEDM'03*, December 2003.)

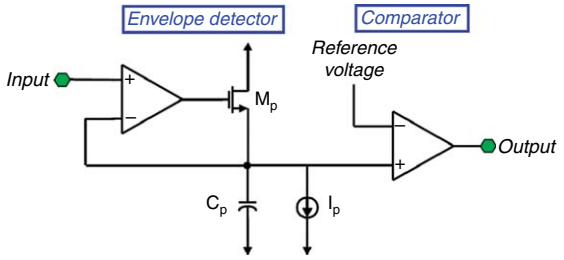


Figure 8 Automatic level control (ALC) circuit. (Source: Lin Y-W, Lee S, Ren Z, Nguyen C T-C 2003 Series resonant micromechanical resonator oscillator. *Technical Digest, IEDM'03*, December 2003.)

an envelope detector that effectively measures the oscillation amplitude and a comparator that feeds back to the gain-setting MOS resistor M_{Rf1} to match the oscillation amplitude to a preset reference level applied to the negative input of the comparator.

Figure 9 presents a plot of phase noise density versus offset from the carrier frequency. It shows a phase noise of $-80\,\text{dBc Hz}^{-1}$ at 1 kHz offset from the carrier, dropping to $-120\,\text{dBc Hz}^{-1}$ at far-from-carrier offsets, with the ALC disengaged. The close-to-carrier noise performance is somewhat poor due to a $1/f^3$ dependence that most likely derives from the Duffing-based limiting of this oscillator. With ALC engaged, this $1/f^3$ noise component is removed, leaving an expected $1/f^2$ close-to-carrier component commonly exhibited by high Q oscillators and allowing a much-improved phase noise density of $-95\,\text{dBc Hz}^{-1}$ at 1 kHz offset. Due to the smaller power level of the oscillator when ALC is engaged, the far-from-carrier phase noise density increased a bit to $-110\,\text{dBc Hz}^{-1}$.

Phase noise of the oscillator is expected to be lower if the power handling of the resonator can be increased. The power handling is defined by the maximum AC input power before the resonator shows Duffing effect. Normally the frequency deviation is defined based on the accuracy requirements of the product. In this chapter, we simply use 10 ppm. At room temperature, the clamped-clamped beam has power handling of about -35 dBm. It could be lower at lower temperature. For free-free beam it is about -20 dBm, while for bulk acoustic type of resonators it could be as high as -10 dBm. However, while pushing

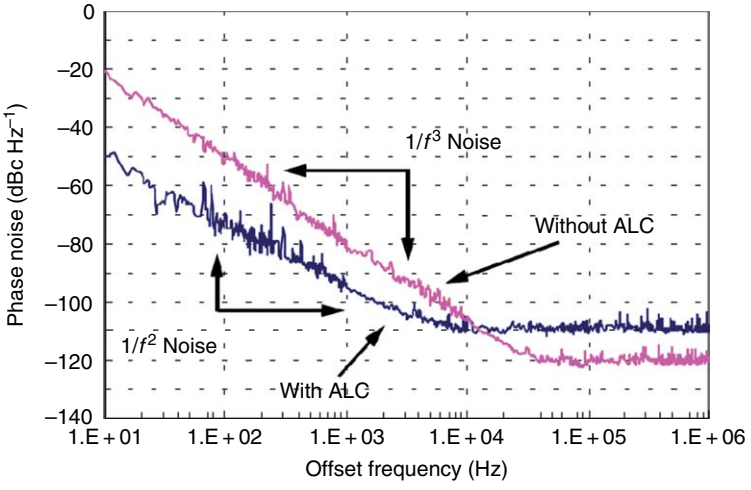


Figure 9 Phase noise plot of a microelectromechanical systems (MEMS) oscillator with and without ALC. (Source: Lin Y-W, Lee S, Ren Z, Nguyen C T-C 2003 Series resonant micromechanical resonator oscillator. *Technical Digest, IEDM'03*, December 2003.)

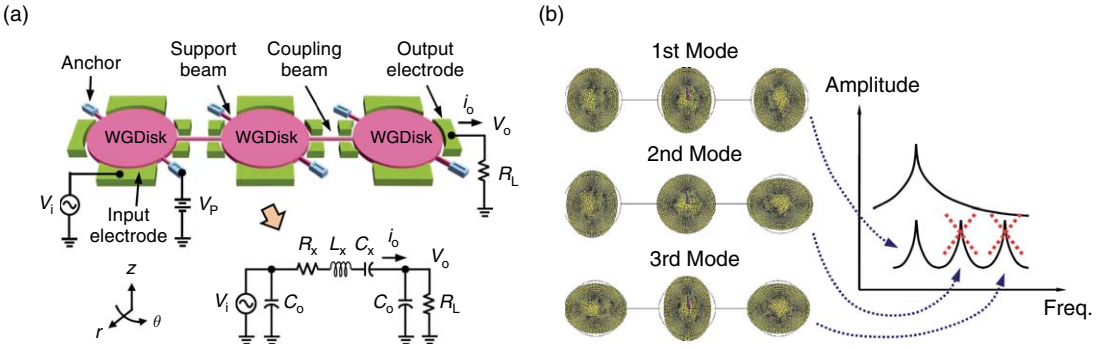


Figure 10 Operation of strong, coupled wine-glass disk resonators. (Source: Lin Y-W *et al.* 2005 Low phase noise array-composite micromechanical wine-glass disk oscillator. *Technical Digest, International Electron Device Meeting*, December 5–7, 2005, pp. 4.)

the power handling higher, there will be trade-offs between other important parameters such as V_P, gap spacing d_0, and R_x (Kaajakari *et al.* 2003, Lin *et al.* 2005).

One way to increase the power handling is by connecting multiple resonators together and forcing them to resonate at the same frequency by design (Clark *et al.* 2006, Demirci *et al.* 2003, Lin *et al.* 2005). As shown in **Figure 10**, three disk resonators operated at wine-glass mode were coupled together. As shown, the mechanical connection of the resonators actually realizes a multipole filter structure that has several modes of vibration. Each modal peak corresponds to a state where all resonators vibrate at exactly the same frequency. **Figure 10(b)** presents ANSYS simulations showing the different modes of this structure, which are distinguishable by the phases between the resonators. Because each mode exhibits unique resonator phasings, a single mode can be selected by choosing the input AC signal to match the phasing of the desired mode. For oscillator applications, the first mode of the array is usually selected. It should be noted that the use of half-wavelength coupling beams serves to spread the modal peaks of the filter structure far apart, which facilitates the selection of one, and only one, of the modes. Once a single mode is selected, the structure practically behaves as a single resonator, but with a current-handling ability equal to the sum of the currents from all constituent resonators.

Figure 11 shows a phase noise of $-123\,\text{dBc Hz}^{-1}$ at 1 kHz offset from the carrier and $-136\,\text{dBc Hz}^{-1}$ at far-from-carrier offsets. This far-from-carrier noise floor is about 4 dB better than that of an oscillator using a single wine-glass disk resonator, verifying the

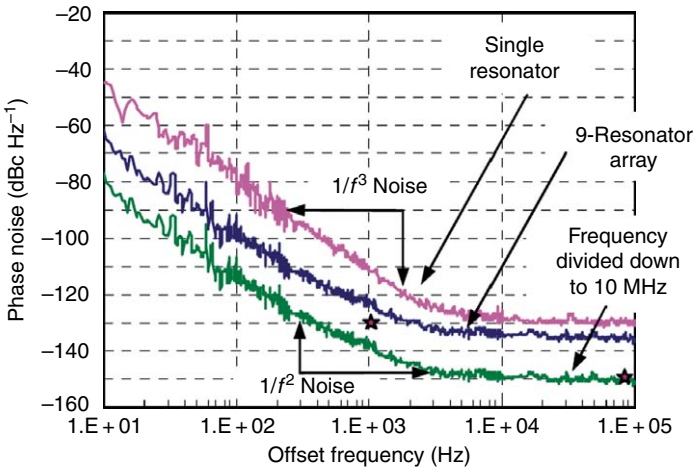


Figure 11 Phase noise of strong coupled wine-glass disk resonator oscillators. (Source: Lin Y-W *et al.* 2005 Low phase noise array-composite micromechanical wine-glass disk oscillator. *Technical Digest, International Electron Device Meeting*, December 5–7, 2005, pp. 4.)

utility of resonator array design. In addition, when a coupled array is utilized the undesired $1/f^3$ noise is removed, due to its increased power handling ability. With $1/f^3$ noise suppressed, an expected $1/f^2$ dependence commonly exhibited by high-Q oscillators then remains, improving the close-to-carrier phase noise at 1 kHz offset from -110 to $-123\,\mathrm{dBc\,Hz^{-1}}$. Dividing down to 10 MHz for fair comparison, this corresponds to $-138\,\mathrm{dBc\,Hz^{-1}}$ at 1 kHz offset and $-151\,\mathrm{dBc\,Hz^{-1}}$ far from the carrier, which more than satisfies (by 8 and 1 dB) the stringent GSM requirement.

3.17.4 Frequency Trimming

As described in previous sections, the frequency of the resonators is determined by both material properties and geometry of the resonators – exactly the same as quartz crystal resonators. Typically, depending on the systems, the oscillator output frequency needs to be within parts per million (e.g., from 1 to 100 ppm) to the target frequency for system applications. However, since semiconductor processes are used for MEMS resonator fabrication, the process variation of film thickness, lithography, and etching made the frequency of the resonators prior to trimming vary between $\pm 1\%$ and $\pm 10\%$, depending on the process control.

Furthermore, it is known that the resonator packaging itself can affect the resonance frequency of a resonator. In other words, the resonance frequency of a resonator can differ before and after encapsulation in a package. If the trimming is conducted before the resonator is packaged, then there is uncertainty as to how much frequency trimming is required.

The resonator frequency can be trimmed after it is packaged by applying voltage pulses to the resonator. This approach is often referred as localized annealing. This causes filament-like heating of portions of the resonator, which results in frequency trimming and improvements in Q (Wang *et al.* 1997). However, the direction of frequency tuning is dependent on the stress of the film. Also localized annealing may introduce unknown aging problems. As a result, this method may not be applied to the trimming that requires ppm level of accuracy.

Since the resonator is small in size, another approach of tuning a MEMS resonator is by creating arrays of resonators with various beam lengths. The variation in beam length results in a difference in resonance frequency between the shortest and the longest beam that is sufficient to account for the typical variation (due to manufacturing tolerances) in the resonance frequency of a resonator (US Pat. No. 6 600 389). This work also discloses that the increment in resonance frequency between adjacent beams is smaller than the targeted frequency variation tolerance of the desired resonator. Therefore, one of the beams will be qualified to serve as the desired resonator. Once that beam is identified, the other beams are disabled.

Material deposition has traditionally been used to trim the resonance frequency of crystals. But traditional material deposition, such as evaporation, has not been considered to be a practical technique for use with MEMS resonators because MEMS resonators are typically much smaller in size than their crystal counterparts. Also, it is desirable to trim a MEMS resonator after it has been packaged. Chiao and Lin (2004) demonstrated a postpackage-pulsed laser deposition frequency trimming for micromechanical resonators as illustrated in **Figure 12(a)**. However, the resonator frequency still failed to reach the accuracy of ppm due to the difficulty in controlling the geometry and the amount of the materials deposited by laser. Furthermore, the quality factor of the resonator is often degraded by additional materials on the resonators because the resonator vibration lost its balance.

As shown in **Figure 12(b)**, a packaged resonator can also be trimmed by directing the electromagnetic energy to the resonator through a transparent portion of the resonator package. The energy removes (e.g.,

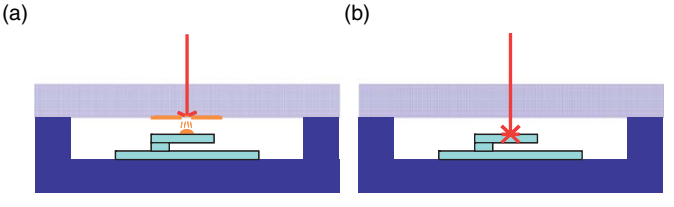


Figure 12 Frequency trimming of packaged resonators via laser enhanced (a) material deposition and (b) material removal.

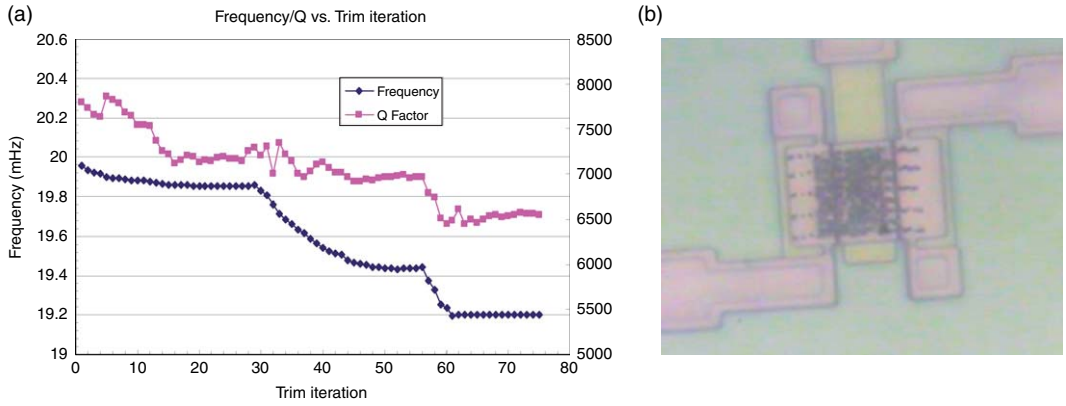


Figure 13 (a) Frequency and *Q* vs. trimming iterations, and (b) trimmed resonator surface.

ablates) the mass at the point(s) of contact on the resonator. Removing the mass from the resonator affects its resonance frequency in a predictable manner. The best electromagnetic energy source, from spot size and energy control point of view, is the femtosecond laser (Hsu and Brown 2007).

Frequency trimming is a function of (1) the amount of mass removed (or deposited) from the resonator and (2) the location(s) on the resonator at which the mass is removed (or deposited). Once the frequency-trimming requirement is determined (i.e., how much change in resonance frequency is required), the trimming is implemented by removing (or depositing) the mass from selected locations on the resonator, because the change in frequency caused by removing (or depositing) the mass is a function of not only the amount of mass but also its location on the resonant element. As a result, a mass-trimming map is developed for identifying the potential mass-trimming sites on the resonator. Each site can be classified as a fine-tuning site or a coarse-tuning site as a function of the magnitude of the change in resonance frequency that is caused by removing (a given amount of) the mass at that site. Fine-tuning sites and coarse-tuning sites tend to group in different regions on the resonator. Moreover, the trimming locations are usually located symmetrically on the resonator surface. This symmetric arrangement has the effect of changing the resonance frequency while the quality factor, *Q*, of the resonator remains substantially unaffected.

As shown in **Figure 13**, resonator trimming was demonstrated by directing femtosecond laser through a transparent potion of the resonator package to remove resonator mass on specific locations. One shot of femtosecond laser creates a 0.5-μm-diameter and 0.3-μm-deep divot. A MEMS resonator with a frequency 4% away from the target was trimmed to 2.5 ppm away from the target. The accuracy is limited by the control of temperature because the resonator has a temperature coefficient of −20 ppm per degree Celsius.

3.17.5 Temperature Compensation

Silicon resonators typically exhibit monotonic temperature-dependent curve. This temperature dependence is mainly due to the temperature coefficient of the Young's modulus of the resonator structure. As a result, depending on the resonator material, the temperature coefficients range from −40 to −15 ppm per degree Celsius. In comparison to the temperature characteristic of quartz crystal resonators, as shown in **Figure 14**, MEMS resonators show a straight line with a slope larger than even the worse case of AT-cut quartz crystals. Based on the experiment results, the temperature coefficient of frequency (TC_f) of the MEMS resonator can be maintained as a straight line from −180 to 150°C.

As defined in Section 3.17.1, the accuracy of the frequency needs to be within 10–100 ppm across all environmental conditions for XO applications. Therefore, TC_f of the MEMS resonators needs to be compensated.

Figure 15 shows the frequency deviation and *Q* of a poly-Si MEMS resonator measured at temperatures ranging from −40 to 100°C. As shown, the frequency line shows a slope close to −20 ppm per degree Celsius. On the other hand, *Q* goes from 6 000 at room temperature to 15 000 at −40°C. As mentioned in Section 3.17.3, resonator vibration

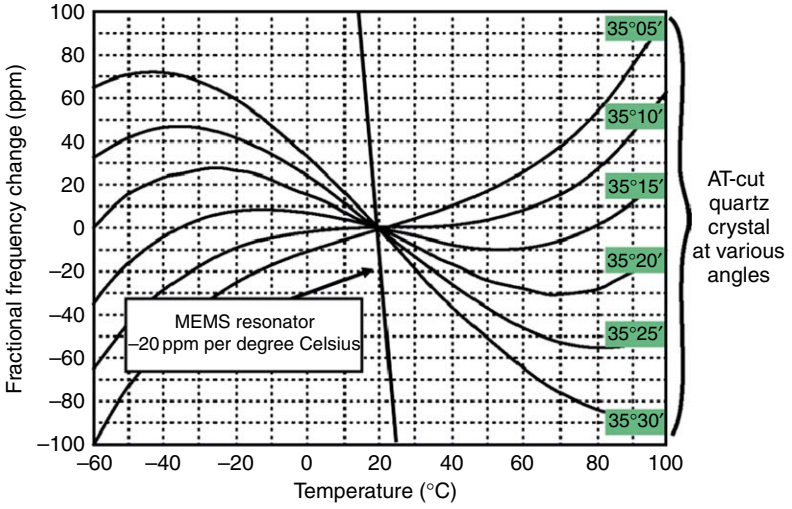


Figure 14 Frequency variations versus temperature of quartz crystals and microelectromechanical systems (MEMS) resonator.

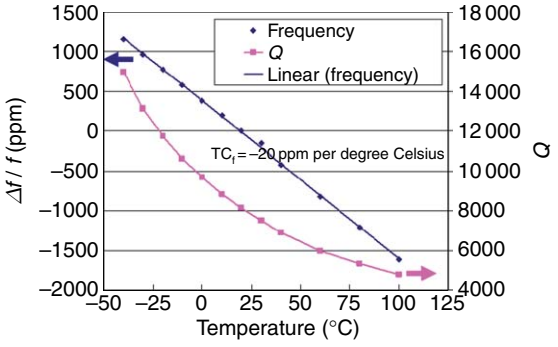


Figure 15 Frequency deviation and *Q* of poly-Si MEMS resonator across the temperature.

amplitude is proportional to Q. Therefore, at low temperature, the vibration amplitude will be 2.5× larger than the amplitude at room temperature, exhibiting Duffing effect thus introducing frequency instabilities. The first task for temperature compensation is to design the sustaining amplifier to prevent resonators from Duffing. This could be done via temperature-dependent gain or Q control circuits.

Temperature compensations for the MEMS oscillators can be categorized into mechanical compensation and electrical compensation. Mechanical compensation means the resonator frequency is adjusted at different temperatures. Electrical compensation, on the other hand, means the resonator frequency is fixed and the frequency adjustment is done on the circuitry.

3.17.5.1 Resonator Compensation

The ideal case of resonator temperature compensation is to design a mechanical structure to adjust the resonator frequency according to the temperature. From Section 3.17.4, we know that the resonator frequency can be expressed as

$$f = f_0\left(1-\frac{k_e}{k_m}\right)^{1/2} = f_0\left(1-\frac{V_P^2 C_0}{k_m d_0^2}\right)^{1/2} = f_0\left(1-\frac{V_P^2 \varepsilon_0 A}{k_m d_0^3}\right)^{1/2} \quad [41]$$

V_P, k_m, and d_0 are parameters that can affect the resonator frequency

$$f = \frac{(k_1 L)^2}{2\pi L^2}\sqrt{\frac{EI}{\rho A}} = 1.013\frac{h}{L^2}\sqrt{\frac{E}{\rho}} \quad [42]$$

So the TC_f of the uncompensated micromechanical resonator is

$$TC_f = \frac{\alpha_E + \alpha_{res}}{2} \quad [43]$$

where α_E and α_{res} are the temperature coefficient of the Young's modulus and the thermal expansion coefficient of the resonator material, respectively. The typical value of the TC_f of a clamped-clamped beam is −20 ppm per degree Celsius. To overcome this negative TC_f of the resonator, parameters in eqn [41] can be adjusted so that the resonator frequency increases with temperature to counteract with the frequency decrease with temperature due to the temperature variation of the Young's modulus.

3.17.5.1.1 Mechanical stiffness

One direct way of temperature compensation is by applying a tensile force along the resonator beam depending on the temperature. The higher the temperature, the larger the tensile force. Thus the tensile force counteracts the negative temperature coefficient of the resonator.

As shown in **Figure 16**, a mechanical compensation structure is designed based on a lateral resonator beam. It consists of a flexural-mode clamped-clamped-like beam anchored to the substrate at one end, but supported on the other by a folded structure comprised of a truss section attached to two outer beams. The outer beams are also designed to be longer than the resonator beam, so that they will expand faster than it with increasing temperature, generating a net tension in the resonator beam, as shown in the figure. This tensile stress serves to increase the beam's resonance frequency and thus oppose frequency decreases caused by the temperature dependence of the Young's modulus, resulting in a smaller overall f_0 excursion over a given temperature range. To ensure that only the resonator beam vibrates when excitation signals are applied, the outer support beams are made much wider than the resonator beam, making them rigid against lateral motions.

In the lateral vibration mode, the beam of **Figure 16** essentially has fixed-fixed end conditions; so the expression for its resonance frequency is identical to that of previous clamped-clamped beam resonators. With a mechanical structure generating an axial stress along the resonator beam, the frequency of the resonator with axial stress σ can be expressed as

$$f = 1.013\frac{W}{L^2}\sqrt{\frac{E}{\rho}}\sqrt{\left(1 + 0.293\frac{L^2}{EW^2}\sigma\right)} \qquad [44]$$

The stress is generated by the thermal expansion difference between the resonator material and the substrate material and the length difference between the resonator beam and the outer compensation beam. The stress is

$$\sigma = E(\alpha_{res} - \alpha_{sub})\Delta T\left(\frac{L_{out}}{L} - 1\right) \qquad [45]$$

The temperature coefficient of frequency

$$TC_f \approx \frac{\alpha_{res} + \alpha_E}{2} + 0.146\left(\frac{L}{W}\right)^2(\alpha_{res} - \alpha_{sub})\left(\frac{L_{out}}{L} - 1\right) \qquad [46]$$

The first term $(\alpha_{res} + \alpha_E)/2$ is the TC_f of uncompensated MEMS resonator. Let us assume that it is -20 ppm per degree Celsius and $TC_f = 0$, we have

$$\frac{L_{out}}{L} = 1 + \frac{136.986}{\alpha_{res} - \alpha_{sub}}\left(\frac{W}{L}\right)^2 \qquad [47]$$

Therefore, TC_f of the resonator is zero when L_{out}/L satisfies eqn [47].

Experimental work of this temperature compensation method shows that the TC_f is reduced from -20 to 2.5 ppm per degree Celsius (Hsu *et al.* 2000). The major drawback of mechanical compensation is that once the geometry of the resonator is defined on the mask, it is hard to compensate for process variations. An array of devices with slightly different L_{out}/L ratio can be fabricated to accommodate the variations.

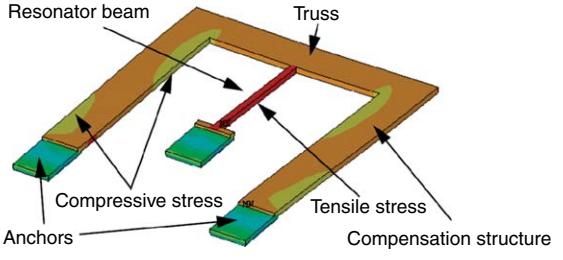


Figure 16 Mechanical temperature compensation structure. (Source: Hsu *et al.* 2000.)

3.17.5.1.2 Electrical stiffness compensation

Resonator frequency changes with bias voltage, V_P, according to eqn [41]. Therefore, it is possible to adjust V_P so that the output frequency remains the same across the temperature. Including the temperature variation into eqn [41], the resonator frequency becomes

$$f_T = f_0(1 - TC_f(T - T_0))\left(1 - \frac{V_{P,T}^2\varepsilon_0 A}{k_m d_0^3}\right)^{1/2} \qquad [48]$$

if f_T is kept the same as f_{T0}, the resonator bias voltage $V_{P,T}$ can be expressed as

$$V_{P,T} = \sqrt{\frac{k_m d_0^3}{\varepsilon_0 A}}\left(\frac{1 - \left(V_{P,T0}^2\varepsilon_0 A\right)/k_m d_0^3}{(1 + TC_f(T - T_0))^2}\right) \qquad [49]$$

Consider 10-MHz clamped-clamped beam as an example. **Figure 17(a)** shows the bias voltage versus frequency of a typical 8 μm × 40 μm beam. In order to generate constant output frequency, a circuit must be designed to supply bias voltage, as plotted in

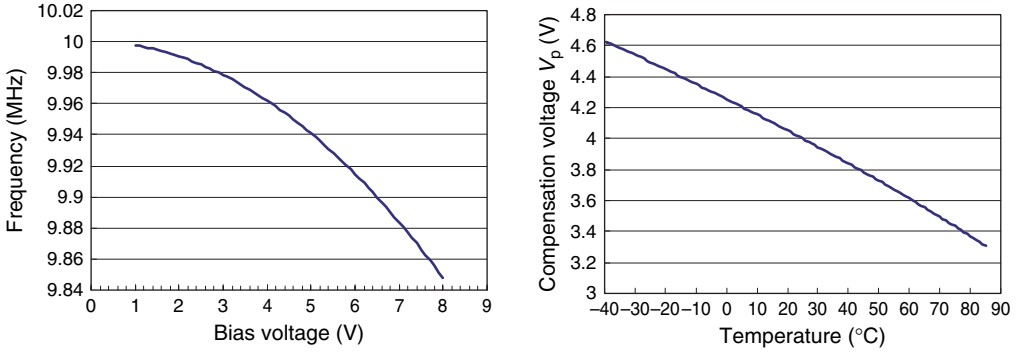


Figure 17 Bias voltage versus frequency for 10-MHz resonator ($d_0 = 1000$ Å, $k_m = 3000$ N, $A = 160\,\mu m^2$).

Figure 17(b). In practice, the variation of d_0 and k_m on the MEMS wafer and V_P accuracy on IC must be taken into consideration. The major drawback of this method is that the variation of R_x is larger. In normal case, as the temperature increases, R_x increases because Q decreases. In order to compensate the resonator frequency, lower V_P will be needed at high temperature, and this low V_P makes R_x even larger than the normal case. Some experimental works are shown in Ho *et al.* (2005).

The other electrical stiffness compensation involves not only bias voltage, but also temperature-dependent gap spacing. Figure 18 presents a cross-sectional schematic conveying the basic concept behind the stiffness-compensated resonator. As shown, this device consists of a polysilicon clamped–clamped (CC)-beam resonator, over which a metal electrode spaced at a distance d_{o2} (e.g., 1000 Å) above the beam has been added as the key instrument for temperature compensation. To simplify the discussion, the anchoring structure for this top electrode is drawn such that it can freely expand in the lateral direction without introducing compressive stresses that might bend the electrode. With this simplification, the overhead electrode in Figure 18 is then effectively supported by blocks free to expand in the vertical direction.

For stiffness compensation, the material comprising these support blocks is chosen to have a thermal expansion coefficient larger than that of the structural material of the mechanical resonator and its anchors. With this selection of materials, when the temperature is increased, expansions are such that the bottom side of the top electrode moves vertically upward faster than the top of the resonator beam, resulting in a net increase in the top electrode-to-resonator gap spacing d_{o2}, as shown in the higher temperature cross section of Figure 18. This increase in the gap spacing then leads to a decrease in the top electrode-to-resonator electrical spring constant and a corresponding increase in the resonance frequency, which then counteracts the decrease in frequency caused mainly by the temperature dependence of the Young's modulus. d_2 can be expressed as

$$d_2 = d_{02} + (\alpha_{electrode} - \alpha_{res}) b_{be} (T - T_0) \qquad [50]$$

Substituting this into eqn [41] and deriving TC_f, we have

$$TC_f = \frac{\alpha_E + \alpha_{res}}{2} + \frac{3}{2}\frac{(V_P - V_C)^2 \varepsilon_0 A}{d_{02}^4 k_r}(\alpha_{electrode} - \alpha_{res}) b_{be} \qquad [51]$$

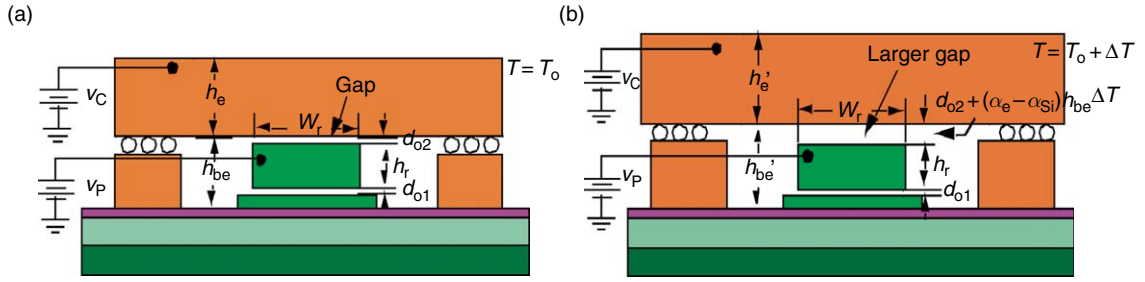


Figure 18 Cross-sectional schematics of a stiffness-compensated resonator (a) at temperature T_o and (b) at an elevated temperature $T_o + \Delta T$.

where V_C is the control voltage applied to the top electrode, k_r is the spring constant of the resonator beam including the effects of the electrical stiffness k_{e1} due to the bottom electrode ($k_r = k_m - k_{e1}$, where k_m is the unbiased mechanical stiffness), h_{be} is the gap between the substrate and the bottom of the top electrode, and $\alpha_{electrode}$ and α_{res} are the thermal expansion coefficients of the electrode and the resonator, respectively. Since the electrical stiffness generated across an electrode-to-resonator gap depends strongly on not only the gap spacing d_{o2} but also the voltage applied across the gap ($V_P - V_C$), the TC_f given by eqn [49] also exhibits a strong dependence on ($V_P - V_C$).

The first term of equation is negative and the second term is positive. Therefore, there must be a combination of $V_P - V_C$ that makes TC_f equals to zero. Plot the frequency variation versus temperature as shown in **Figure 19**, we can see that while $V_P - V_C$ is 8 V, the TC_f is close to zero. In fact, by appropriate adjustment of V_C, the electrical stiffness as a function of temperature can be conveniently optimized to more effectively null the TC_f of the resonator beam – a distinct advantage over other compensation methods.

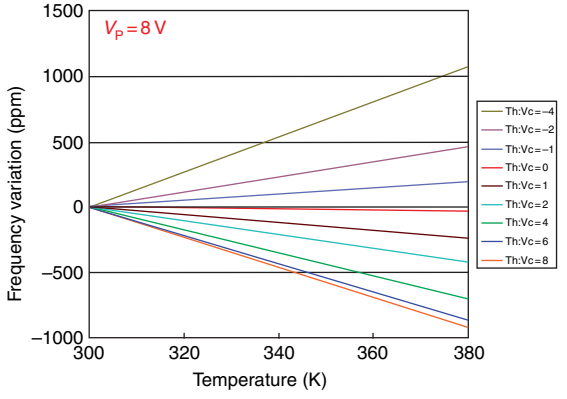


Figure 19 Electrical stiffness compensation plot based on eqn [51].

In practical renditions, the compensation electrode on top of the resonator needs to be designed to avoid bending caused mainly by compressive stresses exerted by the anchors supporting the top electrode. If the top electrode is allowed to bend, the frequency versus temperature function will deviate from the linear behavior predicted by eqn [49]. In particular, depending upon the degree and the rate of bending, the frequency versus temperature curve might look parabolic or might start curving at some high threshold temperatures. Experimental work can be seen in Hsu and Nguyen (2002). The best-achieved compensated resonator shows the TC_f as low as −0.24 ppm per degree Celsius.

3.17.5.2 Digital Trimming and Compensation

Typically the requirement of oscillator frequency accuracy is on the order of 1–10 ppm, while the process variation contributes about 1% frequency variation. Therefore, the basic idea of digital compensation is not only to adjust the frequency across the temperature, but also to set the frequency to the target (Hsu *et al.* 2006).

Figure 20 shows the system diagram of a digital-compensated MEMS oscillator. As shown, a MEMS

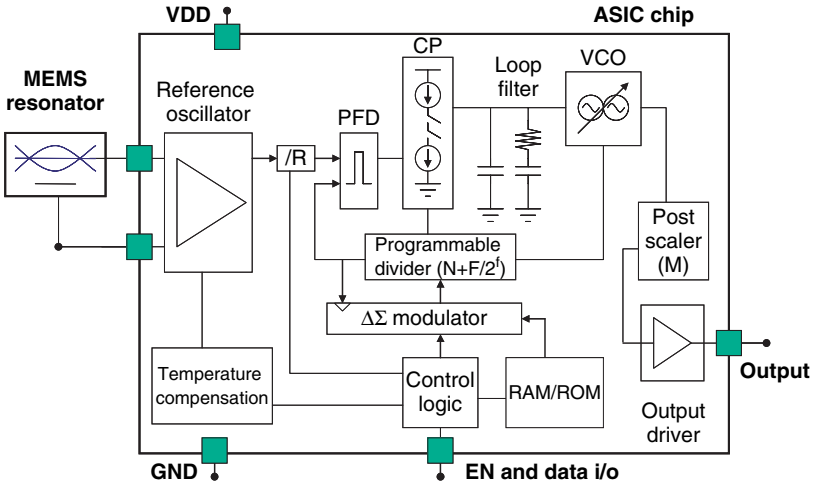


Figure 20 System diagram of a digital compensated MEMS oscillator.

resonator is connected to an IC. The reference oscillator, which is of similar design to that described in Section 3.17.3, accommodates the larger motional resistance and smaller phase shifts seen in MEMS resonators over the temperature range and over all process corner conditions. The output frequency of the reference oscillator is then fed into a fractional-N synthesizer. The main loop for the fractional-N synthesizer consists of a charge pump, a phase-frequency detector, an on-chip loop filter, a wideband voltage controlled oscillator (VCO), and a programmable divider. The programmable divider is modulated by the $\Sigma\Delta$ block. Digital control, programming, and temperature compensation are all monitored by the control circuits. Control logic controls the blocks based on the temperature. All functionality and communication are achieved through four pins, Enable, VDD, GND, and Output, which are the standard oscillator pins.

To calibrate this system, at first, the output frequency of the reference oscillator is measured at room temperature. This frequency is compared with the target frequency. Then the control logic will calculate the numbers to be set in VCO, dividers, and postscaler and store these numbers in the memory. Same steps are performed at another temperature. Based on these two sets of parameters, a look-up table can be easily established simply because the resonator shows monotonic temperature characteristics. The line that connects the two calibration points can set a line with proper slope (TC_f) to extended temperatures. Thus the output frequency of the oscillator is set to the target and is adjusted according to the temperature.

Figure 21 shows the temperature dependence of a digital-compensated programmable MEMS oscillator. It shows total 7 ppm frequency variation in the typical commercial range of temperature (−40°C to 85°C). Compared to digital-compensated programmable quartz XOs with total 25 ppm frequency variation across the temperature, clearly MEMS oscillators are qualified for XO level of applications listed in **Table 1**.

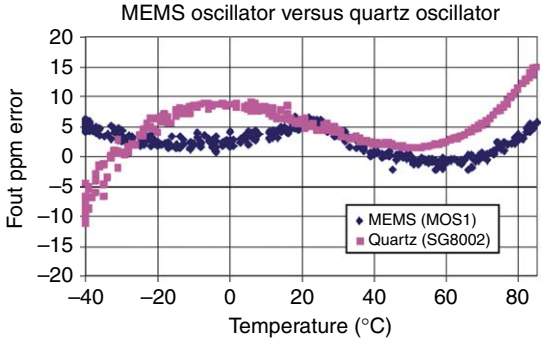


Figure 21 Typical temperature characteristics of a digital-compensated MEMS oscillator.

However, although digital compensation is the best and the easiest temperature compensation method, one drawback of using digital temperature compensation is that the phase-locked loop (PLL) synthesizer and VCO introduced more phase noise into the oscillator output spectrum. This increase in phase noise current prevents MEMS oscillators from having low phase noise and radio frequency (RF) applications. Therefore, further research is needed.

3.17.6 The Future: Research and Business

So far, MEMS oscillators have met all the requirements with respect to phase noise, jitter, temperature stability, power consumption, and reliability (Hsu 2006) for the mainstream 1–125-MHz XO segment. This segment covers high-volume applications including consumer electronics, I/Os such as USB, computing. As indicated in **Table 1**, the next application target for MEMS oscillators is to reach the accuracy requirements of TCXO, which typically exhibits frequency variation of 3 ppm across all the environmental parameters, and $-130\,\mathrm{dBc\,Hz^{-1}}$ at 1-kHz offset and $-150\,\mathrm{dBc\,Hz^{-1}}$ far from carrier of phase noise. Most importantly, the power consumption of the oscillator needs to be low.

Three requirements need to be satisfied in order to promote MEMS oscillators for TCXO applications: (1) temperature stability, (2) low phase noise, and (3) frequency accuracy. MEMS oscillators from different research groups may have satisfied individual specifications, but none has demonstrated meeting three of them at the same time. As a result, this is one of the first and probably the most important research goals for MEMS oscillators.

With the development efforts in universities and in the industry, the MEMS oscillator is getting closer to RF applications. The good news is, due to the reduction trend of cost and size, RF system designers are tweaking system specifications to use low-cost (meaning lower frequency accuracy) and integrated frequency reference sources. As MEMS oscillator researchers are trying to improve the accuracy, system designers are reducing the system

requirements. We believe, in the near future, that these two specs will merge together.

The promise of MEMS resonators has been anticipated for a long time. Finally, the technology is commercially viable and is being rolled out as we speak (by the end of 2006). Industrial analysts have created a new category for MEMS resonator in their reports and expect that this will be a high-volume product. While quartz crystal technology is very important, relevant, and will continue to have its share of the market, that market share will undoubtedly decline over time. MEMS resonators have an opportunity to make a big impact on the market, and will grow and evolve over time. With the ability to put multiple MEMS resonators into footprints the size of quartz crystal resonators, manufacturers will bring new revolutionary products to market.

References

Abdelmoneum M A, Demirci M U, Nguyen C T-C 2003 Stemless wine-glass-mode disk micromechanical resonators. *Proceedings of the 16th International IEEE Micro Electro Mechanical Systems Conference*, Kyoto, Japan, January 19–23, 2003, pp. 698–701

Chiao M, Lin L 2004 Post-packaging frequency tuning of microresonators by pulsed laser deposition. *J. Micromech. Microeng.* **14**, 1742–7

Cioffi K, Hsu W-T 2005 32-kHz MEMS based oscillator for ultra-low power applications. *Technical Digest, IEEE Frequency Control Symposium*, Vancouver, BC, Canada, September 29–31, 2005

Clark J R Hsu W-T, Nguyen C T-C 2000 High-*Q* VHF micromechanical contour-mode disk resonators. *Technical Digest, IEEE International Electronic Devices Meeting*, San Francisco, CA, USA, December 11–13, 2000, pp. 399–402

Clark J R, Hsu W T, Nguyen C T-C 2005 High-*Q* UHF micromechanical radial-contour mode disk resonators. *J. Microelectromech. Syst.* **14(6)**, 1298–310

Clark J R, Pai M, Wissman B, He G, Hsu W-T 2006 Parallel-coupled square-resonator micromechanical filter arrays. *Technical Digest, IEEE Frequency Control Symposium 2006*, Miami, FL, USA, June 4–7, 2006, pp. 485–90

Demirci M U, Abdelmoneum M A, Nguyen C T-C 2003 Mechanically corner-coupled square microresonator array for reduced series motional resistance. *Digest of Technical Papers, Transducers '03*, Boston, MA, USA, pp. 955–8

Fan L S, Crawford L 1993 Spring-softening effect in MEMS microstructures. *Dig. Tech. Pap., IEEE Int. Conf. Solid-State Sensors and Actuators*, Yokohama, Japan, June 1993, pp. 767–70

Franke A E, Howe R T 1999, Post-CMOS integration of germanium microstructures. *Tech. Dig., 12th IEEE Int. Conf. Micro Electro Mechanical Systems*, Orlando, FL, USA, January 17–21, 1999 pp. 630–7

Hao Z, Pourkamali S, Ayazi F 2004 VHF single crystal silicon elliptic bulk-mode capacitive disk resonators; Part I: Design and modeling. *IEEE J. Microelectromech. Syst.* **13(6)**, 1043–53

Ho G, Sundaresan K, Pourkamali S, Ayazi F 2005 Low-motional-impedance highly-tunable I2 resonator for temperature compensated reference oscillators. *Tech. Dig., 18th IEEE Int. Conf. Micro Electro Mechanical Systems*, Miami, FL, USA, Jan 30–Feb 3, 2005, pp. 116–19

Hsu W-T 2006 Reliability of silicon resonator oscillators. *Technical Digest, IEEE Frequency Control Symposium 2006*, Miami FL, USA, June 4–7, 2006

Hsu W-T, Brown A R 2007 Frequency trimming of MEMS oscillators. *IEEE Frequency Control Symposium 2007*, Geneva, Switzerland, June 2007, to be published

Hsu W-T, Cioffi K 2004 Low phase-noise 70-MHz micromechanical oscillators. *International Microwave Symposium*, Fort Worth, TX, USA, June 2004, pp. 1927–30

Hsu W-T, Nguyen C T-C 1998 Geometric stress compensation for enhanced thermal stability in micromechanical resonators. *IEEE International Ultrasonics Symposium*, Sendai, Japan, October 5–8, 1998

Hsu W-T, Nguyen C T-C 2002 Stiffness-compensated temperature-insensitive micromechanical resonators. *Technical Digest, IEEE MEMS Conference*, Las Vegas, NV, USA, January 21–25, 2002, pp. 731–4, to be published

Hsu W-T, Clark J R, Nguyen C T-C 2000 Mechanically temperature compensated flexural-mode micromechanical resonators. *Technical Digest, IEEE Int. Electron Devices Meeting*, San Francisco, California, Dec. 11–13, pp. 493–6

Hsu W-T, Clark J R, Nguyen C T-C 2001a A sub-micron capacitive gap process for multiple-metal-electrode lateral micromechanical resonators. *Tech. Dig., 14th IEEE Int. Conf. Micro Electro Mechanical Systems*, Interlaken, Switzerland, January 21–25, 2001, pp. 349–52

Hsu W-T, Clark J R, Nguyen C T-C 2001b *Q*-optimized lateral FF-beam micromechanical resonators. *Dig. Tech. Pap., 11th Int. Conf. Solid State Sensors and Actuators*, Munich, Germany, June 10–14, 2001

Hsu W-T, Brown A R, Cioffi K 2006 A programmable MEMS FSK transmitter. *Technical Digest, International Solid State Circuit Conference (ISSCC) 2006*, San Francisco, CA, USA, February 7–8, 2006, sec. 16.2, accepted to be published

Kaajakari V, Mattila T, Oja A, Kiihamaki J, Kattelus H, Koskenvuori M, Rantakari P, Tittonen I, Seppa H 2003 Square-extensional mode single-crystal silicon micromechanical RF-resonator. *Proc. Transducers '03, 12th Int. Conf. Solid-State Sensors and Actuators*, Boston, MA, USA, June 8–12, 2003, pp. 425–32

Kim B, Candler R N, Hopcroft M, Agarwal M, Park W T, Kenny T W 2005 Frequency stability of wafer-scale encapsulated MEMS resonators. *Proceedings, Transducers '05*, Seoul, Korea, pp. 1965–8

Lin Y-W, Lee S, Ren Z, Nguyen C T-C 2003 Series resonant micromechanical resonator oscillator. *Technical Digest, IEDM '03,* Washington DC, USA December 2003

Lin Y-W, Lis S, Ren Z, Nguyen C T-C 2005 Low phase noise array-composite micromechanical wine-glass disk oscillator. *Technical Digest, International Electron Device Meeting*, Washington DC, USA, December 5–7, 2005, pp. 4

Nathanson H, Newell W E, Wickstrom R A, Davis J R, Jr. 1967 The resonant gate transistor. *IEEE Trans. Electron. Devices* **ED-14(3)**, 117–33

Nguyen C T-C, Howe R T CMOS micromechanical resonator oscillator*Technical Digest, IEEE International Electron Devices Meeting*, Washington, DC, USA, December 5–8, 1993, pp. 199–202

Pacheco S, Zurcher P, Young S, Weston D, Dauksher W 2004 RF MEMS resonator for CMOS back-end-of-line integration. *Technical Digest, Silicon Monolithic Integrated Circuits in RF Systems, 2004*, September 8–10, 2004, pp. 203–6

Parker E R, Thiebeault J, Aimi M F, Rao M P, MacDonald N C 2005 Inductively coupled plasma etching of bulk titanium for MEMS applications. *J. Electrochem. Soc.* **152**, 675–83

Piazza G, Stephanou P J, Porter J M, Wijesundara M B J, Pisano A P 2005 Low motional resistance ring-shaped

contour-mode AlN piezoelectric micromechanical resonator for UHF applications. *MEMS '05*, Miami, FL, USA, January 30–February 3, 2005

Pourkamali S, Hashimura A, Abdolvand R, Ho G, Erbil A, Ayazi F 2003 High-*Q* single crystal silicon HARPSS capacitive beam resonators with sub-micron transduction gaps. *IEEE J. Microelectromech. Syst.* **12(4)**, 487–96

Pourkamali S, Hao Z, Ayazi F 2004 VHF single crystal silicon elliptic bulk-mode capacitive disk resonators; Part II: Implementation and characterization. *IEEE J. Microelectromech. Syst.* **13(6)**, 1054–62

Roessig T A, Howe R T, Pisano A P, Smith J H 1998 Surface-micromachined 1 MHz oscillator with low-noise Pierce configuration. *8th Solid-State Sensor and Actuator Workshop*, Hilton Head, SC, USA, June 7–11, 1998, pp. 328–32

Tang W C, Nguyen T-C H, Howe R T 1989 Laterally driven polysilicon resonant microstructures. *Sens. Actuators* **20**, 25–32

Vig J 2005 Quartz crystal resonator and oscillators, http://www.ieee-uffc.org/freqcontrol/tutorials/vig2/tutorial2_files/frame.htm

Wang K, Wong A-C, Hsu W-T, Nguyen C T-C 1997 Frequency trimming and *Q*-factor enhancement of micromechanical resonators via localized filament annealing. *Dig. Tech. Pap., v(1), 1997 Int. Conf. Solid-State Sensors and Actuators*, Chicago, IL, USA, June 1997, pp. 109–12

Wang K, Wong A-C, Nguyen C T-C 2000 VHF free-free beam high-*Q* micromechanical resonators. *IEEE/ASME J. Microelectromech. Syst.,* December 2000

Wang J, Butler J E, Feygelson T, Nguyen C T-C 2004 1.51-GHz polydiamond micromechanical disk resonator with impedance-mismatched isolating support. *Proceedings, 17th International IEEE Micro Electro Mechanical Systems Conference*, Maastricht, The Netherlands, January 25–29, 2004, pp. 641–4

Biography

Dr. Wan-Thai Hsu is one of the pioneer researchers and inventors in the area of RF MEMS. Part of his doctoral work led to the birth of Discera Inc. in 2001. He is now Chief Technology Officer of Discera. Since 2001, he has built a strong technical team at Discera focusing on making silicon resonator oscillators a reliable commercial product. The efforts include robust resonator design, volume manufacturing process development, MEMS reliability, technology transfer to foundries, product developments for silicon resonator oscillators, and IP strategies.

He is now leading Discera's R&D team on resonator, filter, packaging, and system developments. He also serves as a principal investigator for multiple government-supported research programs on MEMS filters, system integration, and MEMS reliability.

Dr. Hsu received his BS and MS degrees from National Tsing-Hua University and Ph.D. from University of Michigan, all in Electrical Engineering. Dr. Hsu has published more than 25 technical papers and spoken at multiple international conferences. He owns six US patents and has four patents pending. He received *EE Times* 2007 Innovator of the Year Award.

3.18 MEMS Atomic Clocks

Svenja Knappe, National Institute of Standards and Technology (NIST), Boulder, CO, USA

Published by Elsevier B.V.

Glossary

AM Amplitude Modulation
ASIC Application-Specific Integrated Circuit
CAD Computer-Aided Design
CPT Coherent Population Trapping
CSAC Chip-Scale Atomic Clock
CSAM Chip-Scale Atomic Magnetometer
DBR Distributed Bragg Reflector
DFB Distributed Feedback
DRIE Deep Reactive Ion Etching
DRO Dielectric Resonator Oscillator
EIT Electromagnetically Induced Transparency
EOM Electro-Optical Modulator
FBAR Film Bulk Acoustic Resonator
FM Frequency Modulation
GNSS Global Navigation Satellite Systems (e.g., GPS, Galileo, Golnass)
HBAR High-tone Bulk Acoustic Resonator
ITO Indium Tin Oxide
KOH Potassium Hydroxide
LO Local Oscillator
MCXO Microprocessor-Controlled Crystal Oscillators
MEMS Microelectromechanical Systems
NEMS Nanoelectromechanical Systems
OCXO Oven-controlled Crystal Oscillator
STIRAP Stimulated Raman Adiabatic Passage
TCXO Temperature-Compensated Quartz Crystal Oscillators
UHV Ultra High Vacuum
VCO Voltage-Controlled Oscillator
VCSEL Vertical-Cavity Surface-Emitting Lasers

3.18.1 Introduction to MEMS Atomic Clocks

3.18.1.1 Introduction

Frequency references provide the base for a large number of applications such as digital communication, navigation systems, synchronization of networks, or power distribution. As the amount of data transferred increases and data rates become faster, more stringent requirements are placed on the timing systems. At the same time applications have become more mobile, increasing the demand for small low-power frequency references.

Atomic oscillators have provided the most stable frequency references for more than 50 years (Lyons 1950, Quinn 2005, Townes 1951). They provide unmatched frequency stability over long periods of time (from a few seconds to many years), because their resonance frequency is determined by the energy transition of the atoms – in contrast to crystal oscillators, where the resonance frequency is determined by the length of the crystal and is therefore much more susceptible to temperature changes, for example. But in order not to perturb the resonance frequency of the atoms and to probe it, often very complicated setups are required. Therefore, the large size, cost, and power consumption of most atomic clocks have restricted their use in real-world applications. Nevertheless, smaller versions of optically pumped rubidium standards (Bloch *et al.* 1993, Chantry *et al.* 1996, Ho *et al.* 1998, Koyama *et al.* 2000, McClelland *et al.* 1995, 1996, 1999, Rochat *et al.* 2002, Suzuki *et al.* 1998) with volumes around 100 cm^3 and power requirements of a few watts have become commercial standards. They are manufactured in the tens of thousands for use in cellular telecommunication networks (Kusters and Adams 1999). Space-qualified versions became critical members of the clock ensembles in Global Navigation Satellite Systems (GNSS). But a large number of portable in-field applications require smaller and less expensive frequency references with much lower power consumptions. Temperature-compensated quartz crystal oscillators (TCXOs) are small low-power, low-cost devices with relatively good performance over longer times, which have proven useful in such battery-operated portable applications. Nevertheless, their frequency stability at longer times (one hour to several days) is not sufficient for the requirements of many applications in the civil and military navigation and communication sector (Fruehauf 2001, Lee *et al.* 1996, Murphy and Skidmore 1994, Sturza 1984, Vig 1993). **Figure 1** summarizes typical performances of different commercially available oscillators as a function of averaging time (see Section 3.18.3.2 for an explanation of fractional frequency stability). It can be seen that crystal oscillators can be good frequency references over short times (see Section 3.18.2.3), but they are outperformed by atomic clocks at longer times.

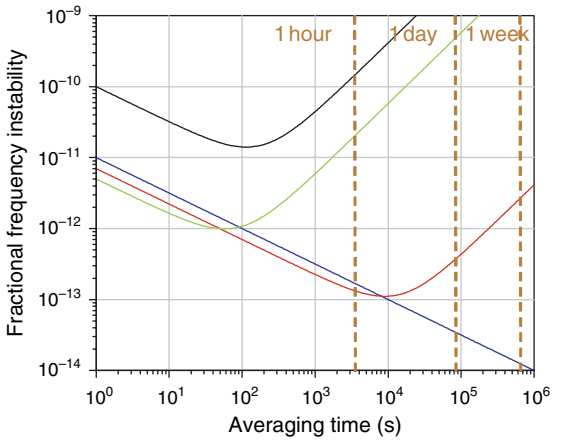


Figure 1 Typical performance of different commercially available oscillators: cesium beam standard (blue), rubidium vapor cell clock (red), oven-controlled crystal oscillator (OCXO, green), temperature-compensated crystal oscillator (TCXO, black).

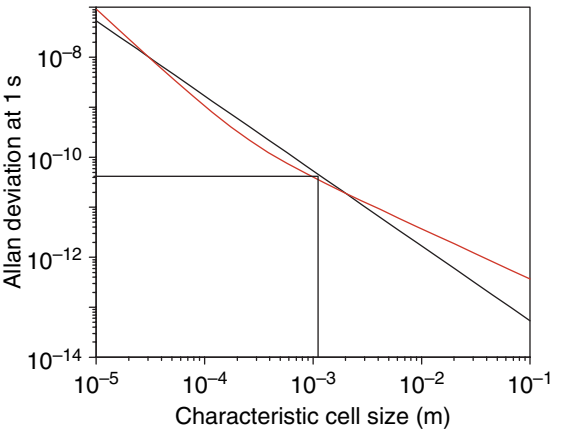


Figure 2 Estimate of the Allan deviation at 1 s as a function of characteristic cell size for a cell with a 100 kPa nitrogen buffer gas (red) or a paraffin wall coating (black).

Furthermore, it can be seen that atomic vapor cell clocks also have drifts over very long timescales (see Section 3.18.3.2.3).

The size of the optically pumped rubidium atomic frequency standards (RAFS) is largely limited by the size of the microwave cavity and the power goes mainly into heating the cell and the lamp. All-optical clocks were proposed years ago (Cyr *et al.* 1993, Levi *et al.* 1997), but the lack of reliable, low-noise, easy-to-use diode lasers at the atomic wavelengths have prevented their implementation into commercial clocks. When single-mode vertical-cavity surface-emitting laser (VCSEL) became available at 852 nm (the D_2 line of cesium) with large modulation bandwidths (King *et al.* 1998), their practicality for coherent population trapping (CPT) spectroscopy was demonstrated (Affolderbach *et al.* 2000). Soon after, the first miniature CPT clock prototype was demonstrated (Kitching *et al.* 2001a, Vanier *et al.* 2004). The idea to combine this CPT spectroscopy with fabrication methods developed for microelectromechanical systems (MEMS) into a chip-scale atomic clock (CSAC) was proposed (Kitching *et al.* 2002). MEMS fabrication techniques would allow for small size in an all-optical design and the use of a VCSEL in combination with small volume for correspondingly lower power requirements. An estimate of the fractional frequency instability of the clock stability at 1 s of integration as a function of cell size is shown in **Figure 2** for a 100-kPa N_2 buffer gas cell (red) and a paraffin wall coating (black) (Kitching *et al.* 2002). It can be seen that the fractional frequency stability of the clock degrades with smaller size, because of the more frequent collisions of the atoms with the cell walls. Nevertheless, stabilities below $1 \times 10^{-10}/\tau^{1/2}$ seem possible with a cell size of 1 mm (for an introduction to Allan deviation as a measure of frequency stability, see Section 3.18.3.2.1).

Since its proposal in 2002, the field of CSACs has rapidly progressed, largely due to the establishment of a CSAC program of the US Defense Advanced Research Projects Agency (DARPA). The goal of this program is to develop a MEMS atomic clock with a frequency instability of 1×10^{-11} at 1 h of integration in a package of volume 1 cm^3 consuming 30 mW of power. While this is still an ongoing project and the goals have not been reached yet, this chapter tries to outline some of the major developments so far. Since a majority of the work presented here was performed under the DARPA project with the goal of product development, many approaches cannot be reviewed here in detail, because no published information exists. Therefore, this review focuses mainly on MEMS clocks based on CPT, as most current approaches follow this route and a lot of published information exists. Nevertheless, other non-CPT approaches were successfully demonstrated and some of these are reviewed in Section 3.18.5.

3.18.1.2 Vapor Cell Atomic Clocks

The atoms in the gas phase are chosen for the best atomic clocks, because it is desirable to have distinct energy levels that are largely unperturbed by the

environment (as compared to the frequency bands of solids). The atoms in the first column of the periodic table, that is, alkali atoms, are favored in many atomic physics experiments because of their comparatively simple energy structure, determined largely by the single valence electron. The lowest energy electric dipole transitions in these atoms are called D-lines for historical reasons, and they correspond to changes in the distance between the valence electron and the nucleus, i.e., the transition $nS_{1/2} \leftrightarrow (n+1)P_{\mathcal{J}}$. The P state is split once again into two components of different total angular momentum $\mathcal{J}$ (**Figure 3(a)**). The atomic ground state ($S_{1/2}$) is split into two hyperfine energy components (F and F + 1), determined by the orientation of the electron spin relative to the spin of the nucleus. The clock frequency is determined by the energy difference between these hyperfine components divided by the Planck's constant h. Finally, the two hyperfine levels can be split into Zeeman sublevels in a magnetic field due to different orientations of the electron spin with respect to the magnetic field (m_F). Two of these Zeeman levels do not change their frequency to first order ($m_F = 0$ levels) in the presence of a magnetic field. These are usually chosen for the clock transition.

In an optically pumped vapor cell clock, one light field, e.g., with frequency ω_1, is present. It excites the atoms from state $|1\rangle$ into the excited state (**Figure 3(a)**). From there the atoms can decay into either of the two ground states. Eventually, a large fraction of atoms will accumulate in $|2\rangle$, a process called optical pumping (Happer 1972). The cell appears transparent because fewer atoms can absorb the light. When a microwave field with frequency $\omega_{\text{microwave}}$, equal to the clock transition frequency of the atom, is applied, the atoms can be repumped from $|2\rangle$ into $|1\rangle$. This microwave resonance can be detected through a change in the absorption of the light field.

Optical pumping in a RAFS, which are probably the most widely used atomic clocks in the world, is usually done with a rubidium discharge lamp (**Figure 3(b)**) (Vanier and Audoin 1989). It creates a broad spectrum of light, which is resonant with the transitions $5S_{1/2} \leftrightarrow 5P$. In order to enable efficient hyperfine pumping, the light needs to be spectrally filtered. This is done by using ^{85}Rb absorption, since the transition from one of the ground states in ^{85}Rb is very close to the one in ^{87}Rb. Therefore, RAFS usually have a ^{85}Rb filter cell in front of the ^{87}Rb clock cell (sometimes they are combined in the same cell though). After the filtered light passes through the ^{87}Rb cell, it is collected on a photodetector. A microwave cavity surrounds the ^{87}Rb cell to provide efficient coupling between the microwave field and the atoms and to ensure a uniform microwave phase over the entire volume of the cell. When the microwave frequency is exactly resonant with the atomic clock frequency of 6.8 GHz, the transmitted light decreases as atoms are pumped back into the depopulated hyperfine state. **Figure 3(c)** shows the transmitted light as a function of microwave frequency. Such a resonance can have Q factors of $\sim 10^8$.

RAFS can reach frequency stabilities in the range of 10^{-11} to $10^{-12}/\tau^{1/2}$ with centimeter cell sizes. At long times their frequency shows slow drifts around 10^{-14} per day; the detailed causes for this are still under investigation, but the diffusion of helium through the cell walls seems to be the one likely candidate

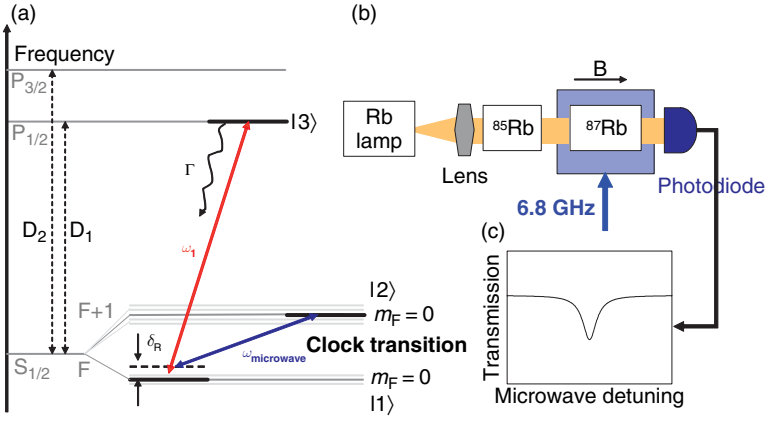


Figure 3 Schematic of the simplified atomic energy level configuration for the observation of (a) optically pumped microwave resonances, (b) spectroscopic setup, and (c) picture of a microwave resonance.

(Camparo 2004). The most power-consuming component in a RAFS is probably the Rb lamp. It has to be heated to high temperatures and usually consumes several hundred milliwatts to watts of power. The size of the RAFS physics package is largely determined by the size of the microwave cavity, since the clock frequency corresponds to ~4 cm. Therefore, it is challenging to miniaturize the current setup below the size of a deck of cards and to reduce the power consumption to below 1 W. In order to reduce the power consumption and the size of a vapor cell atomic clock, it would thus be useful to replace the lamp and the filter cell with a low-power laser and eliminate the microwave cavity.

3.18.1.3 Coherent Population Trapping

Heuristically, coherent population trapping (CPT) can be seen as a destructive quantum interference that happens, when the beat frequency between the two resonant light fields equals the ground state splitting frequency of the atoms. This destructive interference process largely reduces the probability that the atom can be excited by either of the light fields, so it appears dark, i.e., the resonance fluorescence is suppressed. CPT resonances have been studied for three decades. Alzetta *et al.* (1976) observed dark lines in the fluorescence emitted by a beam of sodium atoms. Simultaneously, the phenomenon was theoretically explained for a cascade system (Whitley and Stroud 1976) and for a Λ-System (Arimondo and Orriols 1976, Gray *et al.* 1978). Since then, the phenomenon of CPT and its related effect, electromagnetically induced transparency (EIT) (Fleischhauer *et al.* 2005, Harris 1997, Harris *et al.* 1990, Kasapi *et al.* 1995, Marangos 1998), have been suggested for a large number of applications such as optical bistability (Walls and Zoller 1980), laser cooling of atoms and ions (Aspect *et al.* 1988, Schmidt-Kaler *et al.* 2001), and lasing without inversion (Bentley Jr. and Liu 1999, Kocharovskaya 1992). The steep dispersion at CPT resonance at reduced absorption makes it attractive, for example, for the enhancement of nonlinear effects (Hemmer *et al.* 1995), slow light (Schmidt *et al.* 1996, Vestergaard Hau *et al.* 1999), optical data storage (Liu *et al.* 2001, Phillips *et al.* 2001), and gravitational wave detectors (Mueller *et al.* 1997). Finally, the narrow resonance linewidths (Brandt *et al.* 1997, Erhard *et al.* 2000, Merimaa *et al.* 2003) that can be achieved with CPT make it attractive for metrology applications (Tench *et al.* 1981, Thomas *et al.* 1982) where high-resolution laser spectroscopy can be applied to all-optical atomic clocks (Hemmer *et al.* 1983) and magnetometers (Bloom 1962, Scully and Fleischhauer 1992).

Although alkali atoms have a complicated structure of energy levels, the calculations for a simple system consisting of only three levels show surprisingly good agreement with the measurements (we will limit the introduction here to a very simplified picture for illustrative purposes only; for a review on CPT, see Arimondo 1996a). In these calculations, the so-called lambda system consists of two long-lived ground states $|1\rangle$ and $|2\rangle$, split by the clock frequency Δ_{HFS}, and an excited state $|3\rangle$ (**Figure 4(a)**, resembling the Greek letter Λ). Two light fields E_1 and E_2 with frequencies ω_1 and ω_2 and phases φ_1 and φ_2 can couple the electric dipole transitions $|1\rangle \leftrightarrow |3\rangle$ and $|2\rangle \leftrightarrow |3\rangle$. Both ground states have nearly equal populations in thermal

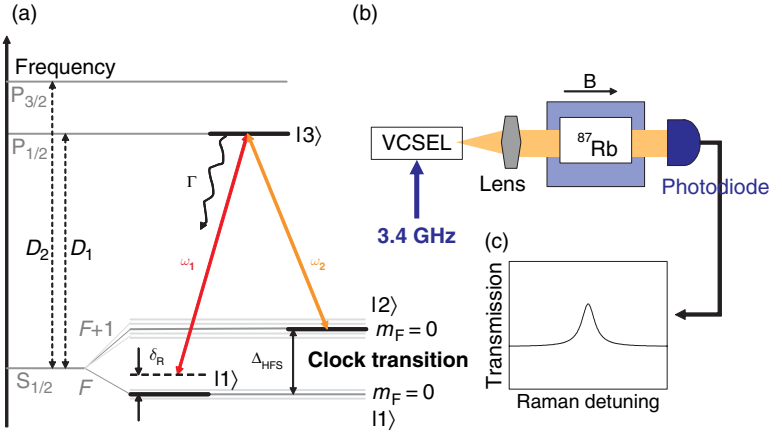


Figure 4 Schematic of the simplified atomic energy level configuration for the observation of (a) coherent population trapping (CPT) resonances (b) spectroscopic setup, and (c) picture of a CPT resonance.

equilibrium. A similar pumping as in microwave clocks can be performed between the ground states $|1\rangle$ and $|2\rangle$, when the two light fields E_1 and E_2 are resonant with the respective optical transitions (**Figure 4(a)**). If the beat frequency $\omega_1 - \omega_2$ is equal to the clock frequency Δ_{HFS} then the atoms can be pumped into a so-called coherent dark state $|NC\rangle$.

For very simplified illustration purposes, we can choose a new basis $\{|C\rangle, |NC\rangle, |3\rangle\}$, where the coupled state $|C\rangle$ and the noncoupled state $|NC\rangle$ are superpositions of the atomic ground states (with notations from **Figure 4(a)**):

$$|C\rangle(t) \propto g_1|1\rangle + g_2 \exp(-i\Delta_{HFS}t + i\phi)|2\rangle \quad [1]$$

$$|NC\rangle(t) \propto g_2|1\rangle - g_1 \exp(-i\Delta_{HFS}t + i\phi)|2\rangle \quad [2]$$

Here, g_1 and g_2 are the so-called Rabi frequencies, which determine the interaction strength of the transition, and they are proportional to the amplitude of the respective light fields combined with atomic parameters. Furthermore, the time-dependent states have a frequency of Δ_{HFS} and a phase of ϕ. It can be shown that if the difference frequency between the light fields $\omega_1 - \omega_2$ is equal to the ground-state splitting frequency Δ_{HFS}, and if the relative phase of the dark state ϕ is equal to the relative phase between the two light fields $\varphi_2 - \varphi_1$, then the probability for an atom to be excited from $|NC\rangle$ by the light fields is zero.

In this case, a ground-state coherence is created that is out of phase with the beat between the driving light fields. The relative phase between the two light fields has to be stable in order for the state to remain dark and not to get an admixture of the coupled state $|C\rangle$. At the same time this places a constraint on the stability of the relative phase only (i.e., the microwave phase), and not on the individual phases of the light fields φ_1 and φ_2.

While this approach gives an intuitive explanation of the existence of a coherent dark state, it does not include the population of this state by spontaneous emission. It also fails to predict detailed lineshapes, dephasing, and decay rates. In order to include these effects, the density matrix approach has proven useful (Arimondo 1996b, Orriols 1979, Vanier *et al.* 1998)) for a three-level (Arimondo 1996a) or a four-level (Vanier *et al.* 2003c) system. When the Raman detuning $\delta_R = \omega_1 - \omega_2 - \Delta_{HFS}$ is scanned around zero, the CPT resonance is detected as a bright line in the transmission of the light fields (see **Figure 4(c)** for the transmission ($\sim 1 -$ absorption)). Many of the features of CPT resonances are similar to the ones of microwave resonances used in RAFS.

Again, the width of the CPT resonance is inversely proportional to the lifetime of the coherence. Since the decay times of the ground-state hyperfine levels are several thousand years for stable alkali atoms, the decoherence is determined by external factors such as collisions, magnetic field inhomogeneities, power broadening, and relative phase instability of the light fields.

To reduce the last broadening mechanism, two phase-stable optical light fields are used with a frequency splitting of a few gigahertz, which is tuned to match the ground-state hyperfine splitting of the atoms. This has been demonstrated by actively phase-locking (Enloe and Rodda 1965, Schmidt *et al.* 1996) two diode lasers, external modulation of one laser (Bouyer *et al.* 1996, Hemmer *et al.* 1983), injection locking of two lasers (Simpson *et al.* 2003), or direct modulation of the diode laser injection current (Cyr *et al.* 1993, Hemmer *et al.* 1993, Levi *et al.* 1997, Myatt *et al.* 1993, Stähler *et al.* 2002). While the first two methods are difficult to implement in a small low-power device for field applications that require ease of use and robustness, the last methods seem to be suitable for this. It requires lasers with large modulation bandwidths for most alkali atoms. Recently, distributed feedback (DFB) lasers have become available at Rb and Cs D_1 and D_2 lines with modulation bandwidths around 10 GHz. A simple and cheap technique was demonstrated by Affolderbach *et al.* (2000), using a single-mode VCSEL modulated at 9.2 GHz, in order to create sidebands for Cs CPT. Despite their broad linewidths of a few tens of megahertz, narrow CPT resonances have been measured. As has been emphasized before, the relative phase stability (given by the modulation) is more important to the CPT resonance than is the overall frequency stability of the light fields. The low operating power of VCSELs, their large modulation bandwidths, and vertical design make them well suited for small, low-power MEMS devices.

When miniaturizing an atomic clock, CPT systems seemed suitable, mainly because of the simplicity of the setup. It is challenging to reduce the size of a microwave cavity very much below the size of the associated microwave wavelength, e.g., 4 cm for ^{87}Rb. In contrast, there is no such size restriction for CPT, since the microwave frequency is carried through the beat frequency of the optical light fields (at least at sizes much larger than the optical wavelength). Since then, small microwave clocks have been demonstrated as well without the microwave cavities (see Section 3.18.5).

3.18.1.4 CPT in Small Vapor Cells

Extensive studies have compared the advantages and disadvantages of CPT over the microwave approach (Lutwak *et al.* 2002, Vanier *et al.* 2003e). Despite many similarities, there are substantial differences between the properties of CPT and optically pumped microwave resonances. First, in optically pumped clocks, a population imbalance must be created between the hyperfine levels. This is usually done by a single light field resonant with one of the transitions, which does not require phase coherence but makes the system more susceptible to light field-induced frequency shifts (Camparo *et al.* 1983, Mileti *et al.* 1998, Orriols 1979, Vanier *et al.* 1998). At the same time CPT resonances tend to have lower contrast than microwave resonances, because background light fields are often present and destructive interference between multiple excited states can exist.

Finally, since the microwave transition is created by a single photon of spin 1, only ground states with magnetic quantum numbers differing by $\Delta m = 0, \pm 1$ can be coupled. In CPT two light fields are used to couple the ground states, which allows for $\Delta m = 0$, ± 1, ± 2 (see Section 3.18.3.2.3.(i)).

In the following section, we consider what happens when the size of the vapor cell is reduced. For simplicity, we assume that all the three dimensions are equal to L. As we reduce the size of the cell, fewer atoms will be interrogated and less light will be absorbed. It has been found that the maximum CPT signal is obtained when roughly 50% of the light gets absorbed by the atoms (Godone *et al.* 2002, Knappe *et al.* 2002). In order to stay close to this optimum condition, the temperature, and thus the vapor pressure of the alkali atoms, needs to be increased for smaller cells.

A major contribution to the CPT linewidth are collisions of the alkali atoms with the cell walls. To reduce this effect, a buffer gas is added to the vapor cell, which increases the time for an alkali atom to reach the cell wall and eliminates residual Doppler broadening through Lamb–Dicke narrowing (Dicke 1953, Vanier and Audoin 1989). Usually, inert gases or molecules such as nitrogen or methane are used as buffer gases, because they have small cross sections σ_{puff} for the ground-state decoherence (Beverini *et al.* 1971). This causes a diffusive motion of the alkali atoms through the cell, and the CPT linewidth contributions can be approximated in the lowest-order diffusion mode by the following equation

$$\Delta \nu_{\text{CPT}}^{\text{diff}} \propto \frac{D_0}{L^2 p} \quad \text{and} \quad \Delta \nu_{\text{CPT}}^{\text{puff}} \propto \sigma_{\text{buff}}\, p \qquad [3]$$

for the diffusion and the buffer gas collision as a function of buffer gas pressure p. Here D_0 is the diffusion constant (see, e.g., Vanier and Audoin 1989, 1992 for values of D_0 and σ_{buff} and various alkali–buffer gas combinations). Thus, the linewidth is determined at low buffer gas pressures by the collisions of the alkali atoms with the cell walls, and at high pressures by collisions with the buffer gas atoms. It is a standard method to choose the buffer gas pressure to operate at the sweet spot of lowest linewidth (Beverini *et al.* 1971, Brandt *et al.* 1997) (**Figure 5**). It has been experimentally shown that the CPT resonance width agrees well with these predictions even at very small cell sizes of 100 µm (Knappe *et al.* 2004a). At the same time it means that when the operation is at the sweet spot, the linewidth increases with cell size like L^{-1} (Kitching *et al.* 2002).

CPT resonances have also been measured in large, but extremely thin cells with thicknesses below the optical wavelength without buffer gases (Sargsyan *et al.* 2006). This was done in a different regime, where the dimensions perpendicular to the laser beam had to remain large so that the linewidth was determined mainly by atoms with a small velocity component in the direction of the laser beam.

Although the buffer gases usually have a small depolarization cross section of the ground state, they cause substantial depolarization in the excited state, causing the P-state to broaden homogeneously at a few megahertz per kilopascal of pressure (Allard and Kielkopf 1982). If the excited-state decoherence rate Γ increases, more laser intensity is required to obtain the same amplitude and linewidth of the CPT resonance.

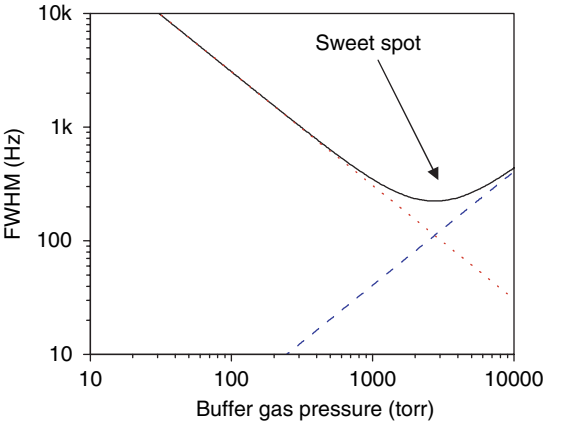


Figure 5 Coherent population trapping (CPT) linewidth as a function of buffer gas pressure in a 1-mm^3 cell filled with ^{87}Rb and N_2 as a result of diffusion to the walls (red dotted line), buffer gas collisions (dashed blue line), and the sum of both (black line).

At the same time, the absorption cross section of the atoms decreases so that the alkali vapor pressure, i.e., cell temperature, is increased to maintain a constant absorption. With an increase in alkali density, the rate of alkali–alkali spin-exchange collisions increases (Beverini *et al.* 1971, Happer 1972), i.e., rate is proportional to the alkali density. At sufficiently high pressures, spin-exchange collisions add a major contribution to the linewidth, rising proportional to the buffer gas pressure. In most vapor cells a molecular buffer gas, e.g., nitrogen, is included to quench the spontaneous emission and prevent radiation trapping (Copley and Krause 1969, Kibble *et al.* 1967).

Finally, the broadening of the excited state puts another constraint on very small CPT clocks. Since the bichromatic light field is usually produced by frequency-modulating a VCSEL at half the ground-state hyperfine frequency, the contrast of the CPT resonance is drastically reduced if the optical linewidth approaches the ground-state splitting frequency (Arimondo 1996a, Nikonov *et al.* 1994, Post *et al.* 2005). At these pressures the carrier and the second-order sidebands become resonant and can cause single-photon transitions out of the dark state (**Figure 6**). Furthermore, the beat frequency between the two first-order sidebands is out of phase with the ones between the carrier and the second-order sidebands. Since both beats have the same frequency, they could create dark states in the atoms, but with $\varphi_1 - \varphi_2$ having a difference of π, they cannot be dark simultaneously (eqns [1] and [2]). The phase relationships between the sidebands are symbolized in **Figure 6** by the direction of the arrows. Both effects reduce the lifetime of the coherence and degrade the stability of the clock.

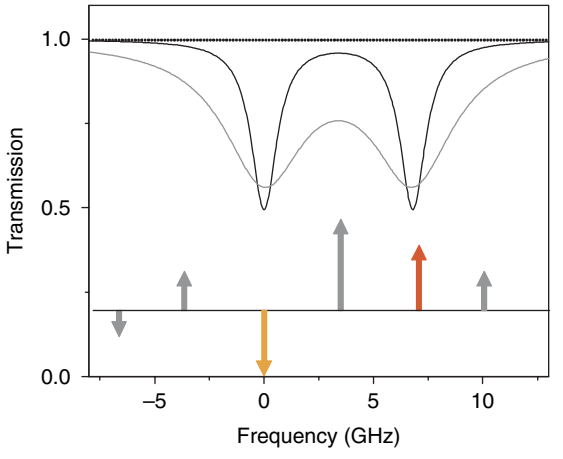


Figure 6 Transmission as a function of laser frequency relative to the transition of an atom with two ground states split by 6.8 GHz without buffer gas (black) and with homogeneously broadened lines (gray). The bottom arrows represent the sidebands created when the laser is modulated at 3.4 GHz and tuned such that the first-order sidebands (red and orange) are in resonance with the two atomic transitions. The direction of the arrows symbolizes the relative phase between them.

When frequency modulation (FM) of the VCSEL current is used, this places an upper limit on the buffer gas pressure. Thus, it is not necessarily favorable to operate at the sweet spot. The various parametric dependencies of the CPT clock resonance signals with cell parameters (such as buffer gas pressure, cell temperature, laser intensity, optical detuning, modulation frequency, and amplitude) have been characterized in detail (Knappe 2001, Knappe *et al.* 2001, Vanier 2001).

3.18.2 Design and Fabrication

3.18.2.1 Introduction

Passive CASCs in their simplest form consist of three parts: a local oscillator (LO) that generates the clock signal, a physics package that compares the frequency of the LO to the internal frequency of the atoms and outputs information about the difference frequency, and control electronics that tune the LO frequency such that the difference frequency is zero. Most CSACs use Cs or ^{87}Rb in their vapor cells, because of their relatively high ground-state splitting frequency (good Q factors), the availability of VCSELs at the right wavelengths, and relatively low vapor pressures.

In CPT-based CSACs, the LO usually produces a signal at half the ground-state hyperfine frequency of these atoms and modulates the injection current of a VCSEL. The frequency of the VCSEL is then tuned such that the two first-order modulation sidebands are in resonance with the two transitions from the ground-state hyperfine components to the excited state. Often, the buffer gas pressure is chosen to be above 5 kPa, so that the excited-state hyperfine structure is unresolved. The light is circularly polarized, attenuated, and sent through the alkali vapor cell. A photodiode detects the transmitted light. A small magnetic field parallel to the laser beam is applied to lift the degeneracy of the ground-state Zeeman levels and the cell is placed inside a magnetic shield to ensure a constant magnetic field strength. A schematic of this simple setup is depicted in **Figure 7**.

When the LO frequency is tuned to exactly half the ground-state frequency between the $m_F = 0$

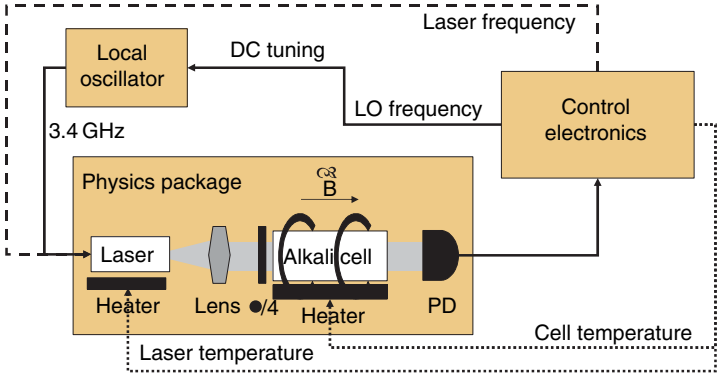


Figure 7 Schematic of a chip-scale atomic clock (CSAC) consisting of a local oscillator (LO), a physics package, and control electronics. The physics package consists of a laser with heater, lens, a quarter waveplate ($\lambda/4$), a vapor cell with heater, a magnetic offset field (B), and a photodetector (PD). The control electronics consist of servo loops for laser and cell temperature, as well as laser and LO frequency.

Zeeman components, the atoms are pumped into a coherent dark state and the transmitted light is reduced. To stabilize the LO frequency onto the top of the CPT resonance, its frequency is modulated at a few kilohertz. Phase-sensitive detection of the transmitted light at this frequency produces a dispersive error signal that can be used to lock the LO to the atomic resonance. A similar technique of phase-sensitive detection is used to lock the laser wavelength onto the center of the optical absorption by modulating the laser current at a low frequency. It is important that both of these modulation frequencies used for locking are well separated to avoid cross-talk between both loops. In addition to the two frequency-lock loops, two temperature servo loops are required to stabilize cell and laser temperatures. In some designs, the laser and the cell are in good thermal contact and one of the four feedback loops is eliminated (Lutwak *et al.* 2004).

3.18.2.2 Physics Package

3.18.2.2.1 Introduction

The physics package takes the instable 3.4-GHz signal from the LO and compares it with the internal frequency of the atoms. It generates an output signal that determines how much the LO frequency differs from the internal frequency of the atoms. In order to miniaturize a MEMS clock, various approaches have been investigated. Most of these, however, include the same general components and differ mostly in the engineered design. The first microfabricated physics package was reported in 2004 by the National Institute of Standards and Technology (NIST) (Knappe *et al.* 2004b). A picture is shown in Figure 8. It consisted of the VCSEL at the bottom, a micro-optical assembly, a vapor cell between two heaters, and a photodiode on the top. The electrical interconnects were wire bonds from the components in the stack to the pads on the baseplate.

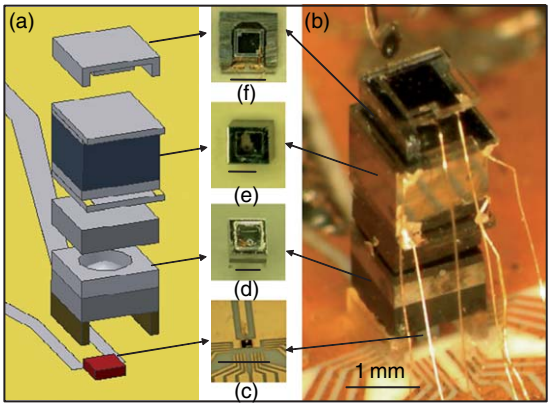


Figure 8 (a) Schematic and (b) photograph of a chip-scale atomic clock (CSAC) physics package consisting of (c) a vertical-cavity surface-emitting lasers (VCSEL), (d) an optics package, (e) a vapor cell with heaters, and (f) a photodetector.

In this approach, all the components were planar and in principle can be fabricated as arrays on individual wafers. While currently the wafers are diced first and the components stacked afterward, the design potentially allows for assembling the wafers first prior to dicing them into many individual physics packages, as indicated in Figure 9. This enables simple exchange of components and will potentially reduce fabrication costs.

Other groups have taken more complex approaches, where the laser beam is not just passed in a straight line to the photodetector. Lutwak *et al.*

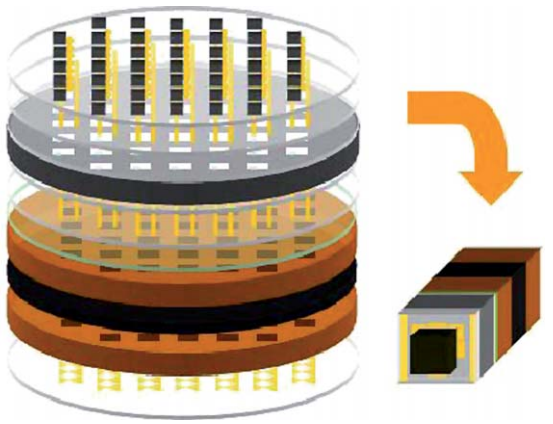


Figure 9 Wafer-level assembly of microelectromechanical systems (MEMS) clocks: wafers of baseplates with vertical-cavity surface-emitting lasers (VCSELs), spacers, neutral density (ND) filters with lenses, spacers, quarter waveplates, heaters, cells, and photodetectors are stacked and diced into individual chip-scale atomic clocks (CSACs) afterward.

(2004), for example, used a diverging beam and retro-reflected it back through the cell, where it was detected by a circular photodiode around the VCSEL die. One group explored a self-powered photon source as a low-power alternative to a VCSEL (Guo and Lal 2003). The kinetic energy of the electrons of 17.1 keV emitted from a radioactive ^{63}Ni film was used to generate photons through collisions with atoms. Photon generation in air, neon, and xenon was demonstrated with the goal to generate a low-power rubidium lamp for CSACs. Cesium, as well as rubidium, has been tried, and the fabrication of vapor cells differs. The cells vary in size between 250 µm and 1.5 mm. In the following sections as one example, we focus mainly on the approach taken by NIST, but there are other important ideas and approaches too.

3.18.2.2.2 Vertical-cavity surface-emitting laser

The VCSEL translates the electrical microwave signal from the LO into an optical beat signal. When the injection current of the VCSEL is modulated, a comb of modulation sidebands is created around the carrier in the optical spectrum spaced by the modulation frequency. The modulation properties of the VCSEL are therefore critical (see, e.g., Larsson *et al.* 2004). In order to interact with the atoms, the wavelength of the VCSEL has to be tuned such that some sidebands are resonant with optical atomic transitions.

VCSELs have been developed for telecom applications with high modulation bandwidths of up to 10 GHz (King *et al.* 1998, Lear *et al.* 1996). In contrast to traditional edge-emitting laser diodes, the light is reflected vertically through the thin layer of the active region, instead of along it. This creates a very short laser cavity of only a few micrometers between two stacks of Bragg mirrors, and only one longitudinal mode can be resonant with the gain medium (Jung *et al.* 1997). Nevertheless, many transverse modes could be resonating simultaneously (Chang-Hasnain *et al.* 1991, Degen *et al.* 2001). In oxide-confined VCSELs the physical restriction through a small aperture of about 3 µm diameter can limit oscillations to a single transverse TEM_{00} mode (Grabherr *et al.* 1997), reducing the output power of these single-mode devices to 5 mW (Jung *et al.* 1997, Seurin *et al.* 2002). Due to their vertical design, the beam profile is circular with a divergence angle of around 15°–25°. At the same time this design can allow two orthogonal polarization modes to be resonant. They can differ in frequency by many gigahertz and the polarization can flip back and forth between the two modes. Several approaches have been proposed to address this problem (Chavez-Pirson *et al.* 1993, Choquette and Leibenguth 1994, Debernardi and Gian Paolo 2003, Mukaihara *et al.* 1993, 1995a, Numai *et al.* 1995, Ostermann *et al.* 2005, Shimuzi *et al.* 1991). Nevertheless, these polarization jumps can still occur in single-mode VCSELs, and mode competition can cause amplitude noise in the single polarizations as well as in the total light power (Kaiser *et al.* 2002, Mukaihara *et al.* 1995b).

The vertical design makes it possible to fabricate large arrays of VCSELs, thus reducing the production costs (**Figure 10**). Their short cavities require Bragg mirrors of high reflectivity ($R > 99$ %), which reduce the sensitivity to optical feedback. Very low threshold currents below 1 mA (Choquette *et al.* 1994, Huffaker *et al.* 1994, Yang *et al.* 1995) have been reported, allowing small power consumption. At the same time, VCSELs have a large tuning of their output frequency with temperature ($\sim$30 GHz K^{-1}) and injection current ($\sim$300 GHz mA), which causes some problems for clock applications and places constraints on the current source and servo loops.

Even though laser linewidths of 3 MHz have been reported (di Sopra *et al.* 1999), most commercially available VCSELs have linewidths of around 50 MHz (Schmidt *et al.* 1996). While this can be a problem for many laser spectroscopy applications (Affolderbach *et al.* 2000), it does not substantially affect the CPT resonances. As mentioned in Section 3.18.1.3, the linewidth of the CPT resonance depends

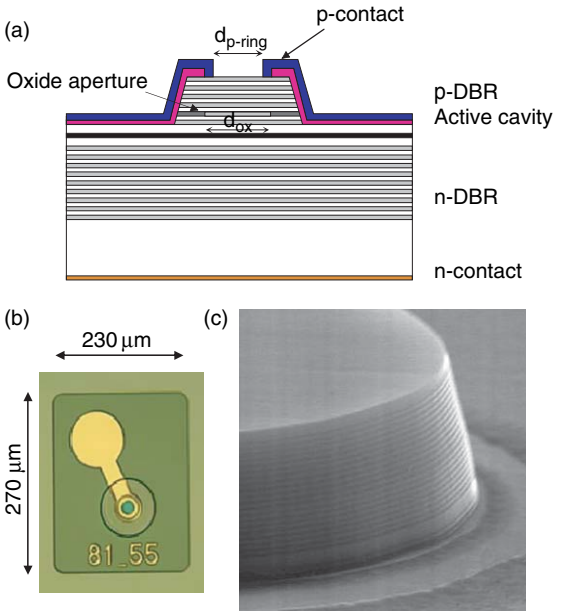


Figure 10 (a) Schematics and (b) photograph of a vertical-cavity surface-emitting laser (VCSEL). (c) Scanning electron microscope (SEM) picture of a VCSEL mesa after etching (before deposition of dielectrics and metal), the mesa height is about 3 µm. (Source: Avalon Photonics, Avalon Photonics Datacom VCSELs, http://www.avap.ch; reproduced with permission from Avalon Photonics Ltd.)

on the relative phase stability between the light fields, that is, the microwave phase, and not on the stability of the individual light fields (Dalton and Knight 1982). Nevertheless, a large VCSEL linewidth can reduce the CPT clock performance because of the large FM noise (Kitching *et al.* 2001c). The FM to amplitude modulation (AM) conversion process (Camparo 1998a) that occurs in the atoms adversely affects the CPT signal.

VCSELs have been developed for telecom applications since the 1980s (Fumio *et al.* 1989, Jewell *et al.* 1989), and reliable single-mode VCSELs have become available at the D_2 lines of Cs (852 nm). Lifetimes beyond one million hours have been reported (Avalon Photonics, Honeywell, Ulm Photonics, Mukoyama *et al.* 2006). Recently, VCSELs at the D_1 lines of Cs (895 nm) and Rb (795 nm), and D_2 lines of Rb (780 nm) and potassium (766 nm) have been produced.

Since their structure consists of more than 100 layers and their fabrication processes and designs vary from one manufacturer to another, the characteristics of single-mode VCSELs vary and directly impact the clock performance (see, e.g., Kwakernaak *et al.* 2004). It is therefore critical to choose the right VCSEL for the specific CSAC design. In order to get high yields, MEMS clocks must conform to stringent requirements regarding epitaxial growth accuracy and uniformity, which is much more demanding than is typically achieved in the best semiconductor fabrication facilities (Serkland *et al.* 2006). Some of the specifications are summarized here:

- Tunable to precise atomic wavelength (preferably 795 or 894 nm) at a given operating temperature, e.g., 75°C, ±5°C (Serkland *et al.* 2006).
- Single longitudinal, transverse, and polarization mode.
- Phase modulation index, >1.8 reachable with radio frequency (RF) powers of −6 dBm at 3.4 GHz (Brannon *et al.* 2005) or 0 dBm at 4.6 GHz (Serkland *et al.* 2006).
- Linewidth, <100 MHz (Serkland *et al.* 2006).
- Power consumption, <3 mW.
- Optical output power, 50 µW–1 mW.

3.18.2.2.3 Vapor cells

The vapor cell can be viewed as the heart of the clock, as it contains the atoms. Besides being a container that does not react with the atoms, the cell has to provide access to the light to and from the atoms. It needs to withstand elevated temperatures, should not be magnetic, and has to be hermetically sealed. For chip-scale vapor cells it should furthermore allow for wafer-level fabrication and filling, since fabrication cost as well as uniformity over an ensemble of cells is of concern. The interior size of the cell should be in millimeters and the exterior volume should be as small as possible. Finally, the geometry of the cell needs to be considered, when integrated with the rest of the physics package.

Several millimeter-sized glass vapor cells have been reported using conventional glass-blowing techniques (Balabas *et al.* 2006) or a focused CO_2 laser (Knappe *et al.* 2003a) (**Figure 11**). Glass alkali cells have been well studied for use in optically pumped clocks. This can make the wall effects due to higher surface-to-volume ratio and thinner walls more predictable. However, the high surface tension of glass, as well as the possibility of the alkali atoms reacting with the glass at high temperatures, makes size reduction to 1 mm challenging.

In an attempt to fulfill the above requirements for a miniature alkali vapor cell, a first fabrication process based on MEMS silicon wafers etching techniques was proposed (Kitching *et al.* 2002). At present, most fabrication techniques for microfabricated cells rely on these ideas, while the actual techniques for filling the

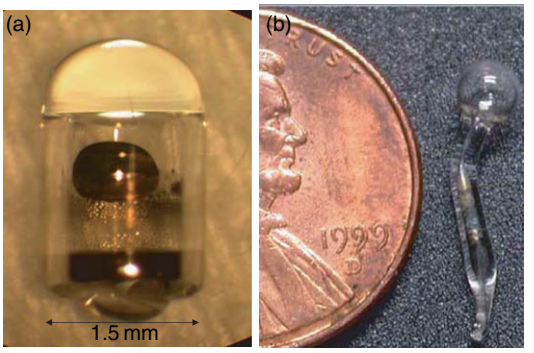


Figure 11 Photographs of small vapor cells: (a) made from a hollow-core glass fiber with a CO_2 laser; (b) made with conventional glass-blowing techniques.

cells with alkali atoms differ significantly. The first MEMS-fabricated vapor cells (Knappe *et al.* 2003b, Liew *et al.* 2004) used double-side polished ⟨100⟩ silicon wafers that were photolithographically patterned inside a clean room (**Figure 12**) and subsequently etched by means of silicon etching techniques (potassium hydroxide (KOH) or deep reactive ion etching (DRIE)) to produce arrays of holes. This allows for a very high level of miniaturization as well as regular arrays of cells with tight tolerances. Cells with holes of 1 mm in size were also successfully made by diamond drilling (Liew *et al.* 2004) or ultrasonic drilling. While surface roughness of the interior walls could be of concern for cells with antirelaxation wall coatings (Alexandrov *et al.* 2002, Bouchiat and Brossel 1966, Robinson *et al.* 1958), no evidence of effects due to rough silicon walls have been published so far for buffer gas cells.

Two early approaches for filling and sealing microcavities with alkali atoms have been demonstrated by NIST (Liew *et al.* 2004). Both used anodic bonding (Wallis and Pomerantz 1969) of Corning 7740 glass (Pyrex™)[1] or borosilicate glass to the etched silicon chips to create a cell preform. The cell was then sealed under a buffer gas atmosphere by anodically bonding a second piece of Pyrex over the top of the preform. In the first method, cesium was injected into the preform with a micropipette in an anaerobic chamber (**Figure 13**). While this is a very simple and inexpensive way to make cells, the chamber environment was difficult to control precisely. First, residual oxygen in the chamber limited the lifetime of cesium and produced large amounts of oxidized cesium inside the cell. Second, other background gases were present inside the chamber and thus inside the cell, which can cause

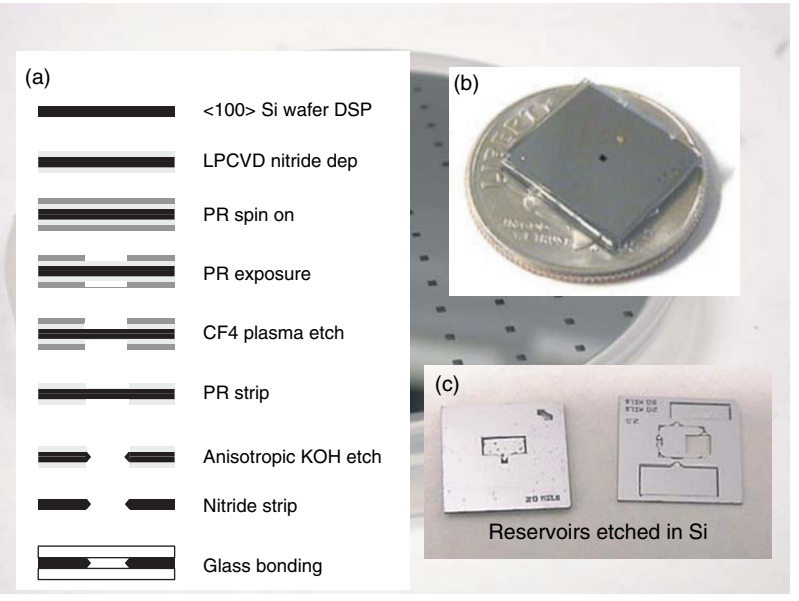


Figure 12 Microelectromechanical systems (MEMS) etching process of microfabricated vapor cells. (a) Steps of a typical potassium hydroxide (KOH) etching process. (b) Photograph of a wafer chip with etched hole and glass bonded onto both sides of the chip. (c) Photograph of chips with channels and reservoirs etched into them in addition to a hole. *Background*: Photograph of a 4″ wafer with an array of holes.

[1]Trade name is stated for technical clarity and does not imply endorsement by NIST. Products from other manufacturers may perform as well or better.

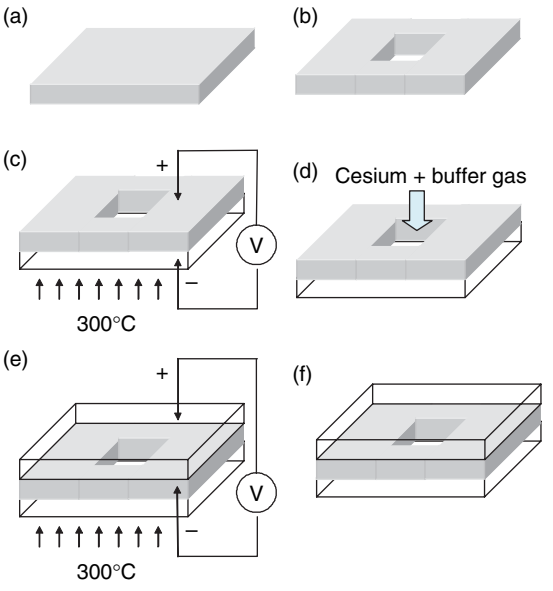


Figure 13 Schematic of cell fabrication based on silicon etching and anodic bonding. A Si wafer (a) with holes etched through it (b) is anodically bonded to a piece of glass (c). It is then filled with alkali atoms and buffer gas (d), and a second piece of glass is bonded on top (e) to form a robust, hermetically sealed vapor cell (f).

problems when trying to reduce the temperature coefficient of the clock by using controlled mixtures of different buffer gases (Vanier *et al.* 1982) (see Section 3.18.4.3 for a more detailed discussion). Very similar filling techniques have been reported subsequently by Kwakernaak *et al.* (2004) and Lutwak *et al.* (2005).

The second filling technique was successfully used in centimeter-sized buffer gas vapor cells before. A stable cesium compound was deposited into the cell and was reacted to produce pure cesium metal by using the following process (Espe 1966):

$$BaN_6 + CsCl \rightarrow BaCl + 3N_2 + Cs$$

The decomposition was initiated by heating the mixture to roughly 200°C inside a UHV chamber. The problem in a small cell arose from the difficulty of pumping off the excess nitrogen gas formed without having all the cesium metal leave the small cell volume at elevated temperatures. Furthermore, the BaCl and the free barium remained inside the cell where the chemical solution was deposited. In bigger cells, this can be desirable because barium can be used as a getter. However, in small cells that are heated to higher temperatures, the gettering rate of nitrogen by barium can cause pressure changes in the cells that result in frequency changes (Knappe *et al.* 2005a). Since the clock frequency of cesium is extremely sensitive to the nitrogen pressure ($45 \pm 2\,\text{Hz}\,\text{kPa}^{-1}$; Beverini *et al.* 1981), the pressure has to be stable at 2×10^{-3} kPa to support a frequency stability of 1×10^{-11}. When using barium inside small buffer gas cells, nitrogen must be completely avoided and instead atoms such as inert gases can be used. But even then it can be difficult to ensure that the chemical reaction has been completed and that no nitrogen is left inside the cell.

One advantage of chloride reduction through barium is that it is relatively easy and cheap to produce small quantities of isotopically enriched ^{87}Rb. Usually, isotopical purity of ^{87}Rb is preferred so that the light is not absorbed by the ^{85}Rb atoms (with 62% natural relative abundance), for which the CPT condition is not fulfilled. In order not to have any residues inside the cell cavity, the alkali atoms can be produced this way inside a glass ampoule and evaporated into the cell cavity through a small nozzle (Knappe *et al.* 2005a). The cell is then closed after filling the chamber with the desired buffer gas mixture. This is a versatile technique, and the cells with Cs, ^{87}Rb, and ^{85}Rb and a variety of buffer gases with pressures <1 torr to 1 atm have been produced this way.

Other approaches to insert the alkali atoms into MEMS cell cavities have also been reported on since then. Radhakrishnan and Lal (2005) encapsulated rubidium in wax micropackets as shown in **Figure 14**. The cell preforms were made by anodic bonding of bulk-micromachined silicon wafers to Pyrex under a xenon atmosphere. One side of the cavity array has a silicon

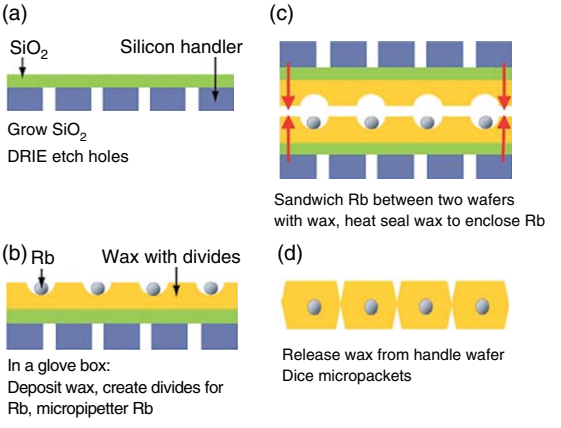


Figure 14 Process to enclose Rb in wax micropackets. (Source: Radhakrishnan S, Lal A 2005 Alkali metal-wax micropackets for chip-scale atomic clocks. *Dig. Tech. Papers – Int. Conf. Solid State Sensors and Actuators and Microsystems, Transducers '05*, pp. 23–6, © 2005; reproduced with permission from IEEE.)

nitride membrane that gets thermally bonded to the wax micropackets, as shown in **Figure 15**. The rubidium is released into the cavity by laser ablation of the silicon nitride membrane. Rubidium absorption was measured in these cells successfully, but the effect of the wax on the ground-state hyperfine coherence has not been studied yet.

A wafer-level process to make arrays of nitrogen buffer gas cells with cesium has been developed by Liew *et al.* (2006a, b). Here, cesium azide (CsN_3) is evaporated through an aluminum shadow mask into cell preforms made from silicon with borosilicate glass windows. The cavities are closed under vacuum by anodically bonding a wafer of borosilicate glass onto the top. The cesium azide is decomposed into cesium and nitrogen by exposing the cells to ultraviolet light. The decomposition is stopped when the nitrogen pressure reaches the desired value.

A technique demonstrated by Gong *et al.* (2006) uses cesium-enriched glass as a source for alkali atoms in microfabricated cells. The glass pieces are melted into a well on the bottom of an anodically bonded cell preform (**Figure 16**). The cell is closed by anodic bonding under the desired buffer gas atmosphere. Finally, the cesium atoms are released from the glass by passing an electrolytic current through the glass at 500°C and 1 kV between a $NaNO_3$ anode and the silicon (cathode).

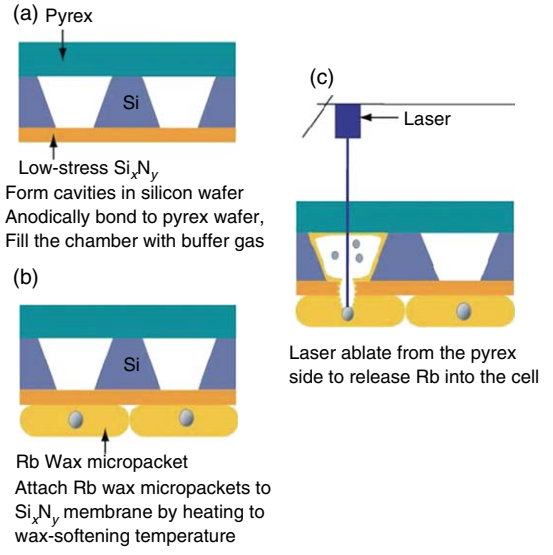


Figure 15 Process outline to realize Rb vapor cells using Rb wax micropackets. (Source: Radhakrishnan S, Lal A 2005 Alkali metal-wax micropackets for chip-scale atomic clocks. *Dig. Tech. Papers – Int. Conf. Solid State Sensors and Actuators and Microsystems, Transducers '05*, pp. 23–6, © 2005; reproduced with permission from IEEE.)

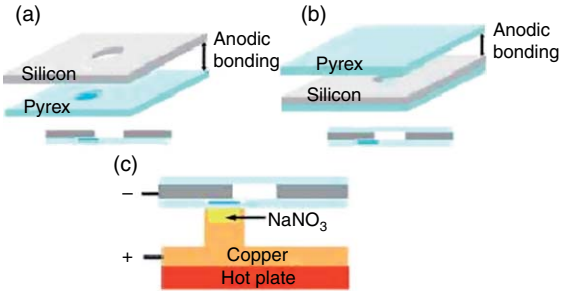


Figure 16 Schematic of the electrolytic cell filling process. (a) A Si wafer with a hole of 2.5 mm diameter is anodically bonded to a Pyrex wafer with a shallow well, 2.5 mm in diameter. Pieces of Cs-enriched glass (dark blue) are melted into the well. (b) Another Pyrex wafer is anodically bonded to the other free surface of the silicon wafer under an argon cover gas. (c) A copper stem, containing molten $NaNO_3$ salt in a basin at the top, is pressed against the Pyrex below the Cs glass well. The stem is attached to the copper base, resting on a hot plate at a temperature of 500°C. A potential of +700 V is applied for a few minutes between the molten $NaNO_3$ anode and the silicon cathode. (Source: Gong F, Jau Y Y, Jensen K, Happer W 2006 Electrolytic fabrication of atomic clock cells. *Rev. Sci. Instrum.* 77, 076101, © 2006; reproduced with permission from American Institute of Physics.)

A MEMS glass-blowing technique has been reported by Erklund and Shkel (2006). As the schematic in **Figure 17(a)** explains, cavities are etched into one side of a silicon wafer using DRIE. A 100-μm-thick piece of borosilicate glass is then anodically bonded over the top of the cavities enclosing the atmospheric pressure. When the wafer is heated to 850°C, that is, above the melting point of the glass, the air pressure inside the cavities deforms the glass into spherical shapes, as shown in **Figure 17(b)**. Afterward, the backside of the cavities is etched to allow for filling with alkali atoms. ^{87}Rb cells have been made this way by using the evaporative filling method described previously and by anodically bonding a second piece of borosilicate glass over the backside in the presence of a nitrogen and xenon atmosphere.

3.18.2.2.4 Optics

Even though the actual designs of the physics packages vary significantly, in almost all cases, there is an optics package that controls the size, intensity, and polarization of the laser beam. Most approaches use circular polarization by passing the beam through a low-order quarter waveplate. Neutral density (ND) filters are a simple way to attenuate the light to the desired power of around 10 μW. At the same time, these attenuators can reduce the optical feedback

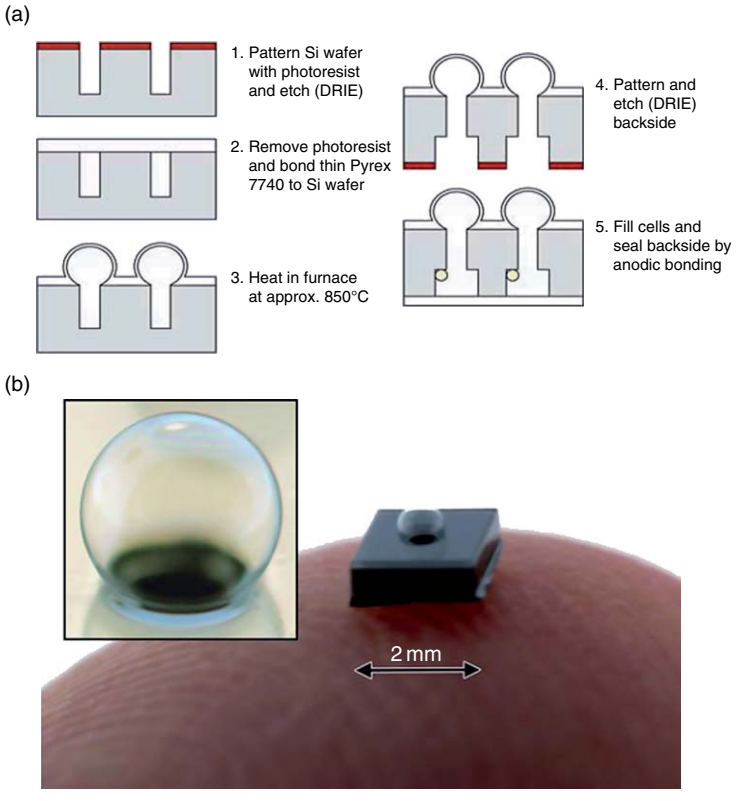


Figure 17 (a) Wafer-level fabrication of glass-blown vapor cells. (b) Photograph of a micro glass cell on top of a silicon wafer. (Courtesy of J. Erklund).

produced by light that gets reflected back into the VCSEL. Some optics packages also use antireflection-coated elements. In some setups, a polarizer is inserted to filter out the undesired polarization modes (Gerginov 2006). While in many designs the laser beam is collimated by a microlens, it has been shown that this is not necessary and a diverging beam can be used instead (Lutwak *et al.* 2004). **Figure 18** shows an example of an optics package consisting of a spacer, a ND filter, an inkjet-printed epoxy lens, and a quarter waveplate through which the beam passes. Other designs use more complicated beam paths by implementing micromirrors in the optics package.

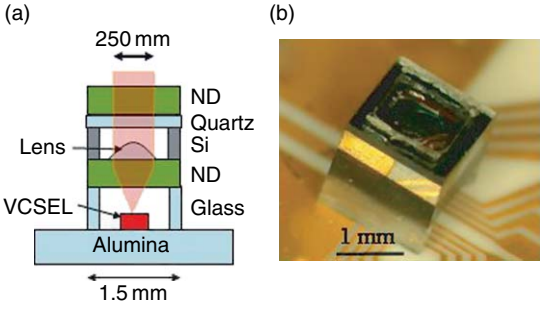


Figure 18 Schematic (a) and photo of a chip-scale atomic clock (CSAC) optics package.

3.18.2.2.5 Heating

In order to achieve sufficient absorption of the laser light over a 1 mm path length, the alkali vapor pressure has to be increased from its room temperature. Because most cells contain a liquid or a solid drop of alkali metal along with the vapor, this is usually done by simply heating the cell. The temperature for 50% absorption of the resonant light depends also on the buffer gas pressure, since buffer gas collisions change the lifetime of the P states (Allard and Kielkopf 1982). For higher pressures and low laser powers, the alkali density has to be increased proportional to the increase in buffer gas pressure to maintain the same absorption. For a 20-kPa nitrogen cell, for example, this temperature is around 75°C for cesium and 90°C for rubidium. Furthermore, it is advantageous to maintain the cell windows at a temperature slightly higher than the body of the cell to minimize the quantity of the alkali atoms that condense on the windows.

Microfabricated heaters are not a new technology (Lai *et al.* 1995, Nguyen and Howe 1993, Ruther *et al.* 2004) and approaches differ significantly. To reduce the magnetic fields produced by the currents flowing through the resistive heaters, patterned traces are often meandered and retraced and resistances are chosen such that only small currents flow. Traces can also be made from transparent materials or patterned such that the laser beam can pass through the center of the cell for better uniformity of heat distribution (Lutwak *et al.* 2004). Furthermore, thin substrate materials with low thermal conductivities are used for better thermal insulation. Substrates, heaters, and temperature sensor materials are chosen to be least magnetic, e.g., platinum, glass, polyimide, indium tin oxide (ITO).

Figure 19 shows an example of microfabricated transparent heaters (Schwindt *et al.* 2006) made from ITO. The heaters are glued directly onto the cell windows, and heat is produced when current flows through the conductive ITO layers. Two glass substrates with patterned ITO are glued together with a nonconductive epoxy to reduce the magnetic fields. The lighter color lines in the heater are where ITO has been removed through laser patterning. After the ITO patterning, gold pads are deposited for wire bonding, and a 2-μm insulating layer of benzocyclobutene is deposited over the remaining exposed ITO so that the electrical contact between the upper and the lower ITO layers is made only through the gold pads. The large arrows show where current enters and exits the heater. The electrically conductive epoxy provides the connection between the upper and the lower ITO. **Figure 20** shows a vapor cell with integrated heaters made from boron-doped silicon. The doped bifilar traces are bulk-micromachined into the silicon cell preform with a high aspect ratio (Liew *et al.* 2005, 2006b).

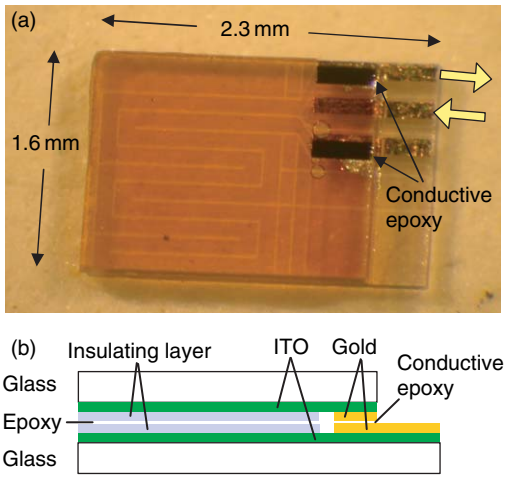


Figure 19 (a) Photograph and (b) schematic side view of laser-patterned indium tin oxide (ITO) heater. (Source: Schwindt P D D, Lindseth B, Knappe S, Shah V, Kitching J 2006 A chip-scale atomic magnetometer with improved sensitivity using the Mx technique. *Appl. Phys. Lett.*)

3.18.2.2.6 Magnetic field control

Most CSACs are operated on the $m_F = 0 \leftrightarrow m_F = 0$ hyperfine transition, to be insensitive to magnetic fields in first order. The magnetic field shifts the neighboring Zeeman sublevels apart with $\sim 7\,\text{kHz}\,\mu\text{T}^{-1}$ for ^{78}Rb and $\sim 3.5\,\text{kHz}\,\mu\text{T}^{-1}$ for Cs. Therefore, a field of a few microtesla in strength is sufficient to separate the neighboring CPT resonances of a few kilohertz linewidth. In order to couple the two 0–0 hyperfine components with circularly polarized light, a longitudinal magnetic field is required.

Until now, not much has been published about how to precisely control the magnetic field in the vicinity of the tiny vapor cell. Nevertheless, magnetic field strength, homogeneity, and direction can be of concern. The amplitude of the CPT 0–0 resonance is proportional to $\cos^2(\theta)$, where θ is the angle between the

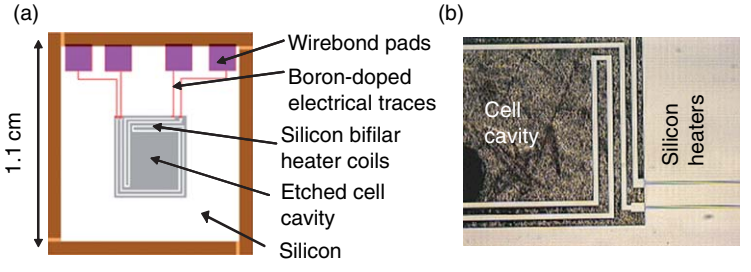


Figure 20 (a) Design of an advanced cell with integrated heaters. (b) Photograph of closeup of the cell cavity. (Source: Liew L, Moreland J, Knappe S, Shah V, Schwindt P, Gerginov V, Kitching J, Hollberg L 2005 Microfabricated alkali atom vapor cells with in-situ heating for atomic-based sensors. *Proc. 3rd Int. Symp. Sensor Science*, Juelich, Germany, pp. 181–3.)

direction of laser propagation and the magnetic field (Wynands and Nagel 1999). Also, because fairly high buffer gas pressures are used in most cells, the atoms can be considered stationary and sample the magnetic field at their location. The CPT resonance is then a convolution of all atoms interrogated and therefore broadens with magnetic field gradients across the cell. But again, if the 0–0 transition is used in buffer gas cells, this effect is reduced because the 0–0 resonance shifts with the magnetic field in second order only. It could be further reduced by using antirelaxation wall-coated cells (Robinson *et al.* 1958), where motional narrowing (Bloembergen *et al.* 1948, Kleppner *et al.* 1962) would suppress the gradient-induced broadening. Finally, the CPT 0–0 resonance frequency shifts with the total magnetic field at $57.5\,\mathrm{mHz}\,\mu\mathrm{T}^{-2}$ for cesium and $42.7\,\mathrm{mHz}\,\mu\mathrm{T}^{-2}$ for ^{87}Rb (see, e.g., Steck 1998, 2001, and references therein).

The externally applied offset field should therefore be made as small as possible to reduce frequency shifts and broadening due to ambient field changes (e.g., through heater currents), but big enough to lift the Zeeman degeneracy beyond the CPT linewidth and to ensure that ambient fields are not rotating the direction of the total field too much (e.g., earth's magnetic field). The external offset field is applied through either magnetic coils or permanent magnets. The latter method has the advantage of reducing the power consumption of the CSAC and the residual magnetic fields because the heater supply leads are not present. The disadvantage is that the CSAC frequency cannot be tuned through the C-field (as is most commonly used in Rb vapor clocks). The physics package is usually shielded from external magnetic fields, for example, the earth's magnetic field, by at least one layer of material with high permeability. It has been shown that shielding factors of around 1000 for one layer and greater than 4×10^6 (Hodby 2006) for three-layer shields of several millimeters in size can be reached.

3.18.2.3 Local Oscillator

3.18.2.3.1 Introduction

Most of the CSACs described so far are passive atomic clocks (see Section 3.18.2.1). The clock frequency is produced by a LO, which is locked to the atomic CPT resonance with a loop bandwidth $\mathrm{BW_{lock}}$. The clock frequency is determined by the atomic ground-state hyperfine frequency, i.e., 6.834 GHz for ^{87}Rb or 9.192 GHz for ^{133}Cs. Using CPT interrogation, a modulation of the VCSEL at half these frequencies is possible. Many applications, however, require a clock with an output frequency of 10 MHz, and frequencies up to 100 MHz can probably be used. Therefore, either the LO frequency has to be divided for the output or a low-frequency LO (10 MHz, for example) should be multiplied for VCSEL modulation. For reaching the goal of 30 mW total power consumption for the CSAC, the output frequency should most likely be the specific gigahertz frequency, determined by the atoms. In the following, we will mostly focus on LOs at microwave frequencies, which are developed for the chip-scale clocks. The list presented here is not complete but is rather a condensed overview.

As a passive standard, the CSAC frequency stability at times shorter than $(\mathrm{BW_{lock}})^{-1}$ is determined mostly by the phase noise of the LO. To reach a clock frequency stability of 1×10^{-11} at 1 h of integration, the single-sideband phase noise power spectral density of the LO needs to be between $-25\,\mathrm{dBc\,Hz^{-1}}$ at 100 Hz offset (Brannon *et al.* 2005, Kitching 2003) and $-43\,\mathrm{dBc\,Hz^{-1}}$ at 300 Hz offset (Römisch and Lutwak 2006) from the carrier. The required output power depends largely on the modulation efficiency of and the coupling efficiency to the VCSEL. It has been shown that for some VCSELs <-6 dBm output power at 3.4 GHz can be sufficient to transfer 60% of the light power into the desired first-order sidebands (Brannon *et al.* 2006b). At 4.6 GHz, the requirement was set to 0 dBm (Römisch and Lutwak 2006). In order to reach the goal of total power of 30 mW for the CSAC, it would be desirable if the DC power consumption of the LO was well below 10 mW and the footprint below $1\,\mathrm{cm^2}$.

A significant concern is the thermal frequency drift of the voltage-controlled oscillator (VCO). Since it is most likely not actively temperature-stabilized, its resonance frequency must not drift out of the tuning range provided by the servo when the ambient temperature varies over the entire operating range of the clock. A maximal thermal frequency shift of around $\pm 10\,\mathrm{ppm\,K^{-1}}$was therefore a goal with a tuning range of a few megahertz. At the same time, the tuning range of the VCO should be small for good resolution and frequency stability, but should be large enough to compensate for manufacturing tolerances. Finally, a supply voltage below 3.5 V is desired and all the components should be of low cost.

The best available oscillators at gigahertz frequencies are dielectric resonator oscillators (DROs) with Q factors of several thousand (Pozar 2001) and even 100 000 for cryogenically cooled sapphire devices (Ivanov and Tobar 2006, Ivanov *et al.* 1998). But their size, weight, and power make them unsuitable

for MEMS clocks. Crystal oscillators can be very good with similar Qs in the frequency range of up to hundreds of megahertz (Vig 1992). Oven-controlled devices (OCXOs) have very good stability, but require several hundred milliwatts of power, but TCXOs or microprocessor-controlled (MCXO) crystal oscillators can operate on a few milliwatts. Nevertheless, multiplying their frequencies into the gigahertz region may require too much power for CSACs. Therefore, considerable research is done on alternative approaches of small, low-power LOs at gigahertz frequencies.

3.18.2.3.2 Oscillator designs for CSAC

Even though the development of VCOs for MEMS clocks has advanced rapidly, not many devices have currently been tested together with CSAC physics packages. One reason is the requirement to operate exactly at the resonance frequency of the atoms.

Popović and coworkers implemented two oscillators at 4.596 GHz and 3.417 MHz (Brannon *et al.* 2005, 2006a) and demonstrated their applicability to a CSAC. The oscillators were based on quarter-wavelength ceramic-filled coaxial resonators. The phase noises of the oscillators were measured to be better than $-35\,\mathrm{dBc\,Hz^{-1}}$ at 100 Hz and $-94\,\mathrm{dBc\,Hz^{-1}}$ at 10 kHz. A weakly coupled varactor diode provides a tuning range of ~3 MHz to compensate for a temperature instability between 0 and $40\,\mathrm{ppm\,K^{-1}}$ over −5°C to +65°C. DC powers of 2.1 and 13.9 mW resulted in −6 dBm output power at 3.4 GHz and −4 dBm at 4.6 GHz. All components were of low cost and were commercially available.

A 3.417-MHz VCO was integrated with a CSAC ^{87}Rb physics package (Brannon *et al.* 2006b) onto a common baseplate. A power of −6 dBm was sufficient to modulate the VCSEL at its desired modulation depth. **Figure 21** is a photograph of the low-power 3.4-GHz oscillator (Brannon *et al.* 2006b) and the CSAC physics package.

Römisch and Lutwak (2006) recently demonstrated a 4.596-GHz oscillator based on a thin-film temperature-compensated resonator (Römisch and Lutwak 2006) using a very similar oscillator design. A single-sideband phase noise power spectral density of $-38\,\mathrm{dBc\,Hz^{-1}}$ at 300 Hz was measured when emitting the required 0 dBm output power, improving at lower powers. With a power consumption below 10 mW and a tunability of $2\,\mathrm{MHz\,V^{-1}}$, this resonator was designed for cesium CSAC.

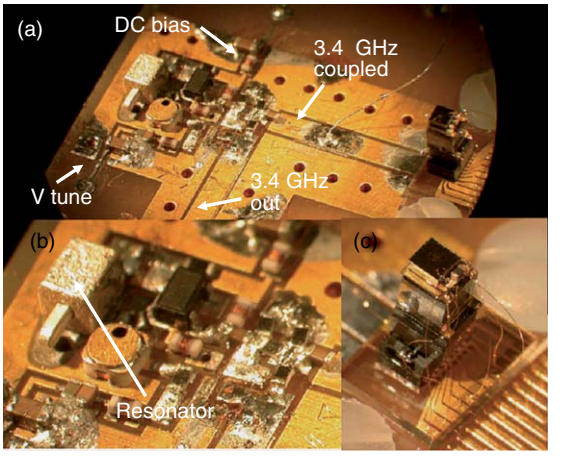


Figure 21 (a) Photograph of a 3.4-GHz voltage-controlled oscillator (VCO) integrated with a chip-scale atomic clock (CSAC) physics package. Inputs are DC bias and tune voltage for the VCO, and laser bias, photodetector bias, and heater currents for the physics package. Outputs are stabilized at 3.4 GHz and diagnostic signals. (b) Closeup view of the VCO and (c) the physics package.

3.18.2.3.3 Other MEMS resonators

The oscillators described so far were able to reach the goals set for the CSAC in terms of frequency stability, power consumption, and size. Nevertheless, fully integrated MEMS oscillators can be advantageous for mass fabrication and possibly advanced performance. A detailed noise analysis of MEMS oscillators was performed (Vig and Yoonkee 1999). In the following, we introduce a few approaches using other types of MEMS resonators.

Totally integrated complementary metal oxide semiconductor (CMOS)-type oscillators are probably the smallest devices, but do not presently fulfill the phase noise requirements (Gierkink *et al.* 2003, Mourant *et al.* 2000). Much progress has been made recently in the development of micromachined mechanical resonators (Nguyen 2005). Unloaded Qs greater than 10 000 have been demonstrated up to 1.5 GHz (Li *et al.* 2004) and temperature stabilities below 1.8×10^{-5} over 25–125°C (Hsu and Nguyen 2002). An advantage of these micromechanical resonators based on polysilicon (Li *et al.* 2004) or diamond (Wang *et al.* 2002) surface micromachining is that their resonance frequency can be defined by their lateral dimension, rather than by their thickness, which allows for easy specification through computer-aided design (CAD). Furthermore, arrays of mechanically coupled 15-MHz resonators have demonstrated an improvement in phase noise of 40 dB by locking all resonator frequencies to one

mode (Lee and Nguyen 2004). Nevertheless, no oscillator using one of these micromachined resonator tank has been demonstrated in the gigahertz region as at the time of this writing (2006).

MEMS techniques lend themselves to the making of gigahertz oscillators for CSACs, as small size allows for faster speed, lower power consumption, and lower cost. In the case of quartz resonators, MEMS fabrication enabled quartz on-chip resonators at 900 MHz with *Q*s greater than 10 000 and strong third overtone modes above 2.5 GHz and *Q*s greater than 7500 (Kubena *et al.* 2005, Stratton *et al.* 2004). Temperature coefficients around $-5\,\text{ppm}\,\text{K}^{-1}$ have been measured in shear strip designs. The fabrication process and a scanning electron microscope (SEM) photograph of a complete ultrahigh frequency (UHF) quartz resonator are shown in **Figure 22** as an example.

Other interesting candidates include thin-film bulk acoustic resonators (FBAR) and high-tone bulk acoustic resonators (HBAR). Frequency-tunable FBAR oscillators at 2 GHz showed a phase noise of $-112\,\text{dBc}\,\text{Hz}^{-1}$ at 10 kHz and an unloaded *Q* better than 500 (Khanna *et al.* 2003). With a supply voltage of 3.3 V, it uses 115 mW of power and had a tuning

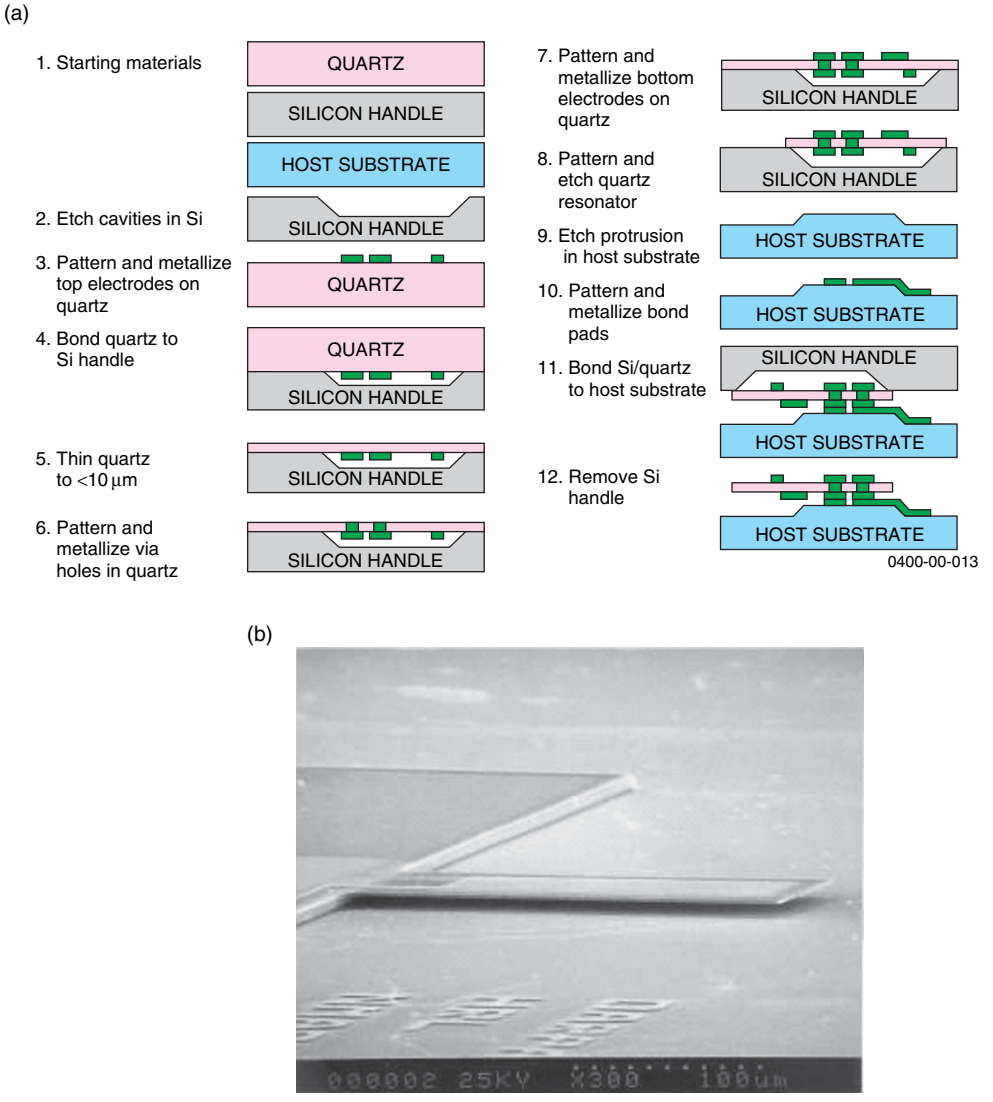


Figure 22 (a) Fabric flow for a microelectromechanical systems (MEMS)-based quartz resonator. (b) Scanning electron microscope (SEM) photograph of a completed ultrahigh frequency (UHF) resonator with 2-mm quartz thickness. (Source: Kubena R L, Stratton F P, Chang D T, Joyce R J, Hsu T Y, Lim M K, M'closkey R T 2005 MEMS-based quartz oscillators and filters for on-chip integration. *Proceedings of the 2005 IEEE International Frequency Control Symposium and Exposition*, pp. 122–27, © 2005; reproduced with permission from IEEE.)

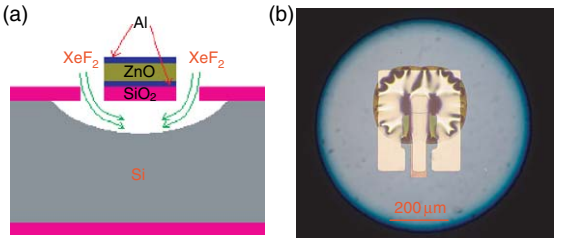


Figure 23 (a) Cross-section schematic of an Al/ZnO/Al/SiO_2 film bulk acoustic resonator (FBAR) and (b) photograph of the fabricated FBAR. (Source: Zhang H, Kim J, Pang W, Yu H, Kim E S 2005 5 GHz low-phase-noise oscillator based on FBAR with low TCF. *The 13th Int. Conf. Solid-State Sensors, Actuators and Microsystems, Dig. Tech. Papers (Transducers '05)*, pp. 1100–1, © 2005; reproduced with permission from IEEE.)

sensitivity of 180–370 kHz V^{-1} (0–10 V tuning range). FBAR resonators generally have temperature stabilities of 20–30 ppm K^{-1}. Pang *et al.* (2005) have reduced the temperature coefficient of a tunable FBAR resonator at 2.8 GHz by 40 ppm K^{-1} by integrating a micromachined air-gap capacitor and Zhang *et al.* (2005) demonstrated a temperature stability of −8.7 ppm K^{-1} with a 5-GHz CMOS-compatible device. Care has to be taken in the design of FBAR oscillators to satisfy the resonance condition for the desired mode only, as they can have closely spaced series and parallel resonances. A schematic and a photograph of a 5-GHz FBAR oscillator are shown in **Figure 23**.

Resonators can be scaled even smaller with nanoelectromechanical systems (NEMS) to allow for Q factors in the range of 1000–100 000 with very low power consumption. Double-clamped beams made from SiC reach microwave frequencies (Ekinci and Roukes 2005).

3.18.2.4 Control Electronics

Besides the physics package and the LO, the CSAC needs miniaturized control electronics for operating the clock by providing the supply currents and voltages for the individual components and for controlling at least four servo loops for laser and cell temperature, laser frequency, and LO frequency. Sometimes, laser and cell are maintained at the same temperature and the two servos can be combined.

To reach the goal of 1 cm^3 total volume for the CSAC, an application-specific integrated circuit (ASIC) will most likely be required. Until now, prototypes of 10 cm^3 have been demonstrated (Gerginov *et al.* 2005, Lutwak *et al.* 2005) with control electronics

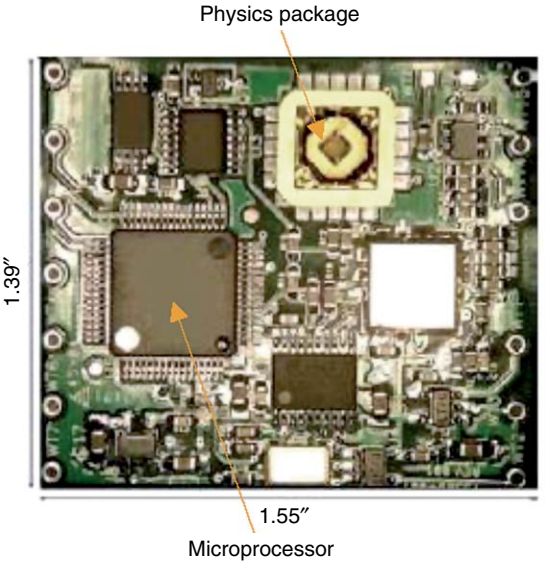


Figure 24 Prototype printed circuit (PC) board for a miniature atomic clock. (Source: Lutwak R, Vlitas P, Varghese M, Mescher M, Serkland D K, Peake G M 2005 The MAC – A miniature atomic clock. *Joint Meeting of the IEEE International Frequency Control Symposium and the Precise Time and Time Interval (PTTI) Systems and Applications Meeting*, Vancouver, BC, Canada, pp. 752–7.)

based on digital microprocessors. **Figure 24**, for example, shows a printed circuit (PC) board of footprint 14 cm^2 containing all the control electronics necessary to run a CSAC. As shown in the block diagram in **Figure 25**, the analog signals from the CSAC are converted into digital signals and processed by the microprocessor. The output is converted into analog signals again, which in turn drive the CSAC.

The temperatures are usually measured by thermistor temperature sensors. A proportional integrator (PI) servo is programmed into the microprocessor chip and the single-sided (heating only) output controls the heater currents. For low power consumption, the cell and the laser are operated above the maximum ambient temperatures, so that no cooling is required.

Since the laser and the LO have to be locked to the center of the optical and microwave resonance lines, respectively, phase-sensitive detection is implemented for these two servos. The laser current and the LO tuning voltage are modulated at a low frequency (500 Hz–30 kHz), and the photodetector voltage is demodulated at this frequency. To prevent cross-talk between the two loops, the two modulation frequencies are usually chosen to be different by an order of magnitude and the demodulation filters have to be tuned accordingly. The phase of the

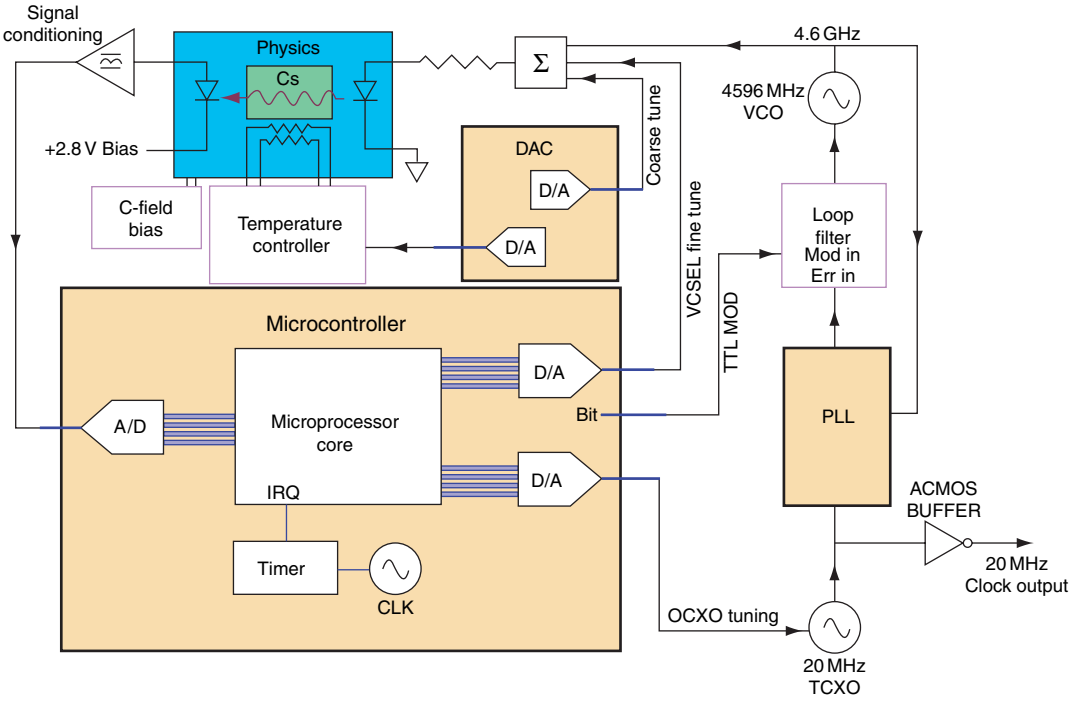


Figure 25 Block diagram of a miniature atomic clock. (Source: Lutwak R, Vlitas P, Varghese M, Mescher M, Serkland D K, Peake G M 2005 The MAC – A miniature atomic clock. *Joint Meeting of the IEEE International Frequency Control Symposium and the Precise Time and Time Interval (PTTI) Systems and Applications Meeting*, Vancouver, BC, Canada, pp. 752–7.)

demodulation has to be chosen precisely for best performance. Due to the high tunability of VCO ($\sim$2 MHz V^{-1}) and VCSEL ($\sim$300 GHz mA^{-1}), highly stable supply voltage and current are needed. In order to run the CSAC off a common voltage supply, for example, two AA batteries (3 V), heater resistances and current supplies have to be designed for minimum power consumption at this voltage.

3.18.2.5 Packaging

When power consumption is the concern, packaging, especially for the physics package, becomes a critical issue. Since laser and cell are heated above the ambient temperature, that is, to at least 75°C for many applications, thermal management and control need to be addressed. Without this control, physics packages will require 100–250 mW just to heat the small vapor cell and tiny laser die (Kitching *et al.* 2004). Most of the heating power is then lost thermally through conduction and convection.

Conduction can be tremendously reduced by increasing the thermal resistance between the physics package and the environment. **Figure 26** shows a CSAC physics package that consumes <10 mW of power at an ambient temperature of 25°C to heat the cell (Lutwak *et al.* 2004). The heat conduction has been reduced to 13 μW K^{-1} by suspending the physics package on a thin web of strained polyimide support tethers. Electrically conductive leads composed of gold and platinum are deposited on top with a titanium adhesive layer to supply the heating current.

Convection is minimized by vacuum packaging. Immersion has been suggested as a lower-cost alternative in some materials of low thermal conductivity. A different approach for a MEMS clock physics package has been proposed by Youngner *et al.* (2005), and also includes a thermal suspension system (**Figure 27**).

One of the major challenges in the thermal management is the requirement to maintain temperature control better than 100 mK over a wide range of ambient temperatures (−40 to +70°C) with a small power budget. This is further complicated by the heat dissipated by the VCSEL (Kitching *et al.* 2004). One solution can be a MEMS thermal conduction switch that can change the thermal resistance between the physics package and the environment (Laws *et al.* 2005). A schematic of a physics package that implements such a thermal switch is shown in **Figure 28**.

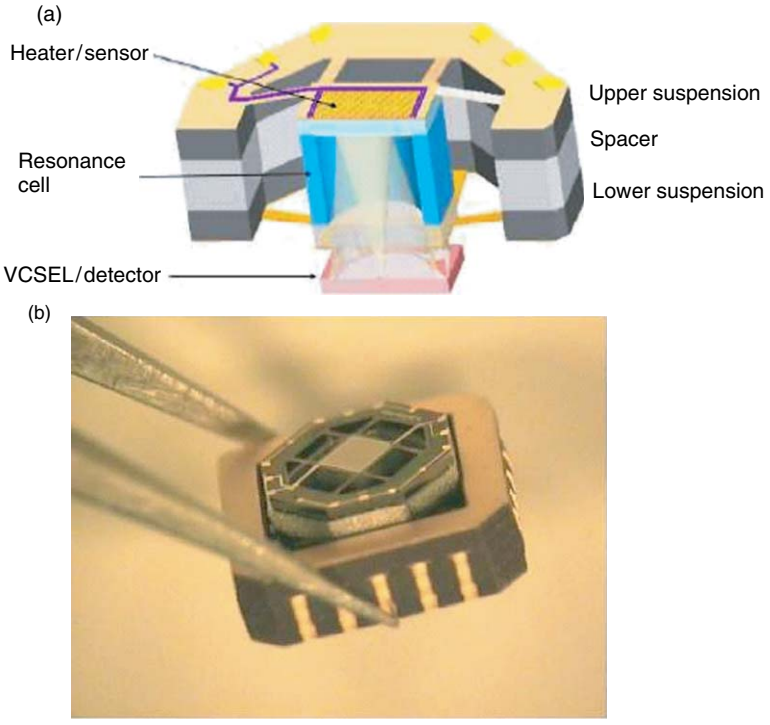


Figure 26 (a) Cutaway view and (b) photograph of a low-power chip-scale atomic clock (CSAC) physics package. (Source: Lutwak R, Deng J, Riley W, Varghese M, Leblanc J, Tepolt G, M.Mescher, Serkland D K, Geib K M, Peake G M 2004 The chip-scale atomic clock – Low-power physics package. *36th Annual Precise Time and Time Interval (PTTI) Meeting*, Washington, DC, USA, pp. 1–14. Honeywell permission.)

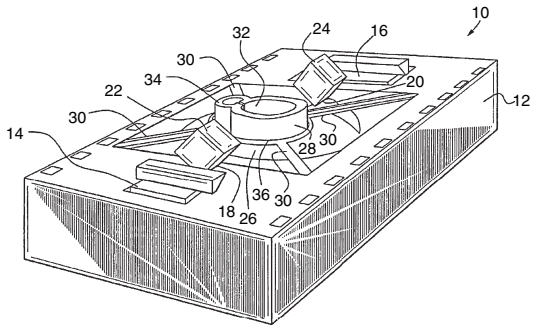


Figure 27 Schematic of a chip-scale atomic clock (CSAC) physics package (10) design including: substrate (12), light source (14), light detector (16), grooves (18, 20), optical processors (22, 24), cavity (26), vapor cell (28, 32, 44), bridge (30), and heater (36). (Source: Youngner D W, Detry J F, Zook J D 2005 MEMS frequency standard for devices such as atomic clocks. *US PTO* 6 900 702 B2.)

3.18.3 Performance

3.18.3.1 Introduction

The long-term goal for most CSAC projects, set by DARPA, is to reach a frequency instability below 10^{-11} at 1 h of integration in a 1 cm^3 package consuming <30 mW of power. Therefore, most CSACs have been characterized mostly by these three specifications: frequency stability at times <1 h, size, and power consumption. Nevertheless, for many applications, other specifications, such as frequency stability at times longer than a week, accuracy, acceleration sensitivity, magnetic field sensitivity, lifetime, no unlocks for long durations, might be at least as important. At the same time, which specifications will be set for the CSAC will strongly depend on the application. In the following sections, we nevertheless focus mostly on the three specifications set by DARPA.

3.18.3.2 Frequency Stability

3.18.3.2.1 Introduction

Atomic clocks are often characterized by their Allan deviation $\sigma_y(\tau)$ (for a review, see Allan 1987), which is a measure of the fractional frequency stability as a function of averaging time τ. Over short averaging times, the stability of many atomic clocks is characterized by white frequency noise. For this type of noise, the Allan deviation decreases as $\tau^{-1/2}$ (**Figure 29**, light gray dots) and can be expressed in

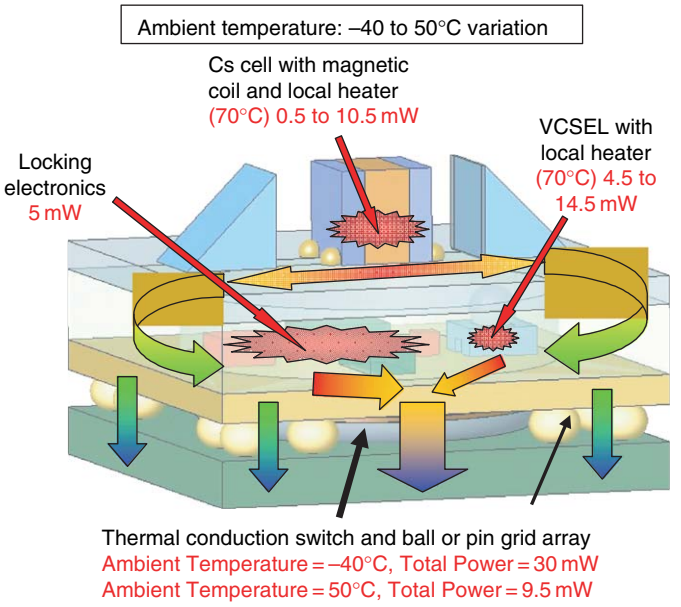


Figure 28 Illustration of heat transfer path through a chip-scale atomic clock (CSAC) design using a thermal conduction switch to reduce overall power consumption. Heat rates are estimated to meet clock power requirements of <30 mW. Radiation and conduction/convection are ignored. (Courtesy of V. Bright.)

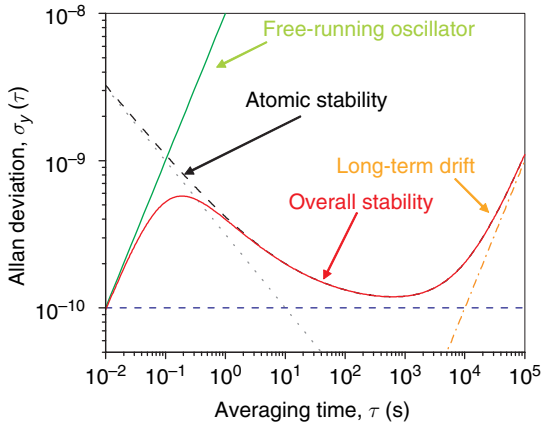


Figure 29 Fractional frequency stability: free-running oscillator (dashed green line), supported by the atoms (dashed black line), when locked to the atomic resonance. The atomic stability here is determined by white frequency noise (light gray dots), flicker noise floor (dashed blue dots), and frequency drift (orange dashes).

terms of signal-to-noise ratio S_{CPT}/N and resonance quality factor Q (Vanier 2001)

$$\sigma_y(\tau) = \frac{\chi}{QS_{\mathrm{CPT}}/N}\frac{1}{\sqrt{\tau}} \qquad [4]$$

Here, $Q = \nu_0/\Delta\nu_{\mathrm{CPT}}$ is defined as the ratio of resonance frequency over linewidth and χ is a parameter of the order of unity related to the method of interrogation. It is desirable to have a high signal-to-noise ratio and a narrow linewidth in order to reach good frequency stability. The fractional frequency stability of these clocks improves over measurement time until it reaches the flicker noise floor, $\sigma_y(\tau) \propto \mathrm{const}$ (**Figure 29**, blue dashes), or until the frequency starts drifting, $\sigma_y(\tau) \propto \tau$ (dashed orange dots). These drifts are usually caused by environmental variations that change the clock frequency over time. A typical Allan deviation determined by the atoms is shown in **Figure 29** (black dashes). In a passive CPT clock, an LO generates a signal close to the clock frequency. The physics package uses this signal as its input to create an output signal that carries information about the frequency difference between the LO frequency and the atomic frequency. This information can then be used to tune the LO frequency so that it always coincides with the one of the atoms (for a review on CPT clocks, see, e.g., Vanier 2005). Thus, the stability of the clock is determined by the LO stability at times shorter than the inverse loop bandwidth, and by the atoms at longer times (**Figure 29**, red line).

For a CSAC, the frequency stability can be predicted in the photon shot-noise limit to be on the order of $\sim 2 \times 10^{-13}/\tau^{1/2}$ for a 1-mm^3 cell and

typical experimental parameters (Vanier *et al.* 2003d). Unfortunately, in an experimental setup, this stability has not been reached, since other noise sources are dominant and the signals are small. For VCSELs, laser amplitude noise and laser frequency noise play a major role. Frequency noise gets converted into amplitude fluctuations through the atomic resonances (Camparo and Buell 1997, Coffer *et al.* 2002). In most CSACs, the VCSEL is locked to the center of a resonance, so that this effect is drastically reduced in comparison to when locked to the side of the resonance (Kitching *et al.* 2001b), but it nevertheless remains significant. Here it can be advantageous to use a high buffer gas pressure where the optical lines are very broad and have a shallow FM–AM conversion slope (Camparo 1998a). Furthermore, in some VCSELs, mode-competition noise (Kaiser *et al.* 2002, Mukaihara *et al.* 1995b) between different polarization modes causes large contributions to the noise on the detector, especially when transformed into amplitude noise by a polarizer. All of these noise sources reduce the frequency stability, so that the best short-term stability reported to date in a CSAC physics package is $4 \times 10^{-11}/\tau^{1/2}$ in a 1-mm^3 cell (Knappe *et al.* 2005b).

Besides the effects that limit the short-term stability of the CSAC, frequency shifts and line shape asymmetries can cause major problems at longer averaging times. They can result from magnetic fields, buffer gas pressure, temperature, light shifts, acceleration, or RF power. It is therefore critical to control these parameters carefully or to find a measurement scheme that reduces the frequency sensitivity to changes in these parameters.

While for many primary frequency standards the absolute clock frequency is critical (Quinn 2005), for most applications of chip-scale clocks this value is of minor interest, as long as it can be calibrated. Nevertheless, it is of great importance that the clock frequency does not change over time, i.e., it needs good stability, but not necessarily good accuracy.

3.18.3.2.2 Short-term frequency stability

The first CSAC physics package demonstrated used the D_2 line of cesium (Knappe *et al.* 2004b). Its short-term frequency stability of $6 \times 10^{-10}/\tau^{1/2}$ was limited by a CPT resonance contrast of 0.9% and a CPT linewidth of 7 kHz. The package had a volume of 9.5 mm^3 and required 75 mW of power at an external baseplate temperature of 46°C. The power budget did not include heating of the laser. A plot of the Allan deviation for this CSAC physics package is shown in **Figure 30** (black squares).

It has been shown for cesium (Lutwak *et al.* 2003) as well as for rubidium (Stähler *et al.* 2002) that the CPT resonance contrast is much higher on the D_1 line than on the D_2 line. The first CSAC physics package using the D_1 line of ^{87}Rb was demonstrated shortly after (Knappe *et al.* 2005b). The short-term frequency instability was $4 \times 10^{-11}/\tau^{1/2}$, the best short-term frequency instability reported so far for this simplest type of CPT CSAC with a 1-mm^3 vapor cell. A plot of the Allan deviation can be found in **Figure 30** (red dots). The improvement can be attributed to a higher CPT resonance contrast of 4.6% as well as to fewer Zeeman sublevels in ^{87}Rb when compared to Cs. Finally, a CSAC physics package using the D_1 line of cesium has also been presented with a short-term frequency instability of $1.4 \times 10^{-10}/\tau^{1/2}$ (Lutwak *et al.* 2005).

When the physics package is integrated with a miniature LO, the performance of the CSAC can change dramatically. In order to reach large tuning ranges, the tunability of the LOs often needs to be around 1 MHz V^{-1}, making them very sensitive to voltage fluctuations on the tuning input. Also, most small low-power LOs have phase noise around −40 dBc at 100 Hz at best, which puts tight constraints on the control system. Nevertheless, short-term stabilities of $2.4 \times 10^{-10}/\tau^{1/2}$ have been demonstrated with an integrated low-power LO physics package design (Brannon

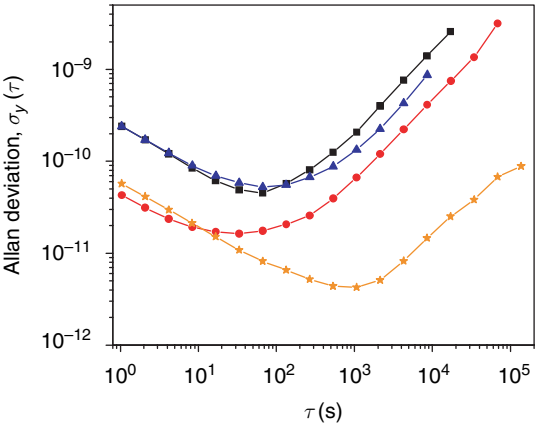


Figure 30 Fractional frequency stability as a function of averaging time for a Cs D_2 chip-scale atomic clock (CSAC) physics package (black squares), ^{87}Rb D_1 CSAC physics package (red dots), CSAC physics package with integrated local oscillator (LO) (blue triangles), and 1-mm^3 ^{87}Rb vapor cell (orange stars).

et al. 2006b) (**Figure 30**, blue triangles). Disciplining the VCO by the use of small control electronics can be challenging, as higher servo bandwidths are desired when compared to oscillators with better phase noise. This is worsened by the fact that several servo loops have to be controlled simultaneously. It has been shown that an integrated CSAC (matchbox size) can nevertheless support a fractional frequency stability of $6 \times 10^{-10}/\tau^{1/2}$ (Lutwak *et al.* 2005).

3.18.3.2.3 Long-term frequency stability

Until now, we have discussed only frequency stability for integration times <100 s. **Figure 30** shows that frequency drifts determine the stability of these CSACs at integration times >100–1000 s. Most drift is directly or indirectly caused by temperature fluctuations. The stabilities presented here were measured in a laboratory environment. It has been shown that under such conditions, MEMS cells can support frequency stabilities below 1×10^{-11} at 1 h and below 1×10^{-10} at one day (Knappe *et al.* 2005a) (see orange stars in **Figure 30**). Nevertheless, it can be expected that in a real-world environment, the frequency stabilities of these CSACs will degrade severely if the shifts are not addressed. In the following, we discuss different parameters that influence the long-term stability.

3.18.3.2.3.(i) Buffer gas pressure

As described in Section 3.18.1.4, a buffer gas is used to narrow the CPT linewidth. At the same time it shifts the hyperfine frequency. This shift is temperature-dependent. A standard way to reduce the temperature coefficient of the clock frequency in commercial optically pumped devices is to use combinations of buffer gases that shift in opposite directions (Arditi and Carver 1961, Vanier *et al.* 1982). The same technique works well for CPT clocks (Knappe *et al.* 2001). However, in small devices, the residual temperature coefficient is proportional to the total buffer gas pressure. Because the pressure in most approaches is nearly an order of magnitude higher than in centimeter-sized devices, coefficients worsen accordingly. Furthermore, the buffer gas combinations are harder to control in microfabricated cells, because the anodic bonding process with its high temperatures can create residual gases inside the cell. Therefore, the cell temperature has to be stabilized precisely. It has been shown, however, that the temperature stability of the cell (as well as the VCSEL) can be improved when direct absorption signals from the atoms are used to determine the temperatures, instead of temperature sensors in the vicinity of the cell (or VCSEL) (Gerginov *et al.* 2006b).

3.18.3.2.3.(ii) Light shifts

Compared to optically pumped clocks, light shifts can be reduced in CPT, because of the ability to create a more symmetric system (Levi *et al.* 2000, Orriols 1979, Vanier *et al.* 1998). Nevertheless, the frequency stability of most CSACs is still limited by light shifts at long averaging periods.

VCSELs have frequency tunabilities of roughly 300 GHz K^{-1} (see Section 3.18.3.2.3). Therefore, it is essential to stabilize their temperatures precisely. This is usually done with small temperature sensors, e.g., thermistors close to the VCSEL die. When temperature gradients between the sensor and the die change the VCSEL junction temperature, for example, due to ambient temperature fluctuations, its output frequency changes. The frequency servo then changes the DC current of VCSEL to pull it back onto resonance, but the output power as well as the VCSEL impedance, that is, modulation index of the light field, is usually different (Gerginov *et al.* 2006b). This in turn generally changes the clock frequency through altered light shifts (Arditi and Carver 1961, Barrat and Cohen-Tannoudji 1961, Mathur *et al.* 1968). It has been shown in a large cell (Levi *et al.* 2000, Vanier *et al.* 1999, Zhu and Cutler 2000, 2001) as well as in a CSAC (Gerginov *et al.* 2006b) that when the right RF modulation index is chosen, the total light shift from the laser carrier and all other sidebands vanishes. A similar technique has been implemented in a microwave clock as well (Affolderbach *et al.* 2006). Furthermore, an additional servo loop can be used to lock the RF power to this point, where no light shifts are present (Shah *et al.* 2006a).

3.18.3.2.3.(iii) Other frequency shifts

The field of MEMS clocks is very young when compared to that of lamp-pumped Rb standards. The use of new materials and components raises many uncertainties onto the performance of these new devices. Careful characterization and engineering over the coming years will be required to turn the current prototypes into reliable off-the shelf devices. In the following, a few of these parameters are summarized.

An important frequency shift is the RF output power of the LO. In an in-field environment where temperatures and coupling to metal objects can change dramatically, the RF power of the VCO is difficult to control. The clock frequency depends critically on the

power distributed over all modulation sidebands through light shifts. The modulation index therefore needs to be stabilized. Two methods have been demonstrated to do this. In a CSAC, the RF power can be locked to the point where the power in the first-order sidebands is maximized (Gerginov *et al.* 2006b). In a table-top CPT setup, the RF power is locked to the point where the total light shift vanishes (Shah *et al.* 2006a). Furthermore, the slow modulation of the LO frequency used for phase-sensitive detection of the CPT resonance can lead to shifts in the clock frequency (Phillips *et al.* 2006).

As mentioned in Section 3.18.2.2.6, the clock frequency is shifted as a function of total magnetic field. At the time of writing (2006), no measurements that characterize the residual shifts that occur after packaging and shielding have been published.

Since MEMS clocks are supposed to work in portable devices, acceleration insensitivity is of importance, and CSACs have not yet been characterized in this way. Nevertheless, two LOs have been specified with 0.2–0.3 ppm g^{-1} at 10 Hz (Brannon *et al.* 2006b) for these clocks.

Furthermore, in optically pumped clocks with their glass-blown vapor cells, diffusion of gases such as helium through the glass walls can cause problems (Bloch *et al.* 2000, Camparo *et al.* 2005). Chip-scale cells have not been characterized in this respect. But thin glass windows and new bonding mechanisms can be of concern. And finally, the reliability of the VCSELs with unknown behaviors in CSACs due to aging, sparks, radiation, etc., is yet to be determined. Nevertheless, initial performance results are encouraging, given the very short development time and the lack of any serious engineering.

3.18.3.3 Power Consumption

The power consumed by the first CSAC physics packages, which had essentially no thermal shielding, was 200 mW (Knappe *et al.* 2004b, 2005b). Nevertheless, their thermal properties have been modeled and characterized (Kitching *et al.* 2004). As discussed in Section 3.18.2.5, the main portion of the power was expended in heating the cell and the VCSEL. Therefore, it can be favorable to use cesium rather than rubidium, because of a higher vapor pressure. It has been convincingly demonstrated that it is possible to reduce heat conduction to below 13 μW K^{-1} in a CSAC physics package and therefore to heat a CSAC cell of volume 8.5 mm^3 to 75°C at an ambient temperature of 25°C with only ~6 mW of power (Lutwak *et al.* 2004). Since Rb has a slightly lower vapor pressure, with the same setup, it would require ~8.5 mW to reach a similar density at 90°C.

As discussed in Section 3.18.2.3, a VCO at 3.417 GHz is required for ^{87}Rb with an output power of at least −6 dBm and at 4.596 GHz for Cs with an output power of ~0 dBm. Two of these have been successfully demonstrated, requiring 2.1 mW (Brannon *et al.* 2006a) and 7.6–10 mW (Römisch and Lutwak 2006) for ^{87}Rb and Cs, respectively.

Control electronics have been developed using microcontrollers, resulting in packages around 10 cm^3 in size (see Section 3.18.2.4). For these, power budgets below 30 mW have been reported (Lutwak *et al.* 2005). For a complete miniature atomic clock package, including three such components, a total power budget of 108 mW was quoted (Lutwak *et al.* 2005). This included a commercial 4.6-GHz VCO. When this is replaced by one of the small low-power VCOs described here, the power budget is expected to be reduced further.

Finally, if an output frequency of 10 MHz is required, the odd output frequency in the gigahertz range has to be converted down. This can be done either by phase-locking a higher-power microwave VCO to a 10-MHz VCO (similar to **Figure 25**) or by adding an external frequency divider. The latter could consist of a fractional-*N* divider locked to a TCXO, for example. Such a device, as depicted in **Figure 31**, has been built on the same footprint as the

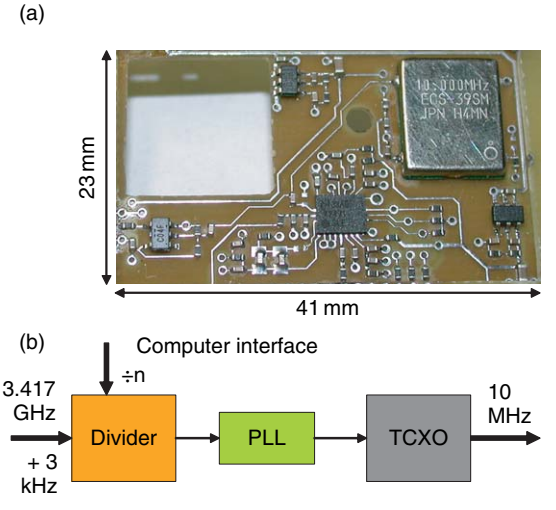


Figure 31 Photograph (a) and schematic (b) of a frequency divider for chip-scale atomic clock (CSAC). The fractional-*n* divider chip is programmed to divide 3.417 GHz down to 10 MHz. The output signal is then phase-locked to a 10-MHz temperature-compensated quartz crystal oscillators (TCXO).

NIST MEMS clock and consumes <30 mW of power (Waltman 2006). The TCXO acts as a cleanup oscillator that filters the modulation sidebands off the signal and reduces the phase noise at higher frequencies.

3.18.3.4 Size

The smallest CSAC physics package so far that consists of VCSEL, optics, cell, and photodiode has been demonstrated by Lutwak *et al.* (2005). It has a volume of 10 mm^3 without packaging. A picture is shown in **Figure 26**. The VCSEL is surrounded by a narrowband photodetector grown on the same substrate. After the diverging beam from the laser passes is attenuated and circularly polarized, it passes through the cesium vapor cell. Then it is retroreflected back through the cell onto the photodetector surrounding the VCSEL. The temperatures of the VCSEL and detector are stabilized together, so that their passbands are tuned simultaneously. When thermal packaging is included, it has a volume of ~1 cm^3. The smallest LO demonstrated to work with a CSAC physics package has a volume of ~0.15 cm^3 (Brannon *et al.* 2006a).

Finally, when integrating the control electronics with physics package and LO, a total volume of 10 cm^3 has been demonstrated (Lutwak *et al.* 2005). While this size is comparable to that of the smallest CPT vapor cell clock based on conventional technology (Vanier *et al.* 2004, 2005), its power is reduced by a factor of 5. With the CSAC technology demonstrated so far it is possible to reduce the volume of the control electronics to below 0.5 cm^3, allowing the CSACs to have volumes of around 1 cm^3.

3.18.4 Advanced Techniques

3.18.4.1 Introduction

Until now, we have mainly discussed the simplest approach to a CSAC: one VCSEL, one beam, one cell, and one photodetector, all in a straight line. The CPT resonance excited in this way couples the $m_F = 0$ ground states to either the $m_F = +1$ or the $m_F = -1$ excited state with circularly polarized light. The short-term performance of the CPT clock can then be characterized by the resonance contrast, defined as the amplitude of the CPT resonance over the background, the resonance width, and the noise. The performance at longer times is determined mostly by frequency shifts induced through time-dependent changes of external parameters. The progress of the CSAC project has triggered various new approaches that can potentially improve the clock performance. In the following section some of these more advanced techniques are discussed.

3.18.4.2 Resonance Contrast

It has been mentioned earlier that the resonance contrast (as well as the CPT width) is more favorable on the D_1 line as compared to the D_2 line. This has been experimentally (and theoretically) demonstrated for rubidium (Stähler *et al.* 2002) as well as for cesium (Lutwak *et al.* 2003). However, when using circularly polarized light on the D_1 line, a new problem arises due to the existence of a second dark state, the so-called stretched state or incoherent trapping state $|S_{1/2}: F = 2I + 1, m_F = \pm F\rangle$. This open system has been successfully modeled theoretically through a system with four levels, or with three levels and a trap (Vanier *et al.* 2003b). **Figure 32(a)** shows the level structure of the D_1 line of ^{87}Rb as an example. As the light intensity is increased, more atoms are pumped into the CPT state, but at the same time, more atoms are also lost into the stretched state, and the fraction of the atoms in the dark state drops again. **Figure 33(a)** shows the CPT absorption contrast (CPT amplitude divided by Doppler absorption) as a function of light intensity for a closed system (black) and a more realistic open system (red). **Figure 33(b)** and **33(c)** shows the CPT resonance width and the CPT resonance amplitude calculated for the same two systems, respectively (Shah *et al.* 2004).

A few different methods have been demonstrated in the past years to circumvent this problem of pumping. The simplest way would be to use linearly polarized light, as a combination of right-hand and left-hand circularly polarized light simultaneously. However, it has been shown (Wynands and Nagel 1999) that this forms a closed-loop system (Korsunsky *et al.* 1999, Kosachiov *et al.* 1992, Lukin *et al.* 1999, Shahriar and Hemmer 1990) on the 0–0 levels, which destructively interferes. A coherent dark state can be formed only in atoms that can couple simultaneously to more than one excited state from one pair of ground states (as in **Figure 32(b)**), if the following two conditions are fulfilled (Buckle *et al.* 1986, Kosachiov *et al.* 1991, 1992, Maichen *et al.* 1995, 1996):

$$\frac{g_1}{g_2} = \frac{g_3}{g_4} \quad \text{and} \quad \varphi_1 + \varphi_2 + \varphi_3 + \varphi_4 = 0 \qquad [5]$$

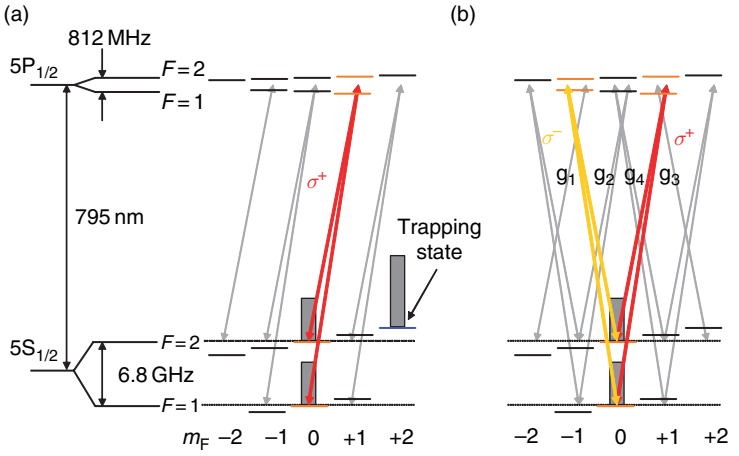


Figure 32 Atomic levels of the ^{87}Rb D_1 line. The levels involved in the coherent population trapping (CPT) coherence are shown in orange. (a) In this open system the circularly polarized light fields (red) optically pump atoms into the trapping state shown in blue. The gray columns indicate the atomic population distribution. (b) Adding an additional pair of light fields (yellow) with opposite circular polarization can eliminate the trapping state. The four optical transitions are coupled by the light fields E_j (with phases φ_j) and Rabi frequencies $g_j, j \in \{1\text{–}4\}$.

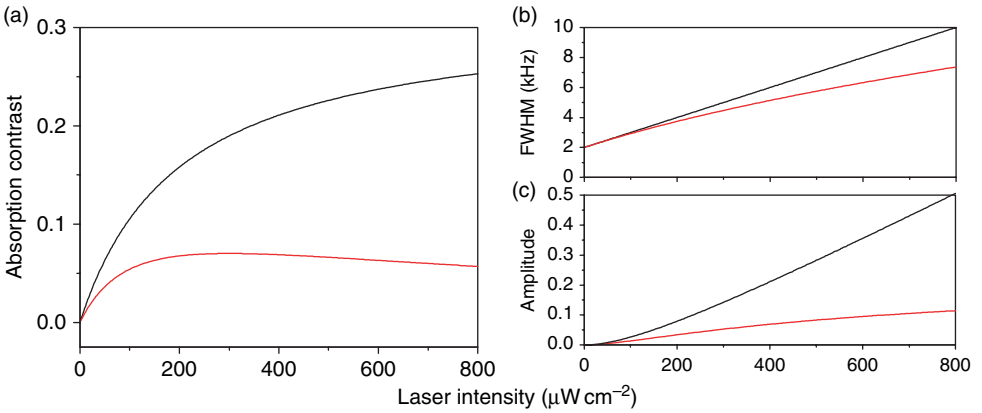


Figure 33 Calculations for a symmetric open (red) and a closed (black) Λ-system. Coherent population trapping (CPT) resonance absorption (a) contrast, (b) width, and (c) amplitude are plotted as a function of laser intensity. (Source: Shah V, Knappe S, Hollberg L, Kitching J 2004 On open Λ systems and the stability of coherent population trapping devices. Unpublished.)

Here, g_j is the Rabi frequency and φ_j is the phase associated with the jth light field, $j \in \{1\text{–}4\}$, as shown in **Figure 32(b)**. For linearly polarized light, on the two $m_F = 0$ ground-state levels $g_1/g_2 = -g_3/g_4$ so that no CPT resonance can be observed.

It has been shown that right-hand and left-hand circularly polarized light can nevertheless be used simultaneously to couple the same pair of ground states, if one beam is delayed by half a microwave wavelength (Affolderbach *et al.* 2002, Alzetta *et al.* 1999). This shifts the relative phase of one beam by π with respect to the other. Jau *et al.* (2004a) have nicely demonstrated that this can be used to increase the contrast of the CPT resonance by eliminating the incoherent trapping state. This transforms the open system back to a closed system (see **Figure 33**). A similar way to create this phase delay is by reflecting the circularly polarized beam back onto itself with opposite circular polarization and by placing the cell a quarter of a microwave wavelength away from the mirror (Affolderbach *et al.* 2002, Alzetta *et al.* 1999, Kargapoltsev *et al.* 2004, Taichenachev *et al.* 2004). It is difficult to integrate either of these methods into a small CSAC due to the delay of half a microwave

wavelength, which is 2 cm for ^{87}Rb. The trapping of atoms in the stretched state can also be circumvented by using two phase-locked single-frequency laser beams with linear orthogonal polarization separated in frequency by the ground-state splitting frequency (Nagel 1999, Zanon *et al.* 2005), but this is not convenient for a compact portable device. One way to implement these ideas into a CSAC physics package has been demonstrated by Shah *et al.* (2006b). They used two VCSEL dies with orthogonal polarizations. The linear light fields were converted into orthogonally circularly polarized fields before entering a microfabricated ^{87}Rb vapor cell. Instead of delaying one optical beam, the microwave phase was delayed by π.

The goal of these very similar techniques is to enhance the resonance contrast and therefore improve the frequency stability of the clock. However, in most cases, the clock is operated at a light intensity where the linewidth is roughly twice the zero-intensity linewidth only (for **Figure 33** this roughly correspond to $200\,\mu W\,cm^{-2}$). Since the linewidths for the open system and the closed system are roughly the same, it would correspond to an increase in CPT resonance amplitude roughly by a factor of 2. Assuming that the noise remains the same, it is questionable whether the more complicated setup justifies an increase in the frequency stability by a factor of 2 (Shah *et al.* 2006b), unless there is another advantage gained at the same time. Especially so, if the CPT amplitude can also be increased by a higher atomic density, until spin-exchange broadening counteracts the gain in frequency stability.

A different way to increase the contrast of CPT resonances has been investigated by Zhu (2003). Here, only a small amount of opposite circularly polarized light is added (without phase delay). While the medium becomes more transparent under CPT conditions for the strong polarization component, the weak one experiences enhanced absorption. Both components acquire opposite phase shifts. A polarization analyzer system is then adjusted such that only a small fraction of the strong polarization component is transmitted, and the relative phase delay in both polarization components is optimized to give a symmetric CPT signal with contrasts above 20%.

When limited to one laser, parallel linear polarizations can be used to couple a coherence between the ground states $|F,\ m_F = \pm 1\rangle$ and $|F+1,\ m_F = \Upsilon 1\rangle$ (Kazakov *et al.* 2005b, Knappe *et al.* 1999, Taichenachev *et al.* 2005, Zhu 2004). These resonances are also insensitive to magnetic fields in first order, but split at higher fields due to the contribution of the nuclear magnetic moment. Since this is a coherence between two states with $\Delta m_F = \pm 2$, it is suppressed at higher buffer gas pressures when the excited-state hyperfine structure is unresolved (Wynands and Nagel 1999).

A different way to accumulate atomic population in the 0–0 states has been demonstrated for optically pumped clocks (Bhaskar 1995) by using a series of microwave pulses. In a similar manner, a population can be optically pumped into one of the $m_F = 0$ states using π-polarized light that is resonant with the transition $|F\rangle \rightarrow |F\rangle$ (Kazakov *et al.* 2005a, Shah *et al.* 2004), where the transition between the $m_F = 0$ states is forbidden. The atoms will then accumulate in this state, while they are probed by circularly polarized light under CPT conditions. This technique also requires the excited-state hyperfine structure to be resolved for efficient pumping and two light fields from orthogonal directions.

3.18.4.3 Buffer Gas

We have seen in Section 3.18.2.3 that the CPT linewidth has a minimum at an optimum buffer gas pressure. From **Figure 5** it can be seen that for a 1-mm cell of ^{87}Rb and nitrogen, this optimum pressure is several atmospheres. At the same time, the buffer gas broadens the optical $S \rightarrow P$ transition. For a frequency-modulated light field, the CPT contrast dramatically reduces when this optical width approaches the ground-state splitting (Arimondo 1996a, Nikonov *et al.* 1994, Post *et al.* 2005), because the carrier and second-order light fields become resonant. They can cause single-photon transitions pumping atoms out of the dark state. They fulfill the dark-state condition as well, but their relative phase differs from the phase between the first-order sidebands. CPT resonances have nevertheless been observed in small cells with several atmospheres of buffer gas, where the ground states were unresolved, when the laser was amplitude-modulated with a $LiNbO_3$ Mach–Zehnder modulator (Jau *et al.* 2004a, Post *et al.* 2005). In most CSACs though, the buffer gas pressure is chosen such that the ground-state hyperfine structure is still resolved, and the linewidth is typically limited by wall collisions.

In table-top systems, antirelaxation wall-coated cells have proven to exhibit extremely narrow linewidths (Bouchiat and Brossel 1966, Robinson *et al.* 1958). Alkenes (C_nH_{2n+2}), such as tetracontane or paraflint, as well as silanes gave good results for Rb

and Cs Zeeman and hyperfine transitions (Budker *et al.* 2005, Frueholz *et al.* 1983, Risley *et al.* 1980, Robinson and Johnson 1982). However, accommodation coefficients (a measure of the average number of wall collisions before decoherence) for hyperfine transitions are found to be worse than for Zeeman transitions by about an order of magnitude (Robinson and Johnson 1982). Nevertheless, it could be advantageous to use wall coatings for small clocks, possibly in combination with low pressures of buffer gases. Advantages could include a reduced temperature dependence (Section 3.18.3.2.3.1), narrower linewidth, lower operating temperatures, or resolved excited-state hyperfine structure. The latter could allow for other interrogation schemes that couple ground states with $\Delta m_F = \pm 2$ (Kazakov *et al.* 2005b, Knappe *et al.* 1999, Taichenachev *et al.* 2005, Zhu 2004), which are suppressed at higher buffer gas pressures (Wynands and Nagel 1999). Research is progressing in this area to find new wall-coating materials that are amenable to microfabrication of vapor cells and that can withstand higher bonding and operating temperatures.

3.18.4.4 Spin Exchange

As cell length decreases, the alkali vapor density usually increases to maintain sufficient absorption through the cell (typically ~50%). At these densities, another broadening mechanism starts to dominate the CPT linewidth, caused by collisions between alkali atoms (Happer 1972, Walker and Happer 1997), as mentioned in Section 3.18.1.3. These so-called spin-exchange collisions conserve the total angular momentum of the pair of alkali atoms. By pumping the atoms into the stretched state with maximum angular momentum, the linewidth of the microwave resonance can be dramatically reduced (Jau *et al.* 2003, 2004b). However, this resonance frequency is strongly dependent on the magnetic field ($21\,\mathrm{kHz}\,\mu\mathrm{T}^{-1}$ for ^{87}Rb and $25\,\mathrm{kHz}\,\mu\mathrm{T}^{-1}$ for Cs), which has to be precisely controlled (see Section 3.18.5.2).

3.18.4.5 Noise Suppression

Most of the noise on the CPT resonance is from the VCSEL through frequency and/or intensity fluctuations (see Section 3.18.2.2.2). When locked to the top of the atomic resonance line, these contributions can be comparable to that of the photon shot noise (Kitching *et al.* 2001c). Furthermore, polarization mode-competition noise can be severe for some VCSELs, but negligible for others. They can be converted into intensity fluctuations when using a polarizer. To reduce the influence of laser-induced noise, two laser beams can be passed through the same vapor cell, inhibiting the same laser noise but different CPT signals.

Gerginov *et al.* (2006a) demonstrated a simple setup in a microfabricated vapor cell of area $2\,\mathrm{mm} \times 1\,\mathrm{mm}$ using two parallel beams, one with circular polarization and another with linear polarization. For the reasons described in Section 3.18.4.2, only the circularly polarized beam causes a CPT resonance. Subtracting the two photodiode signals cancels most of the laser-induced noise, but leaves the full CPT signal. It does not cancel the shot noise, but rather doubles it, so this technique improves the clock stability only if dominated by laser noise, e.g., when locked to the side of the resonance line. Additional noise caused by the CPT resonance itself cannot be reduced in this setup.

Additional noise reduction might be possible, however, with a differential detection technique, where the two beams have orthogonal circular polarization and propagate parallelly through the same region of the cell (Rosenbluh *et al.* 2006) in a similar manner as described in Section 3.18.4.2. The relative phase delay between the two beams here is $\pi/2$, as compared to π previously. This causes the CPT resonance not to have an absorptive Lorentzian lineshape, but rather a combination of absorptive and dispersive Lorentzians. The CPT signal in both beams has an opposite sign for the dispersive part of the signal, so that the difference signal is dispersive in shape. Because the laser noise on both signals is similar, it cancels. A noise suppression of 20 dB has been measured in a table-top setup.

3.18.5 Other CSAC Approaches

3.18.5.1 Introduction

Until now we have primarily discussed passive CSACs based on locking a VCO to a CPT transmission signal. There has been a series of other approaches, of which a few are introduced in the following sections. Unfortunately, not much has been published on these.

3.18.5.2 End-State CSAC

It has been mentioned in Section 3.18.4.4 that microwave resonance broadening due to spin-exchange

collisions between the alkali atoms can be largely reduced when the atoms are pumped into the stretched state (Jau *et al.* 2004b). This can lead to much narrower resonance lines and better stability of CSACs. A gain in sensitivity can be achieved by increasing the cell temperature into a region where spin-exchange collisions usually dominate the linewidth broadening, e.g., high buffer gas pressures. In general, this method works with CPT and microwave interrogation, but for CPT, a second laser beam would be required for optical pumping. Therefore, the scheme was implemented using microwave interrogation (Jau *et al.* 2006).

Since the end-state resonance frequency is very sensitive to magnetic fields, the magnetic field has to be stabilized precisely. This can be done by measuring the Lamor frequency and implementing an additional servo to stabilize the magnetic field (Jau *et al.* 2006) or by alternating between the two opposite stretched states to find the unshifted mean frequency. Furthermore, by choosing the right laser detuning, light shifts have been reduced simultaneously (Jau *et al.* 2006). Potentially, better stabilities could also be reached in microwave-interrogated CSACs, because no high FM is required and therefore lower noise VCSELs might be used.

3.18.5.3 Nanomechanically Regulated CSAC

An RF-interrogated CSAC that uses a mechanical motion to modulate the magnetic field at the location of the ^{87}Rb atoms has been proposed. This oscillation can be induced by using micromechanical choppers that attenuate the magnetic field depending on their position, oscillating at 6.8 GHz, for example (Lee *et al.* 2004).

Even though coupling of alkali atoms to the mechanical oscillation of a MEMS structure has not been demonstrated at gigahertz frequencies yet, it has been shown that double-resonance transitions can occur in ^{87}Rb atoms inside a MEMS vapor cell by using a magnetic cantilever oscillating at 15 kHz (Wang *et al.* 2006).

3.18.5.4 CPT Maser

Under CPT conditions, a coherence is created between the two ground states, oscillating at the difference frequency between them. As a result, the amount of light absorbed by the alkali vapor decreases. At the same time the oscillating magnetization can be detected through stimulated emission in a cavity (Vanier *et al.* 1998). This will happen when a microwave cavity surrounding the cell is tuned to the same ground-state hyperfine frequency. This signal can be used to lock the LO, which generates the modulation on the laser, by synchronous detection (**Figure 34**). While laser noise is not transmitted directly to the microwave signal, a change in sideband intensities and detunings can nevertheless change the output frequency through variations of the ground-state coherence (Godone *et al.* 2000, Levi *et al.* 2000). Using a method similar to the simple CPT clock, the modulation index can be chosen such that the overall light shift is reduced (Vanier *et al.* 1999).

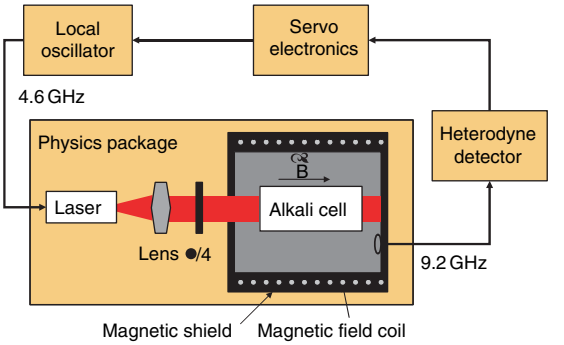


Figure 34 Conceptual setup of a coherent population trapping (CPT) maser.

Furthermore, effects such as cavity pulling, that is, detuning of the cavity with respect to the resonance and shifts due to finite microwave power and phase of the emitted field, can affect the frequency stability of the maser (Godone *et al.* 2000). Even though better stabilities can be potentially reached with a CPT maser (Vanier *et al.* 2003a), the size and the complication of the microwave cavity have prevented it from being used in chip-scale clocks.

3.18.5.5 Raman Oscillator

We have seen in Sections 3.18.1.3 and 3.18.1.4 that the ground-state coherence induced under CPT conditions can be detected in different ways: as a change in the transmission (or fluorescence) or through stimulated microwave emission. Yet another way of detecting the resonance is by detecting the light on a fast photodetector and by detecting the beat signal at the ground-state splitting frequency, e.g., 6.8 GHz for ^{87}Rb (Vukičević *et al.* 2000). No modulation of the VCSEL is necessary to detect this signal, since the second light field is created through stimulated

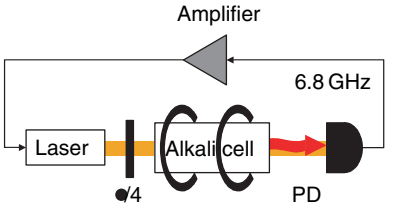


Figure 35 Closed-loop Raman oscillator. The 6.8-GHz signal is fed back to the injection current of the laser.

Raman scattering. However, the linewidth of this resonance can be fairly large in order to achieve enough Raman gain to detect the RF signal, and strong Stark shifts due to high laser intensities can limit the clock stability. This stimulated Raman oscillator design can provide better frequency insensitivity to accelerations when compared to the passive DC CPT clock, because the signal is derived from a purely atomic process.

To increase the signal-to-noise ratio, and allow for lower light intensity, a feedback loop can be included that feeds the amplified 6.8-GHz signal back onto the laser injection current (Vukičević *et al.* 2000) (**Figure 35**). The concept is similar to the one of optoelectronic oscillators as nonatomic systems (Yao and Maleki 1996). This setup requires a rigid structure and low loop delay times to reduce acceleration sensitivity, as changes in the RF phase are translated into frequency changes directly. Strekalov *et al.* (2006) demonstrated a fractional frequency instability of 5×10^{-10} at 1 s with this technique, when using a VCSEL and a 1-cm cell in zero magnetic field. One major problem with this setup is frequency tuning of the VCSEL as the RF power circulating in the loop varies.

Finally, a VCO can be inserted into the loop to modulate the laser (Vukičević *et al.* 2000). The 6.8-GHz signal detected on the photodetector can then be used to control the frequency of the VCO. Fractional frequency instabilities in the 10^{-11} range have been measured this way on the 0–0 transition in a centimeter-sized cell.

3.18.5.6 Ramsey-Type CPT Interrogation

In optically pumped clocks, narrower resonance lines can be achieved, when a (pulsed) Ramsey interrogation (Ramsey 1950, Thomas *et al.* 1982) is used rather than a continuous interrogation. Some of the first CPT resonances were measured in an atomic beam by using Ramsey-type interrogation. In a similar manner, a pulsed CPT clock has been proposed in a vapor cell (Zanon *et al.* 2005). The first pulse creates a ground-state coherence in the atoms followed by a precession in the dark. Coherence is then measured with a second pulse. One advantage of such a system is the reduction of power broadening. In the case of a pulsed technique, while continuous CPT resonances broaden linearly with light power, the width of the Ramsey fringes is given by the inverse time between pulses, e.g., the period during which the coherence precesses in the dark. Unfortunately, the decoherence time in a 1-mm cell is on the order of milliseconds.

3.18.5.7 N-Resonances

While CPT is a two-photon resonance, the so-called *N*-resonance is a three-photon process, which results in similar narrow lines (Phillips *et al.* 2005, Zibrov *et al.* 2005). **Figure 36(a)** shows a level diagram of this scheme. A two-photon resonance is created between the two ground states $|1\rangle$ and $|2\rangle$ by the two red-detuned light fields E_1 and E_2 if the Raman condition is fulfilled, i.e., the difference frequency between these two states is equal to the ground-state splitting frequency $\delta_R = \omega_1 - \omega_2 - \Delta_{HFS} = 0$. This condition is simultaneously probed by the field E_1 resonant with the transition $|2\rangle \rightarrow |3\rangle$. Off Raman resonance, E_1 pumps most of the atoms into the state $|1\rangle$. When tuned onto Raman resonance, however, atoms are pumped back into the state $|2\rangle$, increasing the absorption. This can be seen schematically in **Figure 36(b)**, where the transmission of E_1 is plotted as a function of Raman detuning. The experimental setup is fairly similar to the conventional CPT case with a laser modulated at the full hyperfine ground-state splitting frequency. The major difference is the use of a Fabry–Perot etalon to transmit only E_1 onto the detector in order to reach high resonance contrasts up to 30% (Phillips *et al.* 2005) (contrast is defined here as h_{Nres}/B_{GND}). It has been found that contrasts are higher on the D_2 line as compared to the D_1 line, the opposite of the case of CPT (Novikova *et al.* 2006a).

The shot noise in this system can be potentially lower than in the case of CPT, since the absorption is increased on resonance, and light shifts can be reduced in a way similar to that developed for CPT (Levi *et al.* 2000, Vanier *et al.* 1999, Zhu and Cutler 2000) by choosing a favorable modulation index (Novikova *et al.* 2006b).

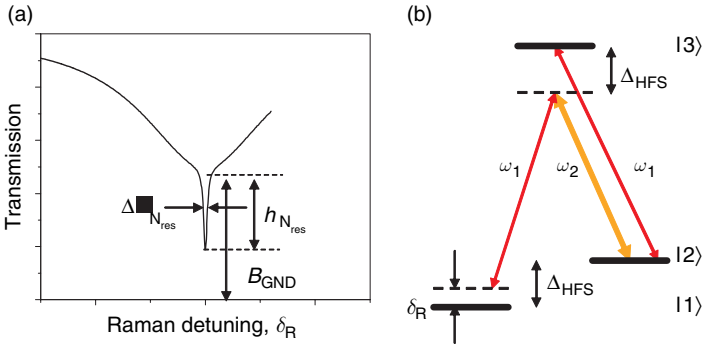


Figure 36 *N*-Resonance scheme. (a) Probe light transmission (ω_1) versus Raman detuning. (b) Three-level diagram showing probe (ω_1) and drive (ω_2) fields with their respective detunings.

3.18.5.8 Others

Recently, a miniature atomic clock was proposed in a solid-state system (White and Hajimiri 2005). The vanadium ion in magnesium oxide shows relatively narrow transitions around 1 MHz linewidth at 1.1 GHz. Such a solid-state atomic clock can be potentially very small and consumes very little power, with frequency instabilities around 10^{-8} to 10^{-9} at 1 s.

CPT spectroscopy is chosen as an all-optical interrogation technique in most approaches. One of the major arguments for this is the lack of a microwave cavity, which constrains the smallest possible size of the clock to a few centimeters. However, it has been shown that microwave resonances can nevertheless be seen in millimeter-sized cells without the use of a cavity, by the use of only small microwave power (Robinson 2003). **Figure 37(a)** shows a photograph of such a setup, where a 1-mm-diameter cell, made from an optical fiber with a CO_2 laser, is inside a copper microwave loop resonator. The magnetic RF field required to cause a clock transition can be roughly estimated to be 0.4 μT for a line of 3 kHz width. Creating this field in a current loop of 2 mm diameter would require roughly 0.6 mA of current, or 20 μW of power into 50 Ω. When using a tuned circuit cavity, the power could be reduced by the quality factor. Unloaded *Q*s around 200 seem possible with simple helical resonators like the one shown in **Figure 37(b)**. Even though these are only rough estimates, it can be seen that it does not necessarily require a high-*Q* large microwave cavity or a huge amount of power to excite a microwave clock resonance. One issue in such a lumped-parameter setup could arise from the stability of the microwave phase. The setup shown in **Figure 37** relies on the near-field and therefore has a decent confinement of the field over the dimension of the cell, but the setup nevertheless needs external shielding to ensure stability of the field.

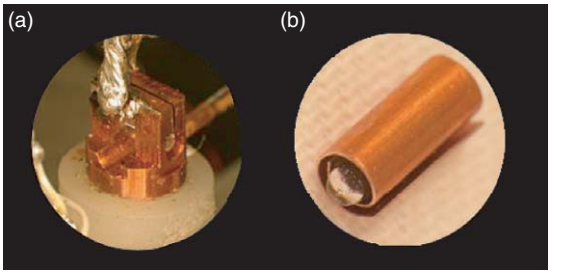


Figure 37 Photograph of a rubidium vapor cell of diameter 1 mm (a) inside a microwave loop and (b) inside a helical resonator (b) with a resonance frequency of 6.83 GHz and an unloaded *Q* of 200 (with and without cells).

3.18.6 Other MEMS Atomic Sensors

The technology discussed in this chapter is not limited to CSACs, but could rather open an avenue to a wide variety of chip-scale atomic sensors. Atomic and molecular transitions have proven useful in applications such as magnetometers (Allred *et al.* 2002, Bloom 1962), gyroscopes (Frazer 1963, Gustavson *et al.* 1997), accelerometers (Clauser 1988), gravimeters (Snadden *et al.* 1998), RF power meters (Camparo 1998b), and optical information storage (Fleischhauer and Lukin 2000, Liu *et al.* 2001). Furthermore, a stable frequency reference might be of value not only in the microwave regime, but in the optical regime as well. Stable optical wavelengths are used in devices such as interferometers, distance measurement, microscopy, and chemical trace detection. Combining wafer-level MEMS fabrication techniques with atomic stabilities and the ability to make absolute measurements without recalibration

could make these chip-scale atomic devices a viable technology for small, low-power applications.

Some chip-scale sensors based on these microfabrication techniques have already been demonstrated already. Two chip-scale atomic magnetometers (CSAMs) were constructed similar to the CASC physics package shown in **Figure 8**. One of them used CPT interrogation on a magnetically sensitive hyperfine transition (Schwindt *et al.* 2004) and reached a sensitivity of 50 pT $Hz^{-1/2}$ at 10 Hz in a 12-mm^3 package consuming 195 mW power. The so-called Mx magnetometer was based on the alignment of atoms in a magnetic field (Aleksandrov *et al.* 1995, Gröger *et al.* 2006). The precession frequency of the atoms was then used as a direct measure of the total magnetic field. The sensitivity was 6 pT $Hz^{-1/2}$ between 1 and 100 Hz bandwidth (Schwindt *et al.* 2006).

Furthermore, a microfabricated saturated absorption spectroscopy (Hänsch *et al.* 1971, Letokhov 1976) setup was demonstrated as an example of an optical wavelength reference with a MEMS cell and miniature optical components (**Figure 38**) (Knappe *et al.* 2006). It shows sub-Doppler linewidths on the order of 25 MHz at 795 nm optical wavelength and could be used to stabilize a diode laser to this transition. Frequency stabilities similar to those in table-top setups are expected.

These are just two examples of applications where microfabricated alkali cells are used in atomic sensors. While alkali atoms are often favored for spectroscopic applications, due to their relatively low vapor pressure and favorable transition frequencies, the technology introduced here is not necessarily limited to alkali atoms. We therefore expect chip-scale atomic devices to find applications in a wide variety of areas where size, cost, and power consumption need to be combined with high precision.

Acknowledgments

I thank John Kitching, Hugh Robinson, Zoya Popović, Leo Hollberg, and Robert Wynands for extensive reading of this manuscript as well as for their helpful and critical suggestions. Alan Brannon and Vishal Shah contributed through useful discussions.

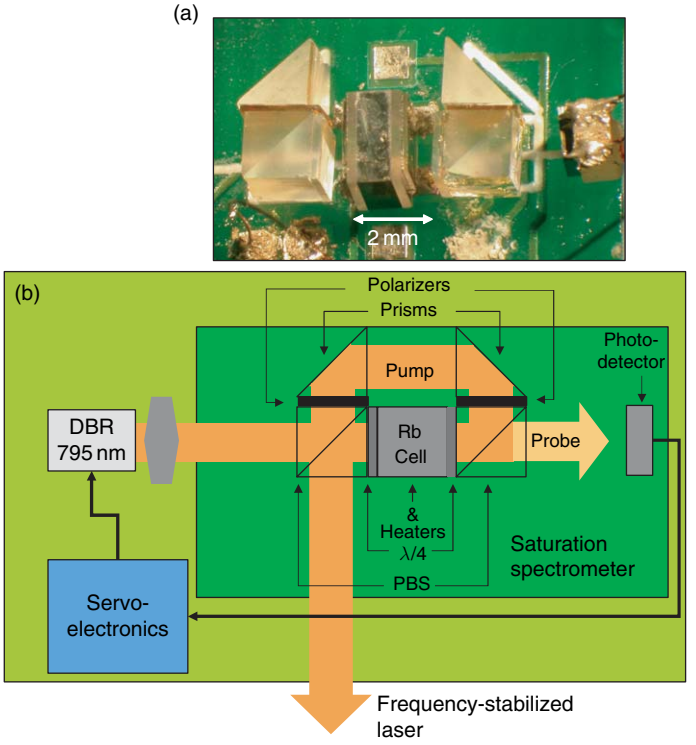


Figure 38 (a) Photograph of a microfabricated saturated absorption spectrometer setup. (b) Schematic of the microfabricated setup (dark green), which consists of a vapor cell with two heaters, two polarizing beamsplitters, two polarizers, two prisms, two quarter waveplates, and a photodetector. The light green shows a schematic of a stable wavelength reference including a laser locked to a stable sub-Doppler absorption resonance.

Commercial Rb standards: See, for example, Symmetricom X-72 Precision Rubidium Oscillator, Stanford Research Systems PRS10 Rubidium Frequency Standard, Accubeat AR-100B Rubidium Frequency Standard, Frequency Electronics FE-5658A Rubidium Atomic Frequency Standard, Temex LCR-900 Low Cost Rubidium; references are for technical clarity and do not imply endorsement by NIST.

References

Affolderbach C, Nagel A, Knappe S, Jung C, Wiedenmann D, Wynands R 2000 Nonlinear spectroscopy with a vertical-cavity surface-emitting laser (VCSEL). *Appl. Phys. B Lasers Opt.* **70**, 407–13

Affolderbach C, Knappe S, Wynands R, Taichenachev A V, Yudin V I 2002 Electromagnetically induced transparency and absorption in a standing wave. *Phys. Rev. A* **65**, 043810

Affolderbach C, Droz F, Mileti G 2006 Experimental demonstration of a compact and high-performance laser-pumped rubidium gas cell atomic frequency standard. *IEEE Trans. Instrum. Meas.* **55**, 429–35

Aleksandrov E B, Balabas M V, Vershovski A K, Ivanov A E, Yakobson N N, Velichanski V L, Senkov N V 1995 Laser pumping in the scheme of an Mx-magnetometer. *Opt. Spectrosc.* **78**, 325–32

Alexandrov E B, Balabas M V, Budker D, English D, Kimball D F, Li C H, Yashchuk V V 2002 Light-induced desorption of alkali-metal atoms from paraffin coating. *Phys. Rev. A* **66**, 042903

Allan D W 1987 Time and frequency (time-domain) characterization, estimation, and prediction of precision clocks and oscillators. *IEEE Trans. UFFC* **34**, 647–54

Allard N, Kielkopf J 1982 The effect of neutral nonresonant collisions on atomic spectral lines. *Rev. Mod. Phys.* **54**, 1103–82

Allred J C, Lyman R N, Kornack T W, Romalis M V 2002 High-sensitivity atomic magnetometer unaffected by spin-exchange relaxation. *Phys. Rev. Lett.* **89**, 130801

Alzetta G, Gozzini A, Moi L, Orriols G 1976 Experimental-method for observation of Rf transitions and laser beat resonances in oriented Na vapor. *Nuovo Cim.* **36**, 5–20

Alzetta G, Botti L, Cartaleva S, Celli R M, Rossi A 1999 Superposition of states and quantum interference in rubidium by diode laser modulation. *Proceedings of SPIE – 10th International Society for Optical Engineering,* Varna, Bulgaria, pp. 195–9

Arditi M, Carver T R 1961 Pressure, light, and temperature shifts in optical detection of 0-0 hyperfine resonance of alkali metals. *Phys. Rev.* **124**, 800–9

Arimondo E 1996a Coherent population trapping in laser spectroscopy.: In: Wolf E (ed.) p. *Progress in Optics. Elsevier Science Publication B V, Amsterdam, The Netherlands, Vol. 35*

Arimondo E 1996b Relaxation processes in coherent-population trapping. *Phys. Rev. A* **54**, 2216–23

Arimondo E, Orriols G 1976 Non-absorbing atomic coherences by coherent 2-photon transitions in a 3-level optical-pumping. *Lett. Nuovo Cim.* **17**, 333–8

Aspect A, Arimondo E, Kaiser R, Vansteenkiste N, Cohentannoudji C 1988 Laser cooling below the one-photon recoil energy by velocity-selective coherent population trapping. *Phys. Rev. Lett.* **61**, 826–9

Avalon Photonics, Avalon Photonics Datacom VCSELs, http://www.avap.ch

Balabas M V, Budker D, Kitching J, Schwindt P D D, Stalnaker J E 2006 Magnetometry with millimeter-scale antirelaxation-coated alkali-metal vapor cells. *J. Opt. Soc. Am. B* **23**, 1001–6

Barrat J P, Cohen-Tannoudji C 1961 Etude du pompage optique dans le formalisme de la matrice densité. *J. Phys. Radium* **22**, 329–36

Bentley C L Jr., Liu J 1999 LWI in a driven lambda three-level atom and effects of the probe laser on EIT. *Opt. Commun.* **169**, 289–99

Beverini N, Minguzzi P, Strumia F 1971 Foreign-gas-induced cesium hyperfine relaxation. *Phys. Rev. A* **4**, 550–5

Beverini N, Strumia F, Rovera G 1981 Buffer gas-pressure shift in the $M_f = 0 - M_f = 0$ ground-state hyperfine line in Cs. *Opt. Commun.* **37**, 394–6

Bhaskar N D 1995 Potential for improving the rubidium frequency standard with a novel optical pumping scheme using diode lasers. *IEEE Trans. UFFC* **42**, 15–20

Bloch M, Pascaru I, Stone C, Mcclelland T 1993 Subminiature rubidium frequency standard for commercial applications. *IEEE International Frequency Control Symposium,* Salt Lake City, UT, USA, pp. 164–77

Bloch M, Mancini O, Mcclelland T 2000 Performance of rubidium and quartz clocks in space. *2000 IEEE Frequency Control Symposium,* Kansas City, MO, USA, pp. 505–9

Bloembergen N, Purcell E M, Pound R V 1948 Relaxation effects in nuclear magnetic resonance absorption. *Phys. Rev.* **73**, 679–712

Bloom A L 1962 Principles of operation of the rubidium vapor magnetometer. *Appl. Opt.* **1**, 61–8

Bouchiat M A, Brossel J 1966 Relaxation of optically pumped Rb atoms on paraffin-coated walls. *Phys. Rev.* **147**, 41–54

Brandt S, Nagel A, Wynands R, Meschede D 1997 Buffer-gas-induced linewidth reduction of coherent dark resonances to below 50 Hz. *Phys. Rev. A* **56**, R1063–66

Brannon A, Breitbarth J, Popović Z 2005 A low-power, low phase noise local oscillator for chip-scale atomic clocks. *2005 IEEE MTT-S International Microwave Symposium,* Long Beach, CA, pp. 1535–8

Brannon A, Breitbarth J, Janković M, Popović Z 2006a Low-power, low phase noise local oscillators for chip-scale atomic clocks. *IEEE Trans. MTT-S,* submitted

Brannon A, Janković M, Breitbarth J, Popović Z, Gerginov V, Shah V, Knappe S, Hollberg L, Kitching J 2006b A local oscillator for chip-scale atomic clocks at NIST. *IEEE Frequency Control Symposium,* Miami, FL, USA, pp. 443–7

Buckle S J, Barnett S M, Knight P L, Lauder M A, Pegg D T 1986 Atomic interferometers – Phase-dependence in multilevel atomic transitions. *Opt. Acta* **33**, 1129–40

Budker D, Hollberg L, Kimball D F, Kitching J, Pustelny S, Yashchuk V V 2005 Microwave transitions and nonlinear magneto-optical rotation in anti-relaxation-coated cells. *Phys. Rev. A* **71**, 012903

Camparo J C 1998a Conversion of laser phase noise to amplitude noise in an optically thick vapor. *J. Opt. Soc. Am. B* **15**, 1177–86

Camparo J C 1998b Atomic stabilization of electromagnetic field strength using Rabi resonances. *Phys. Rev. Lett.* **80**, 222–5

Camparo J 2004 Frequency equilibration and the light-shift effect for block IIR GPS rubidium clocks. *Precise Time and Time Interval (PTTI) Meeting,* Washington, DC, USA

Camparo J C, Buell W F 1997 Laser PM to AM conversion in atomic vapors and short term clock stability. *IEEE Frequency Control Symposium*, Orlando, FL, USA, pp. 253–8

Camparo J C, Frueholz R P, Volk C H 1983 Inhomogeneous light shift in alkali-metal atoms. *Phys. Rev. A* **27**, 1914–24

Camparo J C, Klimcak C M, Herbulock S J 2005 Frequency equilibration in the vapor-cell atomic clock. *IEEE Trans. Instrum. Meas.* **54**, 1873–80

Chang-Hasnain C J, Harbison J P, Hasnain G, Von Lehmen A C, Florez L T, Stoffel N G 1991 Dynamic, polarization, and transverse mode characteristics of vertical cavity surface emitting lasers. *IEEE J. Quantum Electron.* **27**, 1402–9

Chantry P J, Liberman I, Verbanets W R, Petronio C F, Cather R L, Partlow W D 1996 Miniature laser-pumped cesium cell atomic clock oscillator. *IEEE International Frequency Control Symposium,* Honolulu, HI, pp. 1002–10

Chavez-Pirson A, Ando H, Saito H, Kanbe H 1993 Polarization properties of a vertical cavity surface emitting laser using a fractional layer superlattice gain medium. *Appl. Phys. Lett.* **62**, 3082–4

Choquette K D, Leibenguth R E 1994 Control of vertical-cavity laser polarization with anisotropic transverse cavity geometries. *IEEE Photon. Technol. Lett.* **6**, 40–2

Choquette K D, Schneider R P Jr., Lear K L, Geib K M 1994 Low threshold voltage vertical-cavity lasers fabricated by selective oxidation. *Electron. Lett.* **30**, 2043–4

Clauser J F 1988 Ultrahigh sensitivity accelerometers and gyroscopes using neutral atom matter-wave interferometry. *Physica B and C* **151**, 262–72

Coffer J G, Anderson M, Camparo J C 2002 Collisional dephasing and the reduction of laser phase-noise to amplitude-noise conversion in a resonant atomic vapor. *Phys. Rev. A* **65**, 033807

Copley G, Krause L 1969 Trapping and quenching of potassium resonance radiation. *Can. J. Phys.* **47**, 533–9

Cyr N, Têtu M, Breton M 1993 All-optical microwave frequency standard – A proposal. *IEEE Trans. Instrum. Meas.* **42**, 640–9

Dalton B J, Knight P L 1982 The effects of laser field fluctuations on coherent population trapping. *J. Phys. B* **15**, 3997–4015

Darpa Mto Csac Project, http://www.darpa.mil/mto

Debernardi P, Gian Paolo B 2003 Coupled mode theory: A powerful tool for analyzing complex VCSELs and designing advanced device features. *IEEE J. Quantum Electron.* **9**, 905–17

Degen C, Fischer I, Elsäßer W, Fratta L, Debernardi P, Bava G P, Brunner M, Hävel R, Moser M, Gulden K 2001 Transverse modes in thermally detuned oxide-confined vertical-cavity surface-emitting lasers. *Phys. Rev. A* **63**, 023817

Di Sopra F M, Zappe H P, Moser M, Hovel R, Gauggel H P, Gulden K 1999 Near-infrared vertical-cavity surface-emitting lasers with 3-MHz linewidth. *IEEE Photon. Technol. Lett.* **11**, 1533–5

Dicke R H 1953 The effect of collisions upon the Doppler width of spectral lines. *Phys. Rev.* **89**, 472–3

Ekinci K L, Roukes M L 2005 Nanoelectromechanical systems. *Rev. Sci. Instrum.* **76**, 061101

Enloe L H, Rodda J L 1965 Laser phase-locked loop. *Proc. IEEE* **52**, 165–6

Erhard M, Nußmann S, Helm H 2000 Power broadening and Doppler effects of coherent dark resonances in Rb. *Phys. Rev. A* **62**, 061802

Erklund E J, Shkel A M 2007 Glass blowing on wafer level. *J. Microelectromech. Syst.,* **16**, 232–9

Espe W 1966 *Materials of High Vacuum Technology*. Pergamon, Oxford

Fleischhauer M, Lukin M D 2000 Dark-state polaritons in electromagnetically induced transparency. *Phys. Rev. Lett.* **84**, 5094–7

Fleischhauer M, Imamoglu A, Marangos J P 2005 Electromagnetically induced transparency: Optics in coherent media. *Rev. Mod. Phys.* **77**, 633–73

Frazer J T 1963 Optically pumped magnetic resonance gyroscope and direction sensor. *US PTO* 3(103) 621

Fruehauf H 2001 Fast ''direct-P(Y)'' GPS signal acquisition using a special portable clock. *33rd Annual Precise Time and Time Interval (PTTI) Meeting,* Long Beach, CA, USA, pp. 359–69

Frueholz R P, Volk C H, Camparo J C 1983 Use of wall coated cells in atomic frequency standards. *J. Appl. Phys.* **54**, 5613–17

Fumio K, Susumu K, Kenichi I 1989 Room-temperature continuous wave lasing characteristics of a GaAs vertical cavity surface-emitting laser. *Appl. Phys. Lett.* **55**, 221–2

Gerginov V, Knappe S, Schwindt P D D, Shah V, Liew L, Moreland J, Robinson H G, Hollberg L, Kitching J, Brannon A, Breitbarth J, Popovic Z 2005 Component-level demonstration of a microfabricated atomic frequency reference. *Proceedings of the 2005 IEEE International Frequency Control Symposium and Exposition,* Vancouver, Canada, pp. 758–66

Gerginov V, Knappe S, Shah V, Schwindt P D D, Hollberg L, Kitching J 2006 Long-term frequency instability of atomic frequency references based on coherent population trapping and microfabricated vapor cells. *J. Opt. Soc. Am. B* **23**, 593–7

Gerginov V, Knappe S, Shah V, Hollberg L, Kitching J 2007 Laser noise cancellation in single-cell CPT clocks. *Rev. Sci. Instrum.,* accepted

Gierkink S L J, Levantino S, Frye R C, Samori C, Boccuzzi V 2003 A low-phase-noise 5-GHz CMOS quadrature VCO using superharmonic coupling. *IEEE J. Solid-State Circ.* **38**, 1148–54

Godone A, Levi F, Micalizio S, Vanier J 2000 Theory of the coherent population trapping maser: A strong-field self-consistent approach. *Phys. Rev. A* **62**, 053402

Godone A, Levi F, Micalizio S, Vanier J 2002 Dark-line in optically-thick vapors: Inversion phenomena and line width narrowing. *Eur. Phys. J. D* **18**, 5–13

Gong F, Jau Y Y, Jensen K, Happer W 2006 Electrolytic fabrication of atomic clock cells. *Rev. Sci. Instrum.* **77**, 076101

Grabherr M, Jager R, Michalzik R, Weigl B, Reiner G, Ebeling K J 1997 Efficient single-mode oxide-confined GaAs VCSEL's emitting in the 850-nm wavelength regime. *IEEE Photon. Technol. Lett.* **9**, 1304–6

Gray H R, Whitley R M, Stroud C R Jr. 1978 Coherent trapping of atomic populations. *Opt. Lett.* **3**, 218–20

Gröger S, Bison G, Schenker J L, Wynands R, Weis A 2006 A high-sensitivity laser-pumped Mx magnetometer. *Eur. Phys. J. D* **38**, 239–47

Guo H, Lal A 2003 Self-powered photon source. *12th Int. Conf. Solid State Sensors, Actuators and Microsystems (Transducers '03),* Boston, MA, USA, pp. 1474–7

Gustavson T L, Bouyer P, Kasevich M A 1997 Precision rotation measurements with an atom interferometer gyroscope. *Phys. Rev. Lett.* **78**, 2046–9

Häensch T W, Levenson M D, Schawlow A L 1971 Complete hyperfine structure of a molecular iodine line. *Phys. Rev. Lett.* **26**, 946–9

Happer W 1972 Optical pumping. *Rev. Mod. Phys.* **44**, 169–249

Harris S E 1997 Electromagnetically induced transparency. *Phys. Today* **50**, 36–42

Harris S E, Field J E, Imamoglu A 1990 Nonlinear optical processes using electromagnetically induced transparency. *Phys. Rev. Lett.* **64**, 1107–10

Hemmer P R, Ezekiel S, Leiby C C 1983 Stabilization of a microwave-oscillator using a resonance Raman transition in a sodium beam. *Opt. Lett.* **8**, 440–2

Hemmer P R, Shahriar M S, Lamelarivera H, Smith S P, Bernacki B E, Ezekiel S 1993 Semiconductor-laser excitation of Ramsey fringes by using a Raman transition in a cesium atomic-beam. *J. Opt. Soc. Am. B* **10**, 1326–9

Hemmer P R, Katz D P, Donoghue J, Croningolomb M, Shahriar M S, Kumar P 1995 Efficient low-intensity optical-phase conjugation based on coherent population trapping in sodium. *Opt. Lett.* **20**, 982–4

Ho J, Pascaru I, Stone C, Mcclelland T 1998 New Rubidium frequency standard designs for telecommunications applications. *IEEE International Frequency Control Symposium,* Pasadena, CA, USA, pp. 80–3

Hsu W-T, Nguyen C T C 2002 Stiffness-compensated temperature-insensitive micromechanical resonators. *15th IEEE Int. Conf. Micro Electro Mech. Syst.,* Las Vegas, NV, USA, pp. 731–4

Huffaker D L, Shin J, Deppe D G 1994 Low threshold half-wave vertical-cavity lasers. *Electron. Lett.* **30**, 1946–7

Ivanov E N, Tobar M E 2006 Low phase-noise microwave oscillators with interferometric signal processing. *IEEE Trans MTT-S* **54**, 3284–94

Ivanov E N, Tobar M E, Woode R A 1998 Microwave interferometry: Application to precision measurements and noise reduction techniques. *IEEE Trans. UFFC* **45**, 1526–36

Jau Y Y, Post A B, Kuzma N N, Braun A M, Romalis M V, Happer W 2003 The physics of miniature atomic clocks: 0–0 versus ''end'' resonances. *Proceedings of the Annual IEEE International Frequency Control Symposium,* Tampa, FL, pp. 33–6

Jau Y Y, Miron E, Post A B, Kuzma N N, Happer W 2004a Push-pull optical pumping of pure superposition states. *Phys. Rev. Lett.* **93**, 160802

Jau Y Y, Post A B, Kuzma N N, Braun A M, Romalis M V, Happer W 2004b Intense, narrow atomic-clock resonances. *Phys. Rev. Lett.* **92**, 110801

Jau Y Y, Post A, Gong F, Miron E, Kuzma N N, Happer W 2006 Physics and cell manufacturing of miniature atomic clocks. *NICT CSAC Workshop,* Tokyo, Japan

Jewell J L, Scherer A, Mccall S L, Lee Y H, Walker S, Harbison J P, Florez L T 1989 Low-threshold electrically pumped vertical-cavity surface-emitting microlasers. *Electron. Lett.* **25**, 1123–4

Jung C, Jager R, Grabherr M, Schnitzer P, Michalzik R, Weigl B, Muller S, Ebeling K J 1997 4.8 mW singlemode oxide confined top-surface emitting vertical-cavity laser diodes. *Electron. Lett.* **33**, 1790–1

Kaiser J, Degen C, Elsässer W 2002 Polarization-switching influence on the intensity noise of vertical-cavity surface-emitting lasers. *J. Opt. Soc. Am. B* **19**, 672–7

Kargapoltsev S V, Kitching J, Hollberg L, Taichenachev A V, Velichansky V L, Yudin V I 2004 High-contrast dark resonance in sigma(+)–sigma(−) optical field. *Laser Phys. Lett.* **1**, 495–9

Kasapi A, Jain M, Yin G Y, Harris S E 1995 Electromagnetically induced transparency: Propagation dynamics. *Phys. Rev. Lett.* **74**, 2447–50

Kazakov G, Matisov B, Mazets I, Mileti G 2005a Optimized scheme of a rubidium all-optical frequency standard. *Tech. Phys. Lett.* **31**, 1009–10

Kazakov G, Matisov B, Mazets I, Mileti G, Delporte J 2005b Pseudoresonance mechanism of all-optical frequency-standard operation. *Phys. Rev. A* **72**, 063408

Khanna A P S, Gane E, Chong T 2003 A 2GHz voltage tunable FBAR oscillator. *IEEE MTT-S International Microwave Symposium*, Philadelphia, PA, pp. 717–20

Kibble B P, Copley G, Krause L 1967 Effect of imprisonment of radiation in sodium vapor on the measured lifetime of the $3\,^2P$ States. *Phys. Rev.* **153**, 9–12

King R, Wiedenmann D, Schnitzer P, Jaeger R, Michalzik R, Ebeling K J 1998 Single-mode and multimode 2D VCSEL arrays for parallel optical interconnects. *Conference Digest – IEEE International Semiconductor Laser Conference,* Nara, Japan, pp, 103–4

Kitching J 2003 *Local oscillator requirements for chip-scale atomic clocks. Unpublished*

Kitching J, Hollberg L, Knappe S, Wynands R 2001a Compact atomic clock based on coherent population trapping. *Electron. Lett.* **37**, 1449–51

Kitching J, Hollberg L, Knappe S, Wynands R 2001b Frequency-dependent optical pumping in atomic Λ-systems. *Opt. Lett.* **26**, 1507–9

Kitching J, Robinson H G, Hollberg L, Knappe S, Wynands R 2001c Optical-pumping noise in laser-pumped, all-optical microwave frequency references. *J. Opt. Soc. Am. B* **18**, 1676–83

Kitching J, Knappe S, Hollberg L 2002 Miniature vapor-cell atomic-frequency references. *Appl. Phys. Lett.* **81**, 553–5

Kitching J, Knappe S, Schwindt P D D, Shah V, Hollberg L, Liew L, Moreland J 2004 Power dissipation in vertically integrated chip-scale atomic clocks. *Proceedings of the 2004 IEEE International Frequency Control Symposium,* Montreal, Canada, pp. 781–4

Kleppner D, Goldenberg H M, Ramsey N F 1962 Theory of the hydrogen maser. *Phys. Rev.* **126**, 603–15

Knappe S 2001 Dark resonance magnetometers and atomic clocks. *Ph.D. Thesis*, University of Bonn

Knappe S, Kemp W, Affolderbach C, Nagel A, Wynands R 1999 Splitting of coherent population-trapping resonances by the nuclear magnetic moment. *Phys. Rev. A* **61**, 012508

Knappe S, Wynands R, Kitching J, Robinson H G, Hollberg L 2001 Characterization of coherent population-trapping resonances as atomic frequency references. *J. Opt. Soc. Am. B* **18**, 1545–53

Knappe S, Kitching J, Hollberg L, Wynands R 2002 Temperature dependence of coherent population trapping resonances. *Appl. Phys. B Lasers Opt.* **74**, 217–22

Knappe S, Velichansky V, Robinson H G, Kitching J, Hollberg L 2003a Compact atomic vapor cells fabricated by laser-induced heating of hollow-core glass fibers. *Rev. Sci. Instrum.* **74**, 3142–5

Knappe S, Velichansky V, Robinson H G, Liew L, Moreland J, Kitching J, Hollberg L 2003b Atomic vapor cells for miniature frequency references. *Proceedings of the 2003 IEEE International Frequency Control Symposium and PDA Exhibition Jointly with the 17th European Frequency and Time Forum,* Tampa, FL, pp. 31–2

Knappe S, Hollberg L, Kitching J 2004a Dark-line atomic resonances in submillimeter structures. *Opt. Lett.* **29**, 388–90

Knappe S, Shah V, Schwindt P D D, Hollberg L, Kitching J, Liew L A, Moreland J 2004b A microfabricated atomic clock. *Appl. Phys. Lett.* **85**, 1460–2

Knappe S, Gerginov V, Schwindt P D D, Shah V, Robinson H, Hollberg L, Kitching J 2005a Atomic vapor cells for chip-scale atomic clocks with improved long-term frequency stability. *Opt. Lett.* **30**, 2351–3

Knappe S, Schwindt P D D, Shah V, Hollberg L, Kitching J, Liew L, Moreland J 2005b A chip-scale atomic clock based on Rb-87 with improved frequency stability. *Opt. Expr.* **13**, 1249–53

Knappe S, Robinson H G, Hollberg L 2007 Microfabricated saturated absorption laser spectrometer. *Optics Express* **15**, 6293–9

Kocharovskaya O 1992 Amplification and lasing without inversion. *Phys. Rep.* **219**, 175–90

Korsunsky E A, Leinfellner N, Huss A, Baluschev S, Windholz L 1999 Phase-dependent electromagnetically induced transparency. *Phys. Rev. A* **59**, 2302–5

Kosachiov D, Matisov B, Rozhdestvensky Y 1991 Coherent population trapping – Sensitivity of an atomic system to the relative phase of exciting fields. *Opt. Commun.* **85**, 209–12

Kosachiov D V, Matisov B G, Rozhdestvensky Y V 1992 Coherent phenomena in multilevel systems with closed interaction contour. *J. Phys. B* **25**, 2473–88

Koyama Y, Matsura H, Atsumi K, Nakamuta K, Sakai M, Maruyama I 2000 An ultra-miniature rubidium frequency standard. *IEEE International Frequency Control Symposium,* Kansas City, MO, pp. 694–9

Kubena R L, Stratton F P, Chang D T, Joyce R J, Hsu T Y, Lim M K, M'closkey R T 2005 MEMS-based quartz oscillators and filters for on-chip integration. *Proceedings of the 2005 IEEE International Frequency Control Symposium and Exposition,* Vancouver, Canada, pp. 122–7

Kusters J A, Adams C A 1999 Performance requirements of communication base station time standards. *RF Design* May 28–30, 1999

Kwakernaak M H, Lipp S, Mcbride S, Zanzucchi P, Chan W K, Khalfin V B, An H, Whaley J R D, Willner B I, Ulmer A, Li J Z, Davis T, Braun A M, Abeles J H, Post A, Jau Y-Y, Kuzma N N, Happer W 2004 Components for batch-fabricated chip-scale atomic clock. *Technical Digest of 36th Annual Precise Time and Time Interval (PTTI) Meeting,* Washington, DC, USA, pp. 355

Lai S L, Ramanath G, Allen L H, Infante P, Ma Z 1995 High-speed ($10^{4\circ}$C/s) scanning microcalorimetry with monolayer sensitivity (J/m^2). *Appl. Phys. Lett.* **67**, 1229–31

Larsson A, Carlsson C, Gustavsson J, Haglund A, Modh P, Bengtsson J 2004 *Direct high-frequency modulation of VCSELs and applications in fibre optic RF and microwave links*. New J. Phys. **6**, 176. http://www.njp.org/

Laws A, Chang R Y J, Bright V M, Lee Y C 2005 Thermal management for chip-scale atomic clocks. *Proc. ASME/ Pacific Rim Tech. Conf. Exhibit. Integration and Packaging of MEMS, NEMS, and Electronic Systems: Advances in Electronic Packaging,* San Francisco, CA, 2005, pp. 741–6

Lear K L, Mar A, Choquette K D, Kilcoyne S P, Schneider R P Jr., Geib K M 1996 High-frequency modulation of oxide-confined vertical cavity surface emitting lasers. *Electron. Lett.* **32**, 457–8

Lee C, Guo H, Radhakrishnan S, Lal A, Szekely C, Mcclelland T A, Pisano A 2004 A batch fabricated rubidium-vapor resonance cell for chip scale atomic clock. *Solid State Sensor and Actuator Workshop,* Hilton Head Island, SC, USA

Lee H J, Adams C S, Kasevich M, Chu S 1996 Raman cooling of atoms in an optical dipole trap. *Phys. Rev. Lett.* **76**, 2658–61

Lee S, Nguyen C T C 2004 Mechanically-coupled micromechanical resonator arrays for improved phase noise. *Proceedings of the 2004 IEEE International Frequency Control Symposium and Exposition,* Montreal, Canada, pp. 144–50

Letokhov V S 1976 Saturation spectroscopy. In: Shimoda K (ed.) *High-Resolution Laser Spectroscopy.* Springer-Verlag, New York

Levi F, Godone A, Novero C, Vanier J 1997 On the use of a modulated laser for hyperfine frequency excitation in passive frequency standards. *11th Annual European Frequency and Time Forum*, Neuchatel, Switzerland

Levi F, Godone A, Vanier J 2000 Light shift effect in the coherent population trapping cesium maser. *IEEE Trans. UFFC* **47**, 466–70

Li S-S, Lin Y-W, Xie Y, Ren Z, Nguyen C T C 2004 Micromechanical ''hollow-disk'' ring resonators. *17th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS),* Maastricht, Netherlands, pp. 821–4

Liew L A, Knappe S, Moreland J, Robinson H, Hollberg L, Kitching J 2004 Microfabricated alkali atom vapor cells. *Appl. Phys. Lett.* **84**, 2694–6

Liew L, Moreland J, Knappe S, Shah V, Schwindt P, Gerginov V, Kitching J, Hollberg L 2005 Microfabricated alkali atom vapor cells with in-situ heating for atomic-based sensors. *Proc. 3rd Int. Symp. Sensor Science,* Juelich, Germany, pp. 181–3

Liew L, Moreland J, Gerginov V 2006 Wafer-level fabrication and filling of cesium vapor cells for chip-scale atomic devices. *Proceedings of the 20th Eurosensors Conference,* Goteberg, Sweden, paper W1B-P2

Liew L-A, Moreland J, Gerginov V 2007 Water-level filling of microfabricated atomic vapour cells based on thin-film deposition and photolysis of cesium azide. *Appl. Phys. Lett.* **90**, 114106

Liu C, Dutton Z, Behroozi C H, Vestergaard Hau L 2001 Observation of coherent optical information storage in an atomic medium using halted light pulses. *Nature* **409**, 490–3

Lukin M D, Yelin S F, Fleischhauer M, Scully M O 1999 Quantum interference effects induced by interacting dark resonances. *Phys. Rev. A* **60**, 3225–8

Lutwak R, Emmons D, Riley W, Garvey R M 2002 The chip-scale atomic clock – Coherent population trapping vs. conventional interrogation. *34th Annual Precise Time and Time Interval (PTTI) Meeting,* Reston, VA, USA, pp. 539–50

Lutwak R, Emmons D, English T, Riley W, Duwel A, Varghese M, Serkland D K, Peake G M 2003 The chip-scale atomic clock – Recent development progress. *35th Annual Precise Time and Time Interval (PTTI) Meeting,* San Diego, CA, USA, pp. 1–12

Lutwak R, Deng J, Riley W, Varghese M, Leblanc J, Tepolt G, Mescher M, Serkland D K, Geib K M, Peake G M 2004 The chip-scale atomic clock – Low-power physics package. *36th Annual Precise Time and Time Interval (PTTI) Meeting,* Washington, DC, USA, pp. 1–14

Lutwak R, Vlitas P, Varghese M, Mescher M, Serkland D K, Peake G M 2005 The MAC – A miniature atomic clock. *Joint Meeting of the IEEE International Frequency Control Symposium and the Precise Time and Time Interval (PTTI) Systems and Applications Meeting,* Vancouver, BC, Canada, pp. 752–7

Lyons H 1950 The atomic clock – A universal standard for frequency and time. *Am. Scholar* **19**, 159–68

Maichen W, Gaggl R, Korsunsky E, Windholz L 1995 Observation of phase-dependent coherent population trapping in optically closed atomic systems. *Europhys. Lett.* **31**, 189–94

Maichen W, Renzoni F, Mazets I, Korsunsky E, Windholz L 1996 Transient coherent population trapping in a closed loop interaction scheme. *Phys. Rev. A* **53**, 3444–8

Marangos J P 1998 Topical review electromagnetically induced transparency. *J. Mod. Opt.* **45**, 471–503

Mathur B S, Tang H, Happer W 1968 Light shifts in the alkali atoms. *Phys. Rev.* **171**, 11–19

McClelland T, Pascaru I, Shtaerman I, Stone C, Szekely C, Zacharski J, Bhaskar N D 1995 Subminiature rubidium frequency standard: Manufacturability and performance results from production units. *IEEE International Frequency Control Symposium,* San Francisco, CA, pp. 39–52

McClelland T, Pascaru I, Shtaerman I, Stone C, Szekely C, Zacharski J, Bhaskar N D 1996 Subminiature rubidium frequency standard: Performance improvements. *IEEE International Frequency Control Symposium,* Honolulu, HI, pp. 1011–16

McClelland T, Ho J, Pascaru I, Stone C 1999 Rb frequency standard with expanded operating temperature range. *Joint Meeting of the European Frequency and Time Forum and the IEEE International Frequency Control Symposium,* Besancon, France, pp. 137–40

Merimaa M, Lindvall T, Tittonen I, Ikonen E 2003 All-optical atomic clock based on coherent population trapping in ^{85}Rb. *J. Opt. Soc. Am. B* **20**, 273–9

Mileti G, Jinquan D, Walls F L, Jennings D A, Drullinger R E 1998 Laser-pumped rubidium frequency standards: New analysis and progress. *IEEE J. Quantum Electron.* **34**, 233–7

Mourant J M, Imbornone J, Tewksbury T 2000 A low phase noise monolithic VCO in SiGe BiCMOS. *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium,* Boston, MA, pp. 65–8

Mueller G, Mueller M, Wicht A, Rinkleff R H, Danzmann K 1997 Optical resonator with steep internal dispersion. *Phys. Rev. A* **56**, 2385–9

Mukaihara T, Koyama F, Iga K 1993 Engineered polarization control of GaAs/AlGaAs surface-emitting lasers by anisotropic stress from elliptical etched substrate hole. *IEEE Photon. Technol. Lett.* **5**, 133–5

Mukaihara T, Ohnoki N, Hayashi Y, Hatori N, Koyama F, Iga K 1995a Polarization control of vertical-cavity surface emitting lasers using a birefringent metal/dielectric polarizer loaded on top distributed Bragg reflector. *IEEE J. Quantum Electron.* **1**, 667–73

Mukaihara T, Ohnoki N, Hayashi Y, Hatori N, Koyama F, Iga K 1995b Excess intensity noise originated from polarization fluctuation in vertical-cavity surface-emitting lasers. *IEEE Photon. Technol. Lett.* **7**, 1113–15

Mukoyama N, Otama H, Kuwata Y, Nakayama H 2006 850 nm Single mode oxide-confined VCSEL. *NICT Workshop on Chip Scale Atomic Clock Status and Potentiality,* NICT Headquarters, Koganei, Tokyo, Japan

Murphy J, Skidmore T 1994 A low-cost atomic clock: impact on the national airspace and GNSS availability. *Proc. ION GPS-94; 7th Int. Meet. Satellite Division of the Institute of Navigation,* Salt Lake City, UT, USA, pp. 1329–36

Myatt C J, Newbury N R, Wieman C E 1993 Simplified atom trap by using direct microwave modulation of a diode laser. *Opt. Lett.* **18**, 649–51

Nagel A 1999 Precision spectroscopy of coherent dark states in thermal cesium vapor. *Ph.D. thesis*, University of Bonn

Nguyen C T C 2005 MEMS technology for timing and frequency control. *Proceedings of the 2005 IEEE International Frequency Control Symposium and Exposition,* Vancouver, Canada, pp. 1–11

Nguyen C T-C, Howe R T 1993 Microresonator frequency control and stabilization using an integrated micro oven. *Dig. Tech. Pap., 7th Int. Conf. Solid-State Sensors and Actuators (Transducers '93),* Yokohama, Japan, pp. 1040–3

Nikonov D E, Rathe U W, Scully M O, Zhu S Y, Fry E S, Li X F, Padmabandu G G, Fleischhauer M 1994 Atomic coherence effects within the sodium D(1) manifold .2. Coherent optical-pumping. *Quantum Opt.* **6**, 245–60

Novikova I, Phillips D F, Zibrov A S, Walsworth R L, Taichenachev A V, Yudin V I 2006a *N*-resonances for D-1 and D-2 transitions. *Opt. Lett.* **31**, 2353–5

Novikova I, Phillips D F, Zibrov A S, Walsworth R L, Taichenachev A V, Yudin V I 2006b Cancellation of light shifts in an *N*-resonance clock. *Opt. Lett.* **31**, 622–4

Numai T, Kurihara K, Kuhn K, Kosaka H, Ogura I, Kajita M, Saito H, Kasahara K 1995 Control of light-output polarization for surface-emitting-laser type device by strained active layer grown on misoriented substrate. *IEEE J. Quantum Electron.* **31**, 636–42

Orriols G 1979 Non-absorption resonances by non-linear coherent effects in a 3-level system. *Nuovo Cim.* **53**, 1–24

Ostermann J M, Debernardi P, Jalics C, Michalzik R 2005 Polarization-stable oxide-confined VCSELs with enhanced single-mode output power via monolithically integrated inverted grating reliefs. *IEEE J. Quantum Electron.* **11**, 982–9

Pang W, Hongyu Y, Hao Z, Eun Sok K 2005 Electrically tunable and temperature compensated FBAR. *IEEE MTT-S International Microwave Symposium Digest,* Long Beach, CA, pp. 1279–82

Phillips D F, Fleischhauer A, Mair A, Walsworth R L, Lukin M D 2001 Storage of light in atomic vapor. *Phys. Rev. Lett.* **86**, 783–6

Phillips D F, Novikova I, Zibrov S, Smallwood C, Taichenachev A V, Yudin V I, Walsworth R L, Zibrov A S 2005 A novel absorption resonance for atomic clocks. *Proceedings of the 2005 IEEE International Frequency Control Symposium and Exposition,* Vancouver, Canada, pp. 767–73

Phillips D F, Novikova I, Wang C, Walsworth R L, Crescimanno M 2006 Modulation-induced frequency shifts in a coherent-population-trapping-based atomic clock *J. Opt. Soc. Am. B* **22**, 305–10

Photonics U Reliability study, http://www.ulmphotonics.de/

Post A B, Jau Y Y, Kuzma N N, Happer W 2005 Amplitude-versus frequency-modulated pumping light for coherent population trapping resonances at high buffer-gas pressure. *Phys. Rev. A* **72**, 1–17

Pozar D M 2001 *Microwave and RF Design of Wireless Systems*. John Wiley and Sons Inc., New York

Quinn T, (ed.) (2005) Special issue: Fifty years of atomic time-keeping: 1995 to 2005. *Metrologia*, **42**

Radhakrishnan S, Lal A 2005 Alkali metal-wax micropackets for chip-scale atomic clocks. *Dig. Tech. Pap. – Int. Conf. Solid State Sensors and Actuators and Microsystems, Transducers '05,* Seoul, Korea, pp 23–6

Ramsey N F 1950 A molecular beam resonance method with separated oscillating fields. *Phys. Rev.* **78**, 695–9

Risley A, Jarvis S Jr., Vanier J 1980 The dependence of frequency upon microwave power of wall-coated and buffer-gas-filled gas cell ^{87}Rb frequency standards. *J. Appl. Phys.* **51**, 4571–6

Robinson H G 2003 Personal communications

Robinson H, Ensberg E, Dehmelt H 1958 Preservation of spin state in free atom inert surface collisions. *Bull. Am. Phys. Soc.* **3**, 9

Robinson H G, Johnson C E 1982 Narrow ^{87}Rb hyperfine-structure resonances in an evacuated wall-coated cell. *Appl. Phys. Lett.* **40**, 771–3

Rochat P, Leuenberger B, Stehlin X 2002 A new synchronized ultra miniature rubidium oscillator. *IEEE Int. Freq. Contr. Symp,* New Orleans, LA, USA, pp. 451–4

Römisch S, Lutwak R 2006 Low-Power, 4.6-GHz stable oscillator for CSAC. *International Frequency Control Symposium,* Miami, FL, USA, pp. 448–51

Rosenbluh M, Shah V, Knappe S, Kitching J 2006 Differentially detected coherent population trapping resonances excited by orthogonally polarized laser fields. *Opt. Expr.* **14**, 6588–94

Ruther P, Herrscher M, Paul O 2004 A micro differential thermal analysis (/spl mu/DTA) system. *17th IEEE Int. Conf. Micro Electro Mechanical Systems (MEMS),* Maastricht, The Netherlands, pp. 165–8

Sargsyan A, Sarkisyan D, Papoyan A 2006 Dark-line atomic resonances in a submicron-thin Rb vapor layer. *Phys. Rev. A* **73**, 033803

Schmidt O, Wynands R, Hussein Z, Meschede D 1996 Steep dispersion and group velocity below c/3000 in coherent population trapping. *Phys. Rev. A* **53**, R27–30

Schmidt-Kaler F, Eschner J, Morigi G, Roos C F, Leibfried D, Mundt A, Blatt R 2001 Laser cooling with electromagnetically induced transparency: Application to trapped samples of ions or neutral atoms. *Appl. Phys. B* **73**, 807–14

Schwindt P D D, Knappe S, Shah V, Hollberg L, Kitching J, Liew L A, Moreland J 2004 Chip-scale atomic magnetometer. *Appl. Phys. Lett.* **85**, 6409–11

Schwindt P D D, Lindseth B, Knappe S, Shah V, Kitching J 2006 A chip-scale atomic magnetometer with improved sensitivity using the Mx technique. *Appl. Phys. Lett.,* **90**, 081102

Scully M O, Fleischhauer M 1992 High-sensitivity magnetometer based on index-enhanced media. *Phys. Rev. Lett.* **69**, 1360–3

Serkland D K, Peake G M, Geib K M, Lutwak R, Michael Garvey R, Varghese M, Mescher M 2006 VCSELs for atomic clocks. Chun Lei K D C (ed.) *Proceedings of SPIE – The International Society for Optical Engineering*, pp. 613208

Seurin J-F P, Chuang S L, Chirovski L M F, Choquette K D 2002 Novel VCSEL designs deliver high output power. *Laser Focus World* **38**, 119–24

Shah V, Knappe S, Hollberg L, Kitching J 2004 *On open Λ systems and the stability of coherent population trapping devices. Unpublished*

Shah V, Gerginov V, Schwindt P D D, Knappe S, Hollberg L, Kitching J 2006a Continuous light shift correction in modulated CPT clocks. *Appl. Phys. Lett.,* **89**, 151124

Shah V, Knappe S, Schwindt P D D, Gerginov V, Kitching J 2006b Compact phase delay technique for increasing the amplitude of coherent population trapping resonances in open Λ systems. *Opt. Lett.* **31**, 2335–7

Shahriar M S, Hemmer P R 1990 Direct excitation of microwave-spin dressed states using a laser-excited resonance Raman interaction. *Phys. Rev. Lett.* **65**, 1865–8

Shimuzi M, Mukaihara T, Koyama F, Iga K 1991 Polarisation control for surface emitting lasers. *Electron. Lett.* **27**, 1067–9

Simpson T B, Doft F, Golding W M 2003 Using laser diode instabilities for chip-scale stable frequency references. *Precise Time and Time Interval (PTTI) Meeting,* San Diego, CA, USA

Snadden M J, Mcguirk J M, Bouyer P, Haritos K G, Kasevich M A 1998 Measurement of the earth's gravity gradient with an atom interferometer-based gravity gradiometer. *Phys. Rev. Lett.* **81**, 971–4

Stähler M, Wynands R, Knappe S, Kitching J, Hollberg L, Taichenachev A, Yudin V 2002 Coherent population trapping resonances in thermal Rb-85 vapor: D-1 versus D-2 line excitation. *Opt. Lett.* **27**, 1472–4

Steck D 2003 A cesium D line data, http://steck.us/alkalidata

Steck D 2003 A Rubidium 87 D line data, http://steck.us/alkalidata

Stratton F P, Chang D T, Kirby D J, Joyce R J, Tsung-Yuan H, Kubena R L, Yook-Kong Y 2004 A MEMS-based quartz resonator technology for GHz applications. *Proceedings of the 2004 IEEE International Frequency Control Symposium and Exposition,* Montreal, Canada, 2004, pp. 27–34

Strekalov D, Matsko A B, Yu N, Savchenkov A, Maleki L 2006 Application of vertical cavity surface emitting lasers in self-oscillating atomic clocks. *J. Mod. Opt.*, **53**, 2469–84

Sturza M A 1984 GPS navigation using three satellites and a precise clock. *Global Positioning System*. Institute of Navigation, Washington, DC

Suzuki K, Abe Y, Ishihara N, Kudoh H, Fukugawa S 1998 Small-sized Rubidium oscillator. *IEEE International Frequency Control Symposium,* Pasadena, CA, USA, pp. 73–9

Taichenachev A V, Yudin V I, Velichansky V L, Kargapoltsev S V, Wynands R, Kitching J, Hollberg L 2004 High-contrast dark resonances on the D1 line of alkali metals in the field of counterpropagating waves. *JETP Lett.* **80**, 236–40

Taichenachev A V, Yudin V I, Velichansky V L, Zibrov S A 2005 On the unique possibility of significantly increasing the contrast of dark resonances on the D1 line of 87Rb. *JETP Lett.* **82**, 398–403

Tench R E, Peuse B W, Hemmer P R, Thomas J E, Ezekiel S, Leiby C C, Picard R H, Willis C R 1981 2 Laser raman difference technique applied to high-precision spectroscopy. *J. de Phys.* **42**, 45–51

Thomas J E, Hemmer P R, Ezekiel S, Leiby C C, Picard R H, Willis C R 1982 Observation of Ramsey fringes using a stimulated, resonance raman transition in a sodium atomic beam. *Phys. Rev. Lett.* **48**, 867–70

Townes C H 1951 Atomic clocks and frequency stabilization on microwave spectral lines. *J. Appl. Phys.* **22**, 1365–72

Vanier J 2001 Atomic frequency standard. *US PTO 6 320 472*

Vanier J 2005 Atomic clocks based on coherent population trapping: a review. *Appl. Phys. B Lasers Opt.* **81**, 421–42

Vanier J, Audoin C 1989 *The Quantum Physics of Atomic Frequency Standards*. Adam Hilger, Bristol

Vanier J, Audoin C 1992 *The Quantum Physics of Atomic Frequency Standards*. Adam Hilger, Bristol

Vanier J, Kunski R, Cyr N, Savard J Y, Têtu M 1982 On hyperfine frequency-shifts caused by buffer gases – Application to the optically pumped passive rubidium frequency standard. *J. Appl. Phys.* **53**, 5387–91

Vanier J, Godone A, Levi F 1998 Coherent population trapping in cesium: Dark lines and coherent microwave emission. *Phys. Rev. A* **58**, 2345–58

Vanier J, Godone A, Levi F 1999 Coherent microwave emission in coherent population trapping: Origin of the energy and of the quadratic light shift. *Proceedings of the Annual IEEE International Frequency Control Symposium,* Besancon, France, pp. 96–9

Vanier J, Godone A, Levi F, Micalizio S 2003a Atomic clocks based on coherent population trapping: Basic theoretical models and frequency stability. *Proceedings of the Annual IEEE International Frequency Control Symposium,* Tampa, FL, pp. 2–15

Vanier J, Levine M W, Janssen D, Delaney M 2003b Contrast and linewidth of the coherent population trapping transmission hyperfine resonance line in Rb-87: Effect of optical pumping. *Phys. Rev. A* **67**, 065801

Vanier J, Levine M W, Janssen D, Delaney M J 2003d The coherent population trapping passive frequency standard [Rb example]. *IEEE Trans. Instrum. Meas.* **52**, 258–62

Vanier J, Levine M W, Janssen D, Delaney M J 2003e On the use of intensity optical pumping and coherent population trapping techniques in the implementation of atomic frequency standards. *IEEE Trans. Instrum. Meas.* **52**, 822–31

Vanier J, Levine M, Kendig S, Janssen D, Everson C, Delany M 2004 Practical realization of a passive coherent population trapping frequency standard. *IEEE International Frequency Control Symposium,* Montreal, Canada, pp. 92–9

Vanier J, Levine M W, Kendig S, Janssen D, Everson C, Delaney M J 2005 Practical realization of a passive coherent population trapping frequency standard. *IEEE Trans. Instrum. Meas.* **54**, 2531–9

Vestergaard Hau L, Harris S E, Dutton Z, Behroozi C H 1999 Light speed reduction to 17 metres per second in an ultracold atomic gas. *Nature* **397**, 594–8

Vig J R Introduction to quartz frequency standard, SLCET-TR-92-1, http://www.ieee-uffc.org/freqcontrol/quartz/vig/vigtoc.htm

Vig J R 1993 Military applications of high accuracy frequency standards and clocks. *IEEE Trans. UFFC* **40**, 522–7

Vig J R, Yoonkee K 1999 Noise in microelectromechanical system resonators. *IEEE Trans. UFFC* **46**, 1558–65

Vukičević N, Zibrov A S, Hollberg L, Walls F L, Kitching J, Robinson H G 2000 Compact diode-laser based rubidium frequency reference. *IEEE Trans. UFFC* **47**, 1122–6

Walker T G, Happer W 1997 Spin-exchange optical pumping of noble-gas nuclei. *Rev. Mod. Phys.* **69**, 629–42

Wallis G, Pomerantz D 1969 Field assisted glass–metal sealing. *J. Appl. Phys.* **40**, 3946–9

Walls D F, Zoller P 1980 Coherent nonlinear mechanism for optical bistability from three level atoms. *Opt. Commun.* **34**, 260–4

Waltman S 2006 Frequency divider for a chip-scale atomic clock, Unpublished

Wang J, Butler J E, Hsu D S Y, Nguyen T C 2002 CVD polycrystalline diamond high-Q micromechanical resonators. *The 15th IEEE International Conference on Micro Electro Mechanical Systems,* Las Vegas, NV, USA, pp. 657–60

Wang Y-J, Eardley M, Knappe S, Moreland J, Hollberg L, Kitching J 2006 Magnetic resonance in an atomic vapor excited by a mechanical resonator. *Phys. Rev. Lett.* **97**, 227602

White C J, Hajimiri A H 2005 A solid-state atomic frequency standard. *Proceedings of the 2005 IEEE International Frequency Control Symposium and Exposition,* Vancouver, Canada, pp. 940–6

Whitley R M, Stroud C R 1976 Double optical resonance. *Phys. Rev. A* **14**, 1498–513

Wynands R, Nagel A 1999 Precision spectroscopy with coherent dark states. *Appl. Phys. B Lasers Opt.* **68**, 1–25

Yang G M, Macdougal M H, Dapkus P D 1995 Ultralow threshold current vertical-cavity surface-emitting lasers obtained with selective oxidation. *Electron. Lett.* **31**, 886–8

Yao X S, Maleki L 1996 Optoelectronic oscillator for photonic systems. *IEEE J. Quantum Electron.* **32**, 1141–9

Youngner D W, Detry J F, Zook J D 2005 MEMS frequency standard for devices such as atomic clocks. *US PTO 6 900 702 B2*

Zanon T, Guerandel S, De Clercq E, Holleville D, Dimarcq N, Clairon A 2005 High contrast Ramsey fringes with coherent-population-trapping pulses in a double lambda atomic system. *Phys. Rev. Lett.* **94**, 193002

Zhang H, Kim J, Pang W, Yu H, Kim E S 2005 5 5 GHz low-phase-noise oscillator based on FBAR with low TCF. *13th Int. Conf. Solid-State Sensors, Actuators and Microsystems, Dig. Tech. Pap. (Transducers '05),* Seoul, Korea, pp. 1100–1

Zhu M 2003 High contrast signal in a coherent population trapping based atomic frequency standard application. *IEEE International Frequency Control Symposium and PDA Exhibition Jointly with the 17th European Frequency and Time Forum,* Tampa, FL, USA, pp. 16–21

Zhu M 2004 Study of CPT in a vapor cell with optical phase locked diode lasers. *Proceedings of the John Hall Symposium,* Boulder, CO, USA, pp. 156–8

Zhu M, Cutler L S 2000 Theoretical and experimental study of light shift in a CPT-based Rb vapor cell frequency standard. *32nd Annual Precise Time and Time Interval (PTTI) Meeting,* Reston, VA, USA, pp. 311–24

Zhu M, Cutler L S 2001 Coherent population trapping-based frequency standard having a reduced magnitude of total a.c. stark shift. *US PTO 6 201 821*

Zibrov S, Novikova I, Phillips D F, Taichenachev A V, Yudin V I, Walsworth R L, Zibrov A S 2005 Three-photon-absorption resonance for all-optical atomic clocks. *Phys. Rev. A* **72**, 011801

Biography

Svenja Knappe received her diploma in Physics from the University of Bonn, Germany, in 1998. The topic for her diploma thesis was the investigation of single cesium atoms in a magneto-optical trap. She obtained her Ph.D. from the University of Bonn in 2001, with a thesis on "Dark resonance magnetometers and atomic clocks." Since 2001, she has been pursuing research in the Time and Frequency Division at NIST, Boulder, CO. Her research interests include precision laser spectroscopy, atomic clocks and atomic magnetometers, laser cooling, alkali vapor cell technology, applications of semiconductor lasers to problems in atomic physics and frequency control, and miniaturization of atomic spectroscopy.

3.19 Micro-Fuel Cells

Christopher Hebling, Lisbeth Rochlitz, and Thomas Aicher, Fraunhofer Institute for Solar Energy Systems, Freiburg, Germany

3.19.1 Introduction

Micro fuel cell systems have gained an increasing amount of attention as a potential power supply for portable and off-grid electronic applications due to some significant intrinsic advantages as compared with batteries. Depending on the application and the respective boundary conditions of its operation, small fuel cells can be considered both as a substitute and as a supplement to batteries. Fuel cell systems can be customized exactly according to the power demand of the respective electronic device and the geometrical design. The design of a micro fuel cell system has two degrees of freedom, power range and energy capacity, which means that both can be tailored exactly to the requirements of the various electronic devices and the operation time aimed at. However, many aspects, such as technical challenges due to the overall complexity of the fuel cell systems, cost, reliability, and safety issues, as well as the need for a different maintenance required by the customer as compared with that for batteries, should be addressed. In particular the management of the water generated by the electrochemical reaction is one of the major challenges for a reliable and long-term stable fuel cell system.

The advent of fuel cell systems took place when power sources with very sophisticated properties such as no toxic exhaust, low thermal signature, silent mode, variable power output, high energy density, instant refueling, and others were needed for very special applications (submarines, spacecrafts, and premium power). From the mid-1980s, the main emphasis was put on automotive and later on stationary applications with powerful statements about their upcoming market introduction and market penetration. After some disillusion in both areas, the portable and grid-independent fuel cell applications created fairly high expectations due to their intrinsic potential as an energy converter for battery-driven devices.

The major driver for fuel cells is the fact that improvements in battery technology have not kept pace with the growing power demands of particularly consumer data acquisition and communication devices, which again is a result of an increase in intelligence (computation), connectivity (band width), and the promotion of 'always on'. But as it seems now, niche applications with very special and well-defined boundary conditions, such as remote stationary systems (sensors, long-term data acquisition), decentralized microsystem networks, medical

appliances, security cameras, backup power and military systems, will be the first products that are powered by fuel cell systems either in addition to or instead of batteries.

The striking feature of small fuel cell systems is their inherent potential of achieving higher energy densities as compared to secondary batteries, which opens up the possibility to operate the respective devices much longer without recharging. This is due to the high energy density of a wide variety of hydrocarbons, alcohols, or just hydrogen, which can be used as fuel in fuel cell systems either directly or in conjunction with reformers. Unfortunately, in spite of the high specific energy of hydrogen ($32\,W\,h\,g^{-1}$), the use of pure hydrogen has hardly been pursued in micro fuel cell systems due to the challenges of hydrogen storage in high-pressure vessels or metal hydride containers. Liquid fuels such as alcohols are much easier to store and still have a very good energy density: $6.1\,W\,h\,g^{-1}$ ($4800\,W\,h\,l^{-1}$) and $8.6\,W\,h\,g^{-1}$ ($6800\,W\,h\,l^{-1}$) for methanol and ethanol, respectively (Coutanceau *et al.* 2006). Hydrocarbons with even higher energy densities (diesel and gasoline) have to undergo a reforming process in order to generate a hydrogen-rich reformate gas, which can be fed into a proton exchange membrane (PEM) fuel cell.

Due to the independence of the fuel cell system from the fuel container, the amount of energy can be tailored to a given application. In this context, the geometrical adaptability of the fuel cell form factor to a given cavity of the various devices is another important advantage. In addition to this, the system can be an integral part of the electronic device since only the container of fuel has to be refilled or replaced.

3.19.2 The Principles of Fuel Cell Systems

A fuel cell system is an electrochemical device that converts the chemical energy of a fuel directly into electrical and thermal energy with a high conversion efficiency. The fuel cell itself is the basic electrochemical unit where the electrochemical process takes place.

The chemical processes by which the electric current is generated are electrochemical, rather than thermochemical. Since no combustion reaction is involved, fuel cells do not produce any undesirable products normally associated with the oxidation of fossil fuels in conventional energy conversion systems. The fuel cell principle was first demonstrated by William Grove in 1839 but at that time, materials science was to no degree sufficiently developed to realize a functional device.

Generally, fuel cells are classified by their electrolyte: proton exchange membrane fuel cell (PEMFC), direct methanol fuel cell (DMFC) in which a polymer membrane serves as the electrolyte, alkaline fuel cell (AFC) in which the electrolyte is a base such as potassium hydroxide (KOH), phosphoric acid fuel cell (PAFC) in which the electrolyte is phosphoric acid (H_3PO_4), molten carbonate fuel cell (MCFC) in which molten carbonate $(Li,K)CO_3$ serves as the electrolyte transporting carbonate ions, and solid-oxide fuel cell (SOFC) in which the electrolyte is a solid such as yttria (Y_2O_3)-stabilized zirconia (ZrO_2) that transports oxygen ions (Hebling and Yu 2005).

The operation temperature is another way to classify fuel cells. Low-temperature fuel cells have an operation temperature lower than 100 °C. This type includes the PEMFC, the DMFC, and the AFC. The medium-temperature fuel cells have an operating temperature between 100 and 300 °C. This type includes the AFC and the PAFC. High-temperature fuel cells have an operating temperature between 600 and 1000 °C and include the MCFC and the SOFC. But with the advances in various fuel cells, there is a trend that the operating temperatures of fuel cell will be modified within given boundaries. For example, to avoid poisoning the PEMFC with CO, researchers are trying to raise the operating temperature above 100 °C. A higher operating temperature would also be advantageous for the DMFC. Meanwhile, researchers are making an effort to lower the operating temperature of SOFC to simplify the system.

Among the various types of fuel cells, the PEMFC and the DMFC are especially suitable for portable applications due to the low-temperature range and thus the quick-start capability. With the efforts made by research and development institutions and product manufacturers all over the world, micro fuel cells have without doubt a bright future.

On the technical side, the most important tasks are the increase of power and energy density while maintaining the reliability equivalent to the reliability of batteries. Micro fuel cells usually operate under ambient conditions. Reliable operation and simple system architecture are the contradictory requirements for these applications. Thus, to fulfill these requirements, the main technical challenges are to

realize a thermal and water management as passive as possible and to feed suitable reactants. These challenges can be solved for stationary fuel cell systems with active components such as compressors, heaters, cooler, or humidifier, whereas for portable systems only a minimum amount of peripheral elements should be used. The most important functions of the peripheral devices for portable fuel cell systems are to start, control, and stop the fuel supply for the fuel cell stack, to ensure an adequate hydrogen pressure or methanol feeding, and to provide the appropriate air supply for cathode feeding and system cooling. The necessary devices such as valves, pumps, fans, and pressure reducers have to be miniaturized in size and optimized in power consumption. Automatic operation of such a system requires several components that, in many cases, are not yet commercially available and thus need to be developed additionally. **Figure 1** shows a prototype of a 10 W PEM fuel cell system for a digital video recorder, which was developed from the Fraunhofer Alliance Micro Fuel Cells.

Finally, the interaction of all system components with respect to a high efficiency and reliable system performance needs to be controlled in a sophisticated way. This includes the total power management and voltage control with respect to the adaptation of the fuel cell system to the respective electronic device. Since the profile of the electrical load of electronic devices varies during operation, the fuel cell system must be accommodated to the needs of the respective system. Mobile phones, for example, need only some milliwatts in the stand-by mode, short pulses of higher power demand for contacting the next transmitter station, and several watts during the talk time again with additional pulses. Such profiles are a significant challenge for the system architecture of a fuel cell system. A hybrid system combining a fuel cell with a battery or with a super capacitor is considered as an ideal solution for such challenging boundary conditions. If the fuel cell system is to be used as a battery supplement in electronic consumer devices, the same safety standards as for batteries have to be obeyed. Even under abuse, no dangerous situation should occur, as is the case with the common types of batteries such as Ni/Cd, Ni/MH_x, or Li ion. Any kind of maintenance is excluded; however, recharging the fuel cell system by simply replacing an empty storage container by a full one is the change the customer has to accept in his behavior.

In a PEMFC, the reactant and the oxidant are fed into two compartments, which are separated by the membrane electrode assembly (MEA) where electricity is generated due to the electrochemical reaction. A fuel cell consists of reactant gas suppliers (gas diffusion layer), electrochemical active sites (MEA), and the fuel supplier out for the fuel cell (reformer or hydrogen storage). **Figure 2** shows the principle

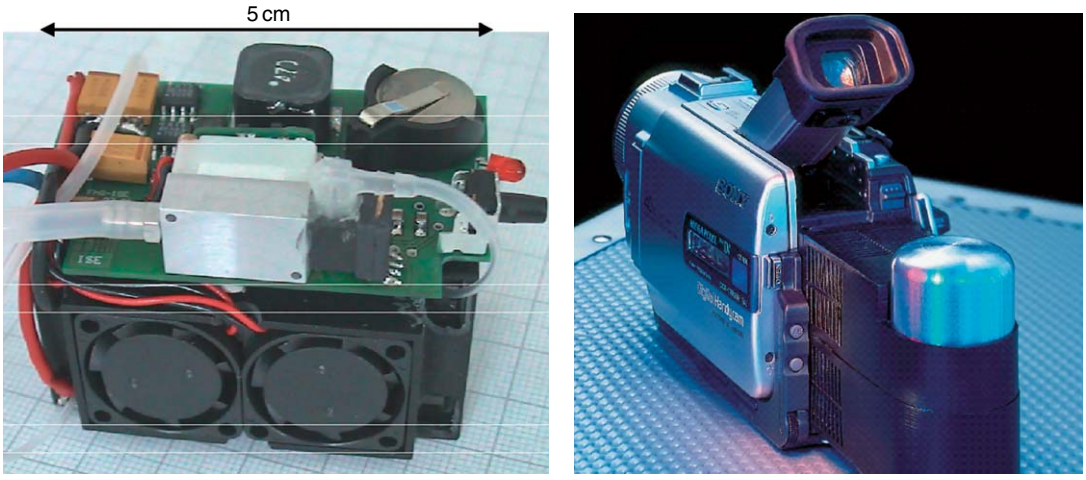


Figure 1 Miniaturized 10 W_p PEM fuel cell system (*left*) integrated into the battery housing of a digital camcorder (*right*). The hydrogen is stored in a metal hydride container located in the back of the fuel cell system. The peripheral components, for example, for microfluidic fuel processing (valves, fans, etc.), define the limits of conventional macroscopic production technology. Further miniaturization requires silicon-based microelectronic fabrication techniques such as microelectromechanical systems (MEMS) technology. (Source: Litterst C, Eccarius S, Hebling C, Zengerle R, Koltay P 2006 *J. Micromech. Microeng.* **16**, 248–53.)

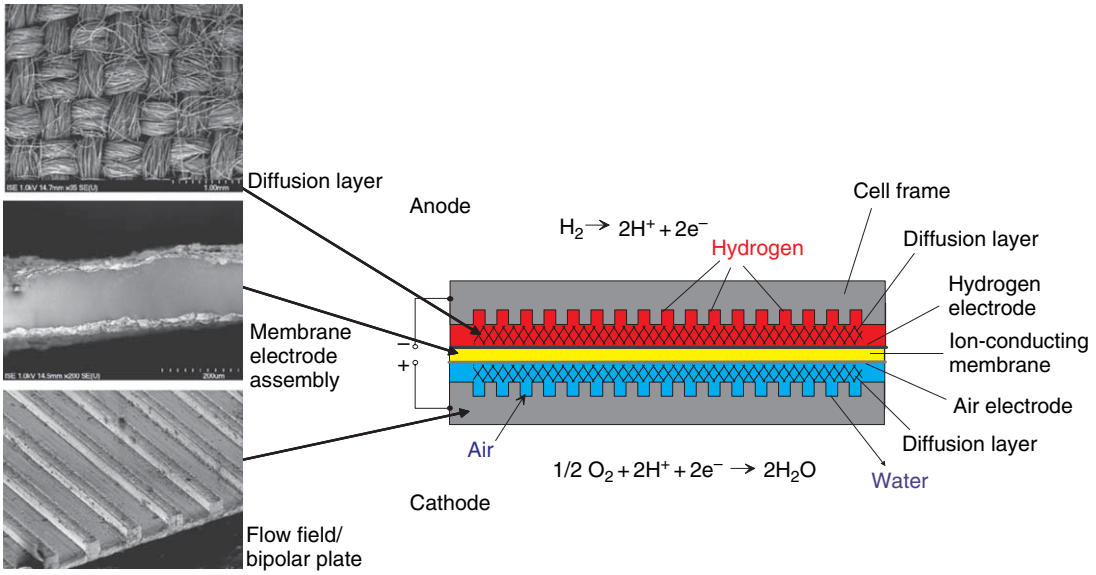


Figure 2 Design of a PEM single fuel cell including the functional units gas diffusion media (GDM), membrane electrode assembly (MEA), and bipolar plate containing the flowfield for the reactants.

setup of a single fuel cell, typically a large number of single cells are stacked on top of each other in order to connect the voltage in series.

The basic electrochemical reactions that take place in a PEMFC are as follows:

At the anode $\varphi_o = 0$ V, at the cathode $\varphi_o = 1.229$ V, which results in a total reaction $E_o = 1.229$ V.

At the anode, in a typical fuel cell design, the hydrogen is broken down into two components: the hydrogen nucleus (proton) and an electron. The electron is transferred from the anode to the cathode through the electric circuit. The proton migrates from the anode to the cathode through the electrolyte. At the cathode, the proton reacts with the electron coming from the load and with the oxygen molecule. They recombine to form water, which is the only chemical product of the fuel cell. The membranes are commonly made of proton-conducting polymers such as perfluorinated ionomers.

The electrochemical reactions that take place at the electrodes need to be facilitated by an electrocatalyst such as platinum or ruthenium. At the two electrodes, the electrochemical reactions occur on the catalyst surface where the electrolyte and the fuel (or oxidant) meet. This interface is often referred to as the triple or three-phase interface, since in some fuel cell designs the fuel (or oxidizer) is gas, the electrolyte is liquid, and the electrode/catalyst surface is solid. The state of this three-phase interface plays a critical role in determining the electrochemical performance of a fuel cell; hence, this triple interface has been the focus of much research in improving the fuel cells.

For the DMFC, the cathode reaction is the same as for the PEMFC cathode, oxygen electrochemical reduction is $\varphi_o = 1.229$ V. While at the DMFC anode, the CH_3OH electrochemical oxidation reaction is $\varphi_o = 0.046$ V, the total reaction is $E_o = 1.183$ V.

Methanol is a liquid fuel that can be oxidized electrochemically in the presence of water. Methanol has a high energy density, is easy to handle, and is completely water-soluble. However, methanol is toxic and the crossover of methanol through the polymer electrolyte membrane results in the poisoning of the cathode catalyst by the formation of a mixed potential. Another problem is the slow anode electrocatalysis of methanol, which requires significantly more noble metal catalyst than the hydrogen PEMFC. As a result, the power density of the DMFC stack at room temperature is currently at least a factor of 5 smaller than that of hydrogen-fed PEMFC.

3.19.3 Micro Fuel Cell Systems Based on MEMS Technologies

Micro fuel cell systems below 10 W and even more in the subwatt range are microchemical systems that offer inherent advantages, such as increased heat

and mass transfer rates, increased safety as a result of smaller volume, enhanced temperature control, on-demand consumption of the reactants, and reduced volume waste streams and thus a very high dynamics for electric load variations, over macrosystems (Kelley *et al.* 2002). Furthermore, integrated microsystems on a chip enable sensing, computing, actuation, control, and communication on a single chip (Lu 2004). Silicon-based micro fuel cell systems are predominantly based on micro direct methanol fuel cells (μDMFC) due to the easy storage and easy handling of liquid fuel as compared to gaseous hydrogen. The anode and the cathode flowfields, usually having channel and rib width on the order of 500 μm and below, are fabricated on Si wafers using the microelectromechanical system (MEMS) technology (Liu *et al.* 2006). Microfabrication permits fine feature resolution, geometry optimization, porosity control in the gas diffusion media (GDM), and the ability to manage the wetting characteristics that are needed to engineer optimal micro fuel cell components (Modroukas *et al.* 2005).

A major target for MEMS-based μDMFCs is to realize a system that is as passive as possible in operation. This means reducing the complexity of the fuel cell stack, methanol sensors, CO_2 separators, electronic controls, methanol feed pump, circulation pump, and pump drivers for the dilution of pure methanol to aqueous methanol solution (usually 2–6 vol.% methanol) to a minimum. In addition to the complexity of such an overall system, the system components add considerable cost to the overall system and consume parasitic electricity, which in turn significantly reduces the net power output. Guo and Faghri (2006) has given an excellent overview of the technologies used by selected developers in order to realize sophisticated μDMFC systems. While some approaches rely on the miniaturization of active components, the major trends pursue passive techniques for driving miniature DMFCs, in particular the cathodic air supply, fuel delivery by capillary forces, and passive CO_2 removal. Fuel storage, CO_2 release, water and thermal management, and orientation-independent operation belong to the main unresolved issues in such systems. In **Figure 3**, a concept that enables fully passive CO_2 gas bubble removal in μDMFC is depicted. The concept is based on a microchannel flowfield with a T-shaped cross section and well-defined tapering angles along the cross section. The tapered channel design creates an intrinsic transport mechanism that removes the gas bubbles from the electrodes by capillary forces only (Litterst *et al.* 2006).

3.19.3.1 Water Management in PEMFC

An excellent stability of the fuel cell operation as well as a high level of miniaturization of the fuel cell system volume is absolutely required for portable electronic applications. These constraints are accompanied by operating conditions that change over time. System start-up, system shut down, and changes in the power demand of the electronic load during operation require an excellent understanding of the transient behavior of fuel cells. In particular, the understanding of the dynamic accumulation and transport of liquid water in the PEMFC is a precondition for a stable operation under varying loads and ambient conditions. Two-phase, dynamic fuel cell models provide insight into the physical phenomena that govern the dynamic operating behavior of a PEMFC.

The complexity of water management in PEMFC systems is due to the fact that the proton exchange membrane is based on cross-linked, sulfonated fluorocarbon membranes and that the electron conductivity of the membrane is strongly affected by the water content. Thus, the membrane

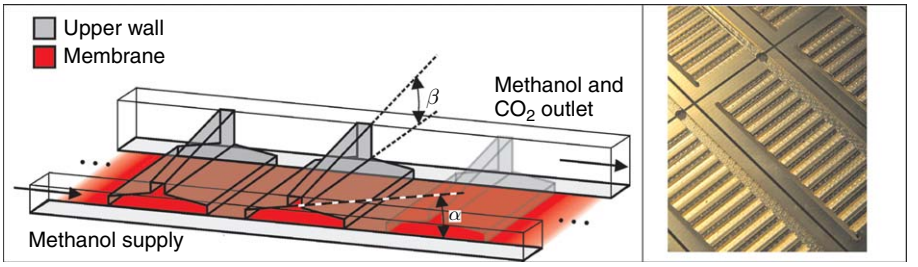


Figure 3 Tapered channel design allowing passive CO_2 gas bubble movement driven by different capillary pressures. (Source: Litterst C, Eccarius S, Hebling C, Zengerle R, Koltay P 2006 *J. Micromech. Microeng.* **16**, 248–53.)

performance will directly follow the water content change, and insufficient water content will lead to a very poor performance and fuel utilization. Since the water content depends both on the local temperature distribution and on the air and hydrogen flow rate, the mathematical description and physical quantification of the time- and spatially-resolved parameters is a demanding challenge, which is being pursued by means of multiphysics and multiscale simulations.

3.19.3.2 Hydrogen Generation and Storage

PEMFCs require hydrogen as feed for their electrochemical energy conversion. The hydrogen can be supplied either in pure form or as a hydrogen-rich gas, containing other compounds as long as they are not harmful to the fuel cell. Pure hydrogen can be provided from either hydrogen storage systems or from reformer systems with subsequent hydrogen separation. A hydrogen-rich gas can be supplied from a fuel processor. Other compounds in the reformer product gas can be nitrogen (if an autothermal reforming (ATR) process is used), carbon dioxide, water, carbon monoxide, and trace amounts of methane and higher hydrocarbons.

The fuel choice is limited by several factors such as the volumetric and gravimetric energy density of the fuel, the weight and lifetime of the fuel container, the conversion efficiency of the chemical energy to electricity, the safety of the fuel and fuel container during refueling and transportation, and finally the fuel distribution infrastructure. In the small-power range, both hydrogen and methanol are very attractive fuels, and also chemical hydrides and even the microreforming of hydrocarbons are moving more and more into the focus of research activities. PEMFCs possess by far higher potential in terms of power density whereas methanol fuel cells dominate from the energy density point of view.

There are several techniques for storing pure hydrogen: as liquid at $-253\,^{\circ}C$ in appropriately isolated vessels, in gaseous form at high pressure, and chemically bound in either metal hydrides or chemical hydrides. Furthermore, the development of carbonaceous materials as another weight-efficient medium for hydrogen storage for fuel cell applications shows promise in the long term. A good overview on hydrogen storage technologies is given by Züttel (2003).

With respect to hydrides, there are two choices:

(1) Metal alloys, which are able to adsorb and desorb hydrogen upon changes in pressure and temperature
(2) Chemical or complex hydrides like $NaBH_4$, which release hydrogen upon heat input or addition of water.

Metal hydride materials, which can be charged and discharged reversibly, are metallic alloys. Hydrogen reacts at elevated temperatures with many transition metals and their alloys to form hydrides. Metal hydrides can absorb large amounts of hydrogen at a constant pressure because of the phase transition. The composition of the alloy determines the temperature and pressure levels for the desorption and the adsorption process. For portable applications, the temperature and the pressure for the desorption process must be close to ambient conditions. These types of alloys, such as AB5, have been under development for many years. The typical hydrogen storage density of these hydride materials is 1.3–1.5 wt.%, corresponding to 150 l of hydrogen per kilogram of hydride material or $450\,W\,h\,kg^{-1}$ or $1050\,W\,h\,l^{-1}$ based on the lower heating value for hydrogen. Some companies have announced that hydrides with 2 wt.% hydrogen will be available in a few years. In addition to the hydride material, there is a container, a filter, and a valve to complete the storage system. These also contribute to the weight and the size of the system, indicating a lower limit for a reasonable size of the container. These hydride materials are considered to be the safest way to store hydrogen since the temperature decreases during discharging and the hydrogen release is decelerated accordingly. The alloys can be charged and discharged several thousands of times with moderate degradation, depending on the purity of the hydrogen used for charging.

Compounds such as chemical or complex hydrides can be used only once. The main difference between complex and metallic hydrides is the transition to an ionic or covalent compound upon hydrogen absorption. The hydrogen in the complex hydrides is often located in the corners of a tetrahedron with B or Al in the center. The negative charge of the anion, $[BH_4]^-$ and $[AlH_4]^-$, is compensated by a cation, e.g., Li or Na. The hydride complexes of borane, the tetrahydroborates $M(BH_4)$, and the tetrahydroaluminates $M(AlH_4)$ are nonflammable and stable in air. They

decompose only at elevated temperatures, often above the melting point of the complex.

Small-scale reformers offer the use of hydrocarbons with their high energy density at reasonable fuel costs. The striking feature is the worldwide availability of fuels such as LPG, gasoline, diesel, or kerosene.

Reformers are used to produce hydrogen from hydrocarbons or alcohols for fuel cells. They consist of the actual reforming reactor, where the fuel is reacted to hydrogen, water, carbon monoxide, and carbon dioxide, with small amounts of methane and other hydrocarbons as by-products. The reforming reaction requires elevated temperatures between 250 °C for methanol and up to 900 °C for liquid hydrocarbons such as kerosene and diesel. In addition, catalysts are used to increase reaction rates and hydrogen yields. In principle there are three reforming processes that can be distinguished by the oxidant. If water is the oxidant the reaction is the so-called steam-reforming (SRF) reaction, which is endothermic. If, on the other hand, oxygen is used as an oxidant, then the reaction is exothermic and is called catalytic partial oxidation (CPOX). If both oxygen and steam are mixed in such a ratio that no heat input is required and no heat release occurs, then the reaction is called autothermal reforming (ATR).

Depending on the fuel cell used in the system, the reformer product gas has to be treated. If PEM fuel cells, which can tolerate up to 20 ppm carbon monoxide only, are utilized, the reformer system also has to comprise shift reactors and further unit operations for removal of carbon monoxide (fine purification) down to the level tolerated by the PEM fuel cell. In **Figure 4** a schematic process flow diagram of a reformer system including gas treatment and fuel cell is depicted.

The feed streams are preheated, evaporated, and superheated up to the operating temperature of the reformer; the heat duty is provided by the gas streams after the reformer and the shift reactors. The reformer also possesses an integrated burner for the fuel cell anode off-gas. In case of a steam reformer, this burner provides the reaction enthalpy, and in the case of CPOX or ATR, this burner provides additional heat duty to preheat the feed streams.

The reformer product is then supplied to one or several shift reactors operating at temperatures between 200 and 350 °C, where the carbon monoxide content of the gas stream is reduced to concentrations of about 0.5–1 vol.% (dry gas). In the subsequent fine purification, the carbon monoxide content is further reduced by either preferential CO oxidation (PROX) or selective CO methanation. Both processes are explained in more detail in the Section 3.19.4. The treated reformate enters the anode of the fuel cell where the electrochemical oxidation of hydrogen occurs under the release of electrical energy and heat. The anode off-gas still contains unreacted hydrogen

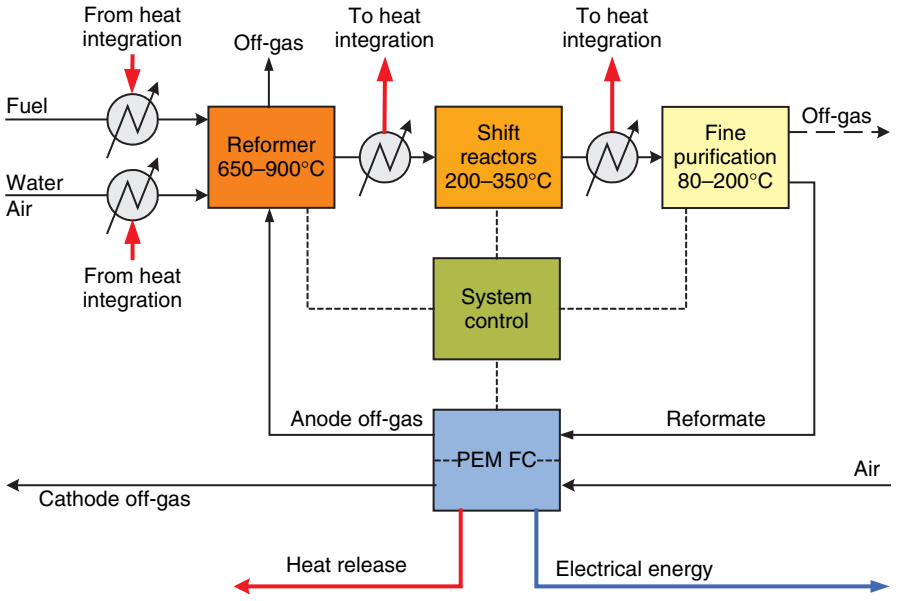


Figure 4 Process flow diagram of a reformer system with PEM fuel cell.

along with traces of methane and is, therefore, combusted in the reformer to increase system efficiency.

The entire reformer/fuel cell system is controlled by a control system that controls operating parameters in the reformer system and the fuel cell. System control for such a system is very complex and is the objective of numerous R&D activities.

Microreformers for hydrogen supply to PEMFCs for portable electronic devices and off-grid power supply are well investigated and have been developed at numerous academic and industrial research centers. They are viewed as replacement for batteries.

Typical applications range from military soldier power supply, where high energy densities are of foremost importance (Palo *et al.* 2002, Patil *et al.* 2004), to portable electronic devices such as cell phones, handhelds, or laptop computers (Hallmark *et al.* 2006, Kawamura *et al.* 2006, Kundu *et al.* 2006). Auxiliary power units (APU) are also designed based on microreforming technologies (Edlund 2005). The PEMFC is the fuel cell of choice for these applications due to low temperatures and fast dynamic operation. However, it requires hydrogen supply from a microreformer.

Some authors predict a significant potential for high energy densities. Palo *et al.* (2002) predict methanol SRF/PEMFC systems to have the potential to reach $1500\,\mathrm{W\,h\,kg^{-1}}$ by 2005 and up to $3000\,\mathrm{W\,h\,kg^{-1}}$ by 2008 for $15\,\mathrm{W_e}$ power output. However, these numbers have to be taken with care, because they always depend on the tank size, or in other words on the amount of stored energy.

3.19.4 Introduction to Reforming Technologies

Ahmed and Krumpelt (2001) give a brief overview of fuel processing for hydrogen generation. The conversion of fuels to hydrogen is carried out by one of the three major techniques – steam reforming (SR), partial oxidation (POX) reforming, and autothermal reforming (ATR).

SRF is probably the most common method for producing hydrogen in the chemical process industry:

$$\mathrm{fuel}\left(\mathrm{C}_n\mathrm{H}_m\mathrm{O}_p\right) + \mathrm{steam} \rightarrow \mathrm{carbon\ oxides} + \mathrm{hydrogen},\ \Delta H_\mathrm{R} > 0 \qquad [\mathrm{I}]$$

In this process, steam reacts with the fuel in the presence of a catalyst at elevated temperatures producing hydrogen, carbon monoxide, and carbon dioxide. Steam reformers are well suited for long periods of steady-state operation and can deliver relatively high concentrations of hydrogen (>70% on a dry basis). The primary SRF reaction is strongly endothermic, and the reactor performance is typically limited by heat transfer rather than by reaction kinetics.

POX reformers process the fuel with a substoichiometric amount of oxygen. The oxidation reaction yields heat generation causing high temperatures of the product gas:

$$\mathrm{fuel}\left(\mathrm{C}_n\mathrm{H}_m\mathrm{O}_p\right) + \mathrm{air} \rightarrow \mathrm{carbon\ oxides} + \mathrm{hydrogen} + \mathrm{nitrogen},\ \Delta H_\mathrm{R} < 0 \qquad [\mathrm{II}]$$

The heat generated from the oxidation reaction increases the gas temperature to above 1000°C. The oxidation step may be conducted with or without a catalyst.

Autothermal reformers combine the heat effects of the POX and the SRF reactions by feeding the fuel, water, and air together into the reactor:

$$\mathrm{fuel}\left(\mathrm{C}_n\mathrm{H}_m\mathrm{O}_p\right) + \mathrm{air} + \mathrm{steam} \rightarrow \mathrm{carbon\ oxides} + \mathrm{hydrogen} + \mathrm{nitrogen},\ \Delta H_R \approx 0 \qquad [\mathrm{III}]$$

This process is carried out in the presence of a catalyst, which controls the reaction pathways and thereby determines the relative extents of the oxidation reforming and SRF reactions. The SRF reaction absorbs part of the heat generated by the oxidation reaction, limiting the maximum temperature in the reactor. The net result is a process that is slightly exothermic to compensate for heat losses. In order to achieve the desired conversion and product selectivity, an appropriate catalyst is essential. **Figure 5** below gives a simplified overview of the three reforming technologies presented.

The three different technologies result in different product gases, as shown in **Table 1**. Depending on the downstream energy conversion device (fuel cell, internal combustion engine, turbine, and burner), the dilution of product gases with nitrogen is not favored, for the fuel cell then has a higher gas throughput to achieve the same electrical output. Therefore, the advantage of the faster start-up in ATR over SRF is negated, so most of the microreformer fuel cell systems apply SRF technology.

Of the three reforming pathways, the SRF process yields the highest hydrogen concentration in the

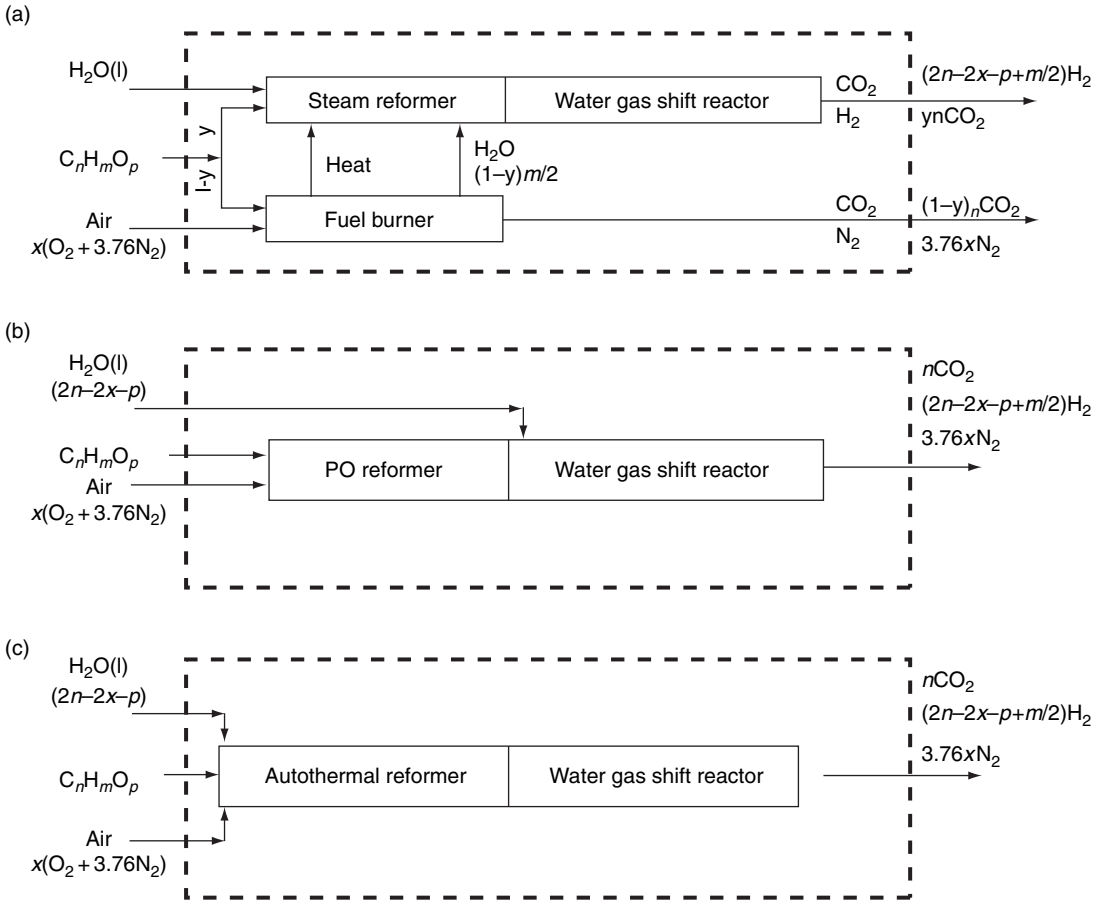


Figure 5 Schematics of idealized fuel processors, showing feeds and products from (a) steam reforming (SRF), (b) partial oxidation (POX) reforming, and (c) autothermal reforming (ATR). (Source: Ahmed S, Krumpelt M 2001 Hydrogen from hydrocarbon fuels for fuel cells. *Int. J. Hydrogen Energy* **26**, 291–301.)

Table 1 Comparison of dry product gases (vol.%) from the three different processes

Gas	*SRF*	*POX with air*	*ATR*
H_2	75–78	24–30	32–35
CO	8–10	14–18	9–11
CO_2	10–13	0.6–2	8–10
CH_4	2–5		0.2–1
N_2	–	43–52	44–48

product. However, for practical applications, POX and ATR processes are more attractive, according to Ahmed and Krumpelt, because they possess the potential to be more energy-efficient and the hardware can be smaller and lighter (Ahmed and Krumpelt 2001). In ATR technology, no extra burner is needed theoretically but then the anode off-gas from the fuel cell needs to be treated in an after burner to prevent combustible gases from being set free.

The carbon monoxide and, if necessary, carbon dioxide are removed from the reformate gas by a variety of reactions and scrubbing techniques, such as the water gas shift (WGS) reaction:

$$CO + H_2O \leftrightarrow CO_2 + H_2, \quad \Delta H_R < 0 \qquad [IV]$$

methanation:

$$CO + 3H_2 \rightarrow CH_4 + H_2O, \quad \Delta H_R < 0 \qquad [V]$$

PROX:

$$CO + 1/2\,O_2 \rightarrow CO_2, \quad \Delta H_R < 0 \qquad [VI]$$

and CO_2 absorption in amine solutions, and pressure swing adsorption.

Apart from reforming technologies, there is also the pyrolysis process in which no oxygen is provided

to the fuel, so it splits directly into hydrogen and carbon when heated to temperatures above 800 °C.

$$C_xH_y \Leftrightarrow xC + \frac{y}{2}H_2, \quad \Delta H_R < 0 \qquad [VII]$$

Up to now pyrolysis is considered to be economically feasible for large-scale plants only. The disadvantages lie in the formation of carbon, which has to be removed in a batch process, and also the efficiency, which lies below that of SRF (Ledjeff-Hey *et al.* 2000). A disadvantage in propane pyrolysis is that 30–50% of the energy content of the fuel are discarded with the carbon.

To choose the appropriate hydrogen generation process, the pros and cons in size, product gas purity, manufacturing process with respect to the size, weight, and complexity of the system, power density, total system efficiency, and cost need to be evaluated. For example, methanol reformer PEM fuel cell systems can reach higher power densities than DMFC stacks. Thus, building larger PEM fuel cell stacks with a greater electrical power output and still achieving acceptable costs instead of using a larger DMFC is favored for systems in which power density is the key criterion (Edlund 2005).

3.19.4.1 Differences Between the Micro- and Macroworlds

In this overview the word microsystem is used for a miniaturized integrated system with integrated sensors, controllers, and actors whose characteristic dimensions are on the micrometer scale (Kasper 1999). The word macrosystem is used for systems with characteristic dimensions >1 cm. A general comparison between micro- and macroworlds is given for instance by Hessel *et al.* (2005).

Advantages due to smaller size. With decreasing dimensions the gradients of physical parameters such as temperature, pressure, and density, which represent the driving force for heat and mass transfer and diffusion, increase. This results in improved heat and mass transfer, as evidenced in higher heat transfer coefficients in micro heat exchangers and shorter mixing times in micromixers. Thus, in microapplications with exothermal reactions, local hotspots are avoided through improved heat transfer.

Larger surface area-to-volume ratio. The maximum surface area-to-volume ratio in macroreactors from the production industry is $100\,m^2\,m^{-3}$, whereas in microchannels, values as high as $50\,000\,m^2\,m^{-3}$ can be achieved. This is an advantage for all catalytic reactions, for the active area increases noticeably. The change in surface area-to-volume ratio results in a limit for small-dimensional combustion (Peterson 2003). This is due to the surface area decreasing to the square and the volume decreasing to the cube with decreasing length. Fast thermal losses of a forming nucleus of combustion require a minimum surface area-to-volume ratio. These correlations are also found in nature in the limits in the size of warm-blooded animals. On the other hand, the increase of surface area-to-volume ratio also leads to an increase of heat losses (Kawamura 2006).

Analogies in the micro- and macroworlds. To calculate the behavior of gases and liquids in microchannels, the Navier–Stokes equations seem to be applicable. Calculated friction coefficients in laminar flows show good accordance with those of the macroworld. Small deviations can result from measurement errors that occur when experiments are carried out in this small scale (Amador *et al.* 2004).

Generally speaking, it can be stated that there will almost always be laminar flow conditions in microchannels. Although in the microworld the critical Reynolds numbers for laminar flow lie between 200 and 900, i.e., lower than in the macroworld, occurring Reynolds numbers mostly lie far below this limit. The critical Reynolds number changes due to a higher sensitivity towards surface roughness (Ducrée 2005).

Differences between the micro- and macroworlds. When placing a solid-state body in a cooler or a hotter environment, it cools down or heats up, resulting in an inner temperature gradient, represented by the dimensionless Biot number:

$$Bi = \frac{\alpha L}{\lambda} \qquad [1]$$

which describes the ratio of heat transfer α on the outer surface to inner heat conduction λ over a certain characteristic length L.

As L decreases, the Biot number decreases. For $Bi << 1$ the internal temperature gradient becomes very small, so homogeneous temperatures can be assumed. Therefore, smaller reactors are less affected by thermal tension and can tolerate faster temperature changes (Madou 1997).

3.19.5 Microchannel Reformer

Most conventional microreformers are manufactured using conventional welding techniques, the catalysts being inserted as pellets, honeycombs, or sometimes directly applied onto the reactor surface of heat exchangers, if they are used as reactors. These microreformers produce power in the range of 1–100 W_e. Another concept is the use of microchannel reactors for microreformers. In these reformers, dimensions of flow channels and other reactor internals, such as packed beds of catalyst particles, flow distributors, and mixing devices, are on the order of hundreds of micrometers or even below. They are called microchannel reformers. This overview of microreformers concentrates on microchannel reformers.

Numerous developments of small-scale reformers with internal microstructures are reported in the literature. Key players in microchannel reforming research and development are companies such as Motorola Labs (USA), Samsung (Korea), Matsushita Electric Works (Japan), Casio (Japan), research institutes such as Battelle (USA), Korea Institute of Energy Research (Korea), IMM GmbH (Germany), FZK-IMVT (Germany), EPFL (Switzerland), and universities such as University of Michigan (Dept. Chem. Eng.), Lehigh University (USA), Dalian Institute of Chemical Physics (China), Yonsei University (Korea), Korea University (Korea), Sungkyunkwan University (Korea), Seoul National University (Korea), Kyoto University (Japan), Tohoku University (Japan), Kogakuin University (Japan). This list is by no means comprehensive.

The published papers reflect a fraction of the activities going on in academia and especially in industry in the field of microreformers. A good overview of the literature on microchannel reformer development up to 2002 is given by Gariilidis *et al.* (2002). More current developments are summarized in a book by Hessel *et al.* (2005) where the authors describe various kinds of microchannel reformers, gas cleanup concepts, catalytic combustor design, and fully integrated systems.

Microchannel reformer capacities range from several hundred milliwatts (**Figure 6**) to one hundred watts (**Figure 7**) based on the electric power output a fuel cell could generate with the reformer product gas flow. With parallel installation of several microchannel reformers, capacities can reach the several kilowatt range (Tonkovich *et al.* 2004). Fuels commonly used in microreformers are liquid fuels such as methanol, ethanol, gasoline, isooctane, diesel, diesel surrogates, and gaseous fuels like LPG, and natural gas (methane).

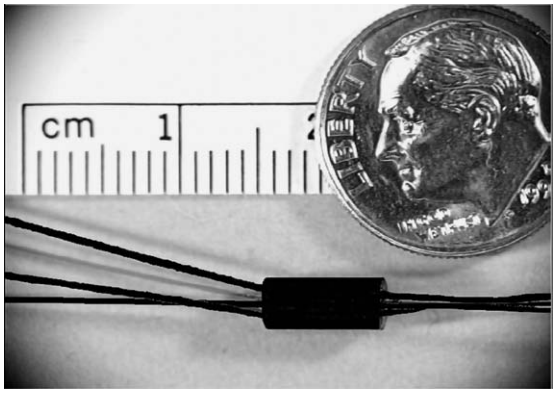


Figure 6 Battelle's 100 mW_e fully integrated methanol steam reformer including vaporizer, and catalytic combustor. (Source: Holladay J D, Jones E O, Phelps M, Hu J 2002 *J. Power Sources* **108**, 21–7.)

Over the past five years, the majority of research and development teams have concentrated on steam reformers for methanol for several reasons:

(1) Methanol can be reformed at relatively low temperatures between 250 and 350°C, which allows the use of silicon and ceramics as reactor material
(2) It reduces the requirements for insulation.
(3) Due to the low reforming temperature the carbon monoxide content of a methanol steam reformer product gas is on the order of 0.5–2 vol.%. Therefore, a subsequent shift reactor is not required.

In case of using PBI-membrane PEMFC that operates at elevated temperatures, no further gas cleanup is necessary, because they tolerate carbon monoxide concentrations up to 2 vol.%. In the case of standard low-temperature PEMFC, the carbon monoxide content has to be reduced down to 20 ppm by PROX, methanation, or Pd membranes. A few researchers report work on microchannel reforming using ATR or CPOX and on fuels other than methanol (Aartun *et al.* 2004, Hu *et al.* 2003, Kolb *et al.* 2007, Ryi *et al.* 2005, 2006, Tadd *et al.* 2005, Thormann *et al.* 2006, Tonkovich *et al.* 2004).

Aside from the key components of any reformer system, the actual reformer reactor, one or several WGS reactors, a carbon monoxide-reducing step (methanation, PROX, or Pd membrane), a vaporizer, a catalytic combustor, and various peripheral components, e.g., pumps, valves, pressure/temperature sensors, electrical heaters, and system controls, are required.

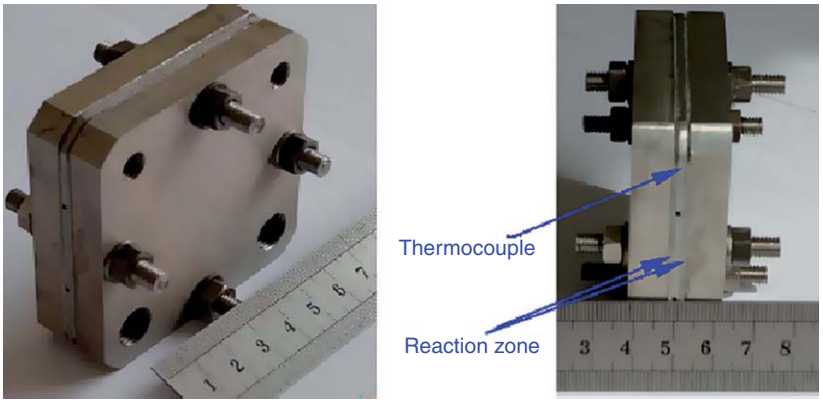


Figure 7 Microchannel methanol autothermal reformer (100 W_e) by Chen and coworkers. Combustor and vaporizer are not integrated. (Source: Chen G, Li S, Yuan Q 2007 *Catal. Today* **120**, 63–70.)

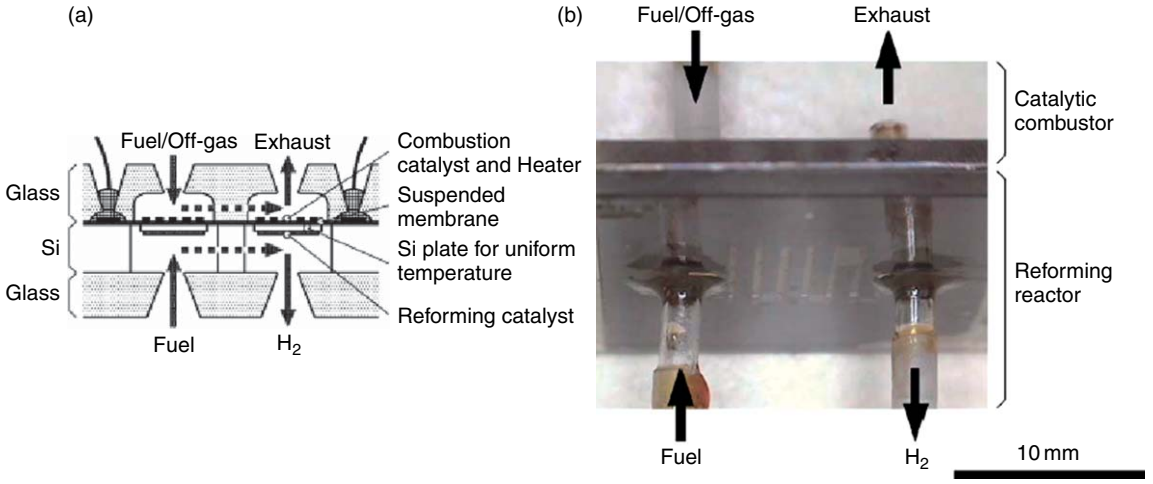


Figure 8 Microchannel reactor (a) with integrated catalytic combustor (b). (Source: Tanaka S, Chang K S, Min K B, Satoh D, Yoshida K, Esashi M 2004 MEMS-based components of a miniature fuel cell/fuel reformer system. *Chem. Eng. J.* **101**, 143–9.)

Most researchers focus on the design, fabrication, and testing of the microchannel reactors only, with integrated catalytic burners (**Figure 8**) or electric heaters (**Figure 9**) for heat supply to vaporizer and endothermic SRF reactor. A few teams, mainly with a more industrial background, develop complete integrated reformer systems with heat exchangers, preheaters, vaporizers, and catalytic combustors (Hallmark *et al.* 2006, Holladay *et al.* 2002, Holladay et al. 2004a, b, Kwon *et al.* 2006, 2007, Shin *et al.* 2006) – see **Figures 6** and **10**.

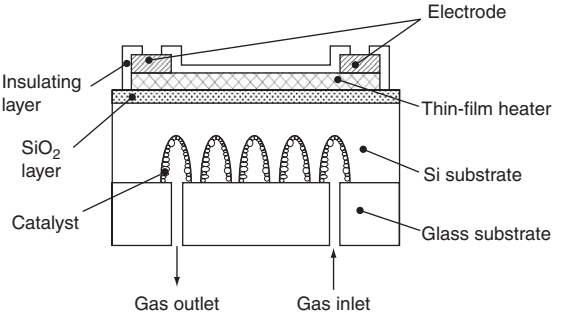


Figure 9 Microchannel reformer with electric film heater. (Source: Kawamura Y, Ogura N, Yamamoto T, Igarashi A 2006 *Chem. Eng. Sci.* **61**, 1092–101.)

3.19.5.1 Overview

Table 2 gives on overview of the developments reported in most recent papers. It provides

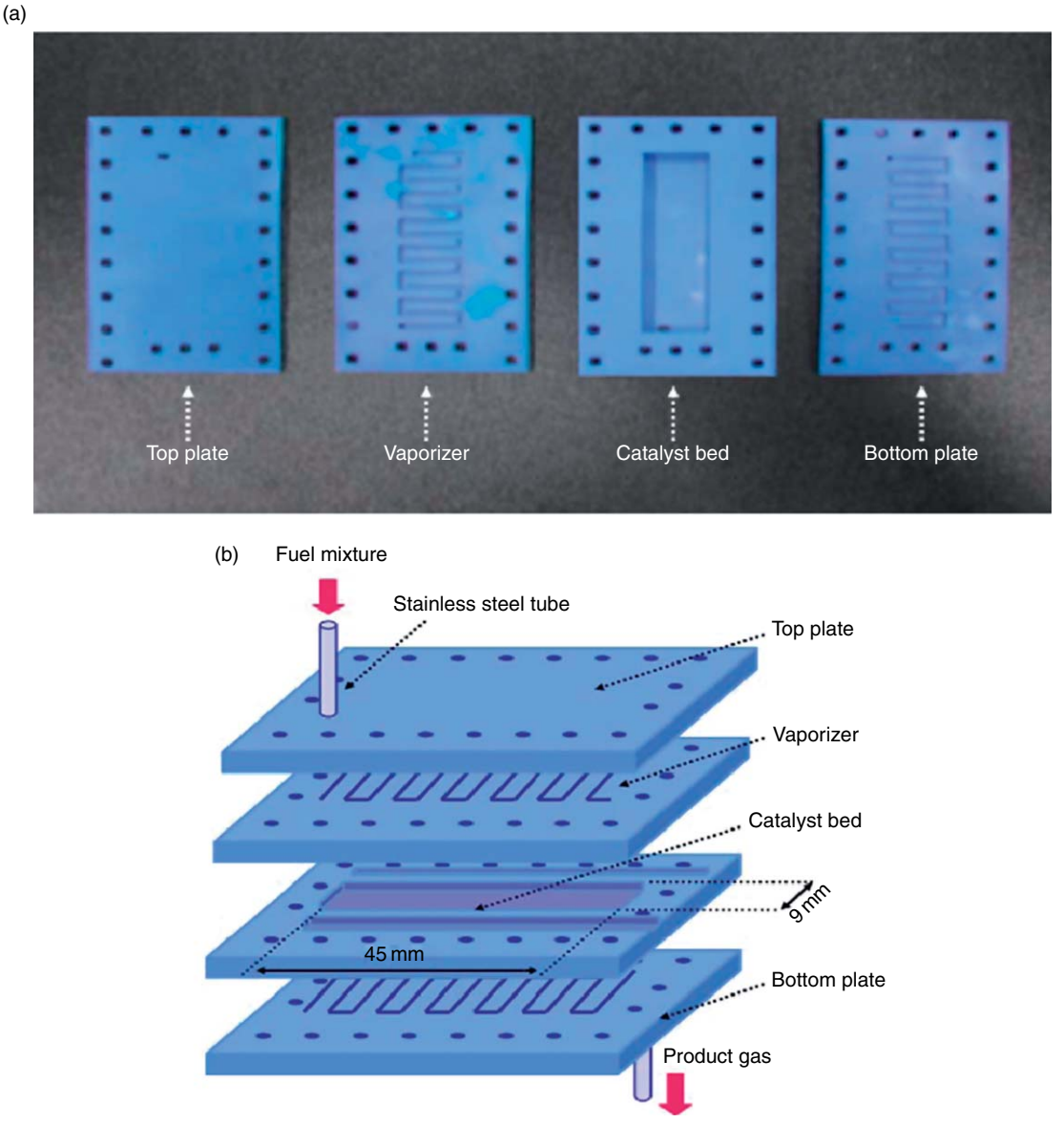


Figure 10 Methanol microchannel reformer with integrated vaporizer and electrical heaters. (Source: Shin Y, Kim O, Hong J-C, Oh J-H, Kim W-J, Haam S, Chung C-H 2006 *Int. J. Hydrogen Energy* **31**, 1925–33.)

information about the power output of the reformer, the fuel used, the reforming process (SRF, ATR, and CPOX), the type of gas cleanup (if any), peripheral components, reactor material (metal, silicon, and ceramic), and the source of the catalyst (proprietary development or commercial supplier). At the bottom of the table, references that deal with gas cleanup only, such as WGS, PROX, methanation (meth.), and Pd membranes (Pd) are cited. A question mark denotes that the particular information is not provided in the reference.

3.19.5.2 Catalyst and Fabrication

Many groups develop proprietary catalysts for reforming, PROX, WGS, and methanation. They all deal with the coating of reactor walls (adhesion problems) and catalyst distribution on the wash coat. Commercial catalyst suppliers are BASF AG, Süd-Chemie AG, and Johnson Matthey Plc.

A few studies have been published that include comparisons of packed bed and microchannels for microreforming reactors (Kundu *et al.* 2006, Park *et al.* 2005, Tadd *et al.* 2005) and PROX (Srinivas *et al.*

Table 2 Summary of most recent papers on microchannel reformers

	Power	*Fuel*	*Reforming process*	*Gas cleanup*	*Peripheral components*				*Material*	*Catalyst development*	
					Catalytic combustor	*Electrical heaters*	*Vaporizer*	*Heat exchanger*			
Reformer											
Holladay *et al*. (2002)	100 MW_e	MeOH			X	–	X	X	SS	Prop	
Palo *et al*. (2002)	15 W_e	MeOH	SRF	–	X	–	X		SS	Prop	
Hu *et al*. (2003)		Diesel/ gasoline surrogates	SRF	–	X	–	X	X	SS	Prop	
Aartun *et al*. (2004)	?	C3	ATR, CPOX		–	–			FeCrA	Prop	
Bravo *et al*. (2004)	–	MeOH	SRF	–	–				Silica cap.	Prop	
Holladay *et al*. (2004a, b)	600 mW_e	MeOH	SRF	Meth.	X	–	X		SS	Prop	
Presting *et al*. (2004)	–	?	SRF	Pd					Por Si		
Reuse *et al*. (2004)	?	MeOH	SRF	–	X	–			FeCrA	Comm	
Tanaka *et al*. (2004)	200 mW	MeOH	SRF	–	X	–		Eject.	Si	?	
Tonkovich *et al*. (2004)	?	CH_4	SRF	–	X	–			Incon.	Prop	
Park *et al*. (2004, 2005)	15 W_e	MeOH	SRF	–	X	–			SS	Comm	P. bed
Pattekar and Kothare (2004, 2005)	10–20 W_e	MeOH	SRF	–	–				Si	Comm	P. bed
Tadd *et al*. (2005)	?	IsoC8	ATR	–	–	–			FeCrA	Prop	P. bed
Ryi *et al*. (2005, 2006)	26 W_e	CH_4	SRF	–	X	–			Incon.	Prop	
Conant *et al*. (2006)	–	MeOH	SRF	–	–				Silica cap.	Comm	
Ehrich *et al*. (2006)		EtOH	SRF	–					SS	Prop	
Hallmark *et al*. (2006)	1–5 W_e	MeOH	SRF	(PBI)			X		LTCC	prop	
Kawamura *et al*. (2006)	1 W_e	MeOH	SRF	–		X			Si	Prop	

Kim and Kwon (2006a, b)		MeOH	SRF	–	–					Glass	Prop		
Kothare (2006)	8–10 W_e	MeOH	SRF	WGS, Pd					T sens	Si	?		
Kundu *et al*. (2006)	2.4 W_e	MeOH	SRF	–		X	X			Si	Comm	P. bed	
Park *et al*. (2006)	0.1 W_e	MeOH	SRF	Pd	X					Si	Comm		
Shin *et al*. (2006)		MeOH	SRF	PROX		X	X			LTCC	Comm	P. bed	
Thormann *et al*. (2006)	?	Diesel/ gasoline surrogate	SRF	–						SS	Prop		
Yoshida *et al*. (2006)	10 W	MeOH	SRF	–	X		X			Si	Comm		
Chen *et al*. (2002, 2007)	100 W_e	MeOH	ATR	–	–	–				SS	Prop		
Kwon *et al*. (2006, 2007)	20 W_e	MeOH	SRF	PROX		X				Si	Comm	P. bed	
Kolb *et al*. (2007)	Up to 5 kW	MeOH, diesel, LPG, isoC8	SRF, ATR	WGS, PROX	X	–	X	X		SS	Prop, comm		
Kundu *et al*. (2007)	0.5 W_e	MeOH	SRF	–	–	X	X			SS	Comm		
Gas cleanup													
Karnik *et al*. (2003)	–	–	–	WGS, Pd							Prop		
Mukherjee *et al*. (2007)	–	–	–	WGS	–					Glass	Prop		
Görke *et al*. (2005)	–	–	–	Meth.	–					SS	Prop		
Delsman *et al*. (2004)	–	–	–	PROX	–					SS	Prop		
Chen *et al*. (2004)	–	–	–	PROX	–					SS	Prop		
Srinivas *et al*. (2004)	–	–	–	PROX	–					Si	Prop	P. bed	

Reforming process: SRF, steam reforming; ATR, autothermal reforming; CPOX, catalytic partial oxidation. Gas cleanup: WGS, water gas shift; meth., methanation; Pd, palladium membrane. Periphery: eject., ejector; T sens, temperature sensor on chip. Material: SS, stainless steel; Incon., Inconel; FeCrA., FeCr alloy; Si, silicon; LTCC, low-temperature cofired ceramic; silica cap., silica capillaries; por Si, porous silicon. Catalyst development: prop, proprietary; comm., commercial; p. bed, packed bed.

2004). Packed beds supposedly have high-pressure drops, even despite new designs that lead to radial geometries (Pattekar and Kothare 2005). However, other sources do not state significant differences between both reactor concepts, as reported, for example, for the PROX reaction by Srinivas *et al.* (2004). Holladay *et al.* (2002) investigated different supports for methanol SRF catalysts. They looked into foams from pure metals, alloys, or ceramics. Kundu and coworkers used a packed powder (packed bed) in a microchannel reactor, as can be seen in **Figure 11**.

Palladium membrane fabrication in microchannel reactors is described by Karnik *et al.* (2003), Presting *et al.* (2004), and Kothare (2006). These membranes are used for separation of hydrogen from a reformate gas, with the copper support acting as a WGS catalyst.

For SRF of methanol, Cu/ZnO on an alumina substrate (Al_2O_3) is commonly used and is described

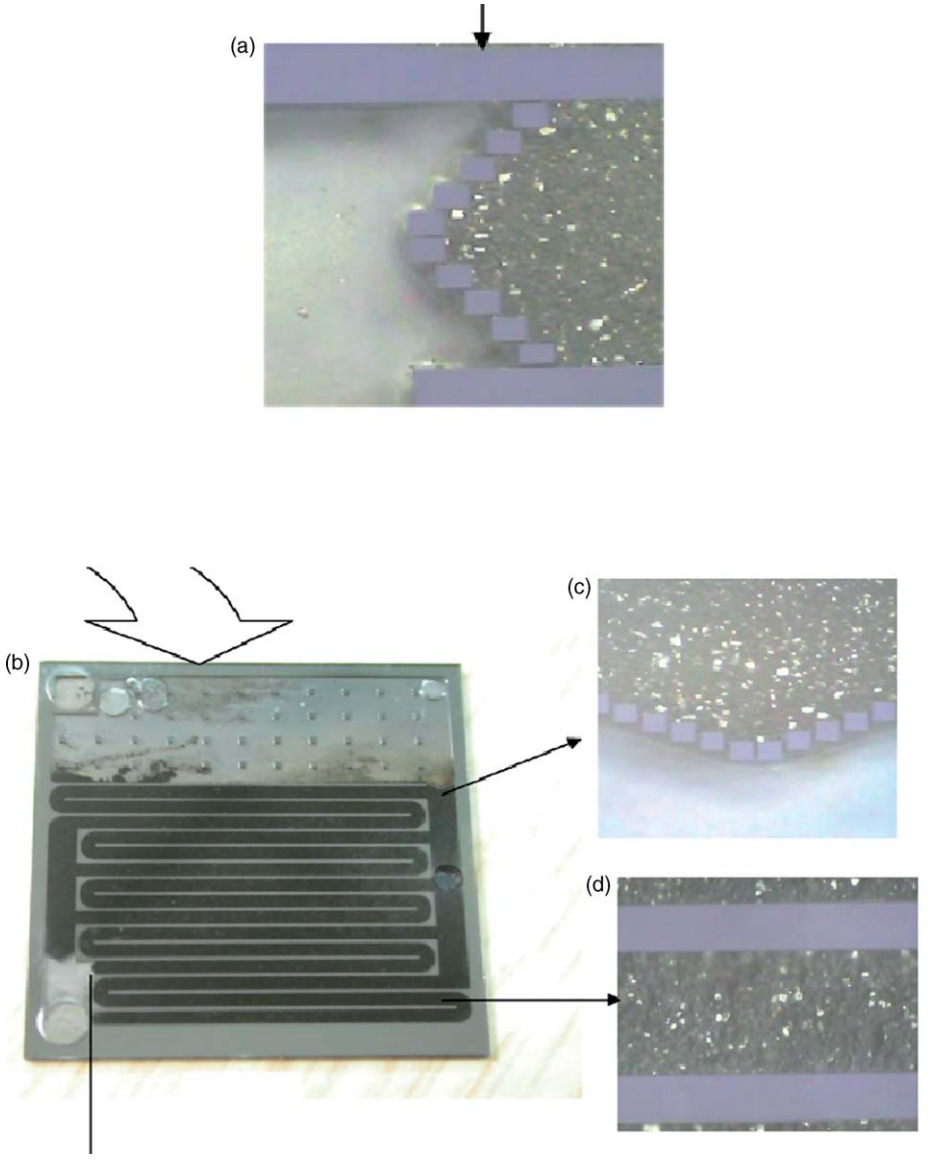


Figure 11 Packed powder in microchannel reactor. (Source: Kundu A, Jang J H, Lee H R, Kim S H, Gil J H, Jung C R, Oh Y S 2006 *J. Power Sources* **162**, 572–8.) (a) Photograph of serpentine patterned silicon wafer with catalyst. (b) Microscopic imaging of filter at the inlet point of the reactor part. (c) Microscopic image of catalyst packing inside the channels. (d) Microscopic image of filter at the outlet of the reactor part.

as a good catalyst with high activity and high carbon dioxide selectivity. Unfortunately, the catalyst does not tolerate hot spots, liquid water, and exposure to air while warm. All of these restrictions pose difficulties in the design, fabrication, and operation of reformers with that catalyst. According to Hallmark *et al.* (2006) the temperature limitations of the Co/ZnO-based catalyst probably pose the greatest challenges, because manufacturing of metal and ceramic reactor typically requires at least one high-temperature step (welding, brazing or diffusion bonding for metal, and cofiring for ceramic). However, these high temperatures significantly reduce the activity of the Cu/ZnO catalyst. Therefore, the catalyst coating of the reactor must take place after the high-temperature fabrication step for this catalyst. This is achieved by coating the reactor walls after reactor fabrication and impregnating the wash coat with catalyst solution afterwards. Depending on the wall material, wash coating faces severe adhesion problems. In this context, Hallmark *et al.* (2006) teamed up with BASF AG to develop a novel palladium catalyst that can withstand the high temperatures of the cofiring production step. This allows the catalyst to be applied to green tape prior to cofiring and, thereby, facilitates the production process.

3.19.5.3 Microchannel Design

Typical dimensions of microchannels reported are on the order of 100–500 μm wide, 100–300 μm deep, and 20–60 mm long. Many times the reformer and gas cleanup reactors are integrated in plate heat exchangers with counter current, cocurrent, and cross flow. Also, single–flow, meandric-shaped channels are investigated (Kawamura *et al.* 2006, Shin *et al.* 2006). Double meanders of reformer and vaporizer with catalytic combustor on the backside of the plate have been reported as well (Park *et al.* 2006). For more fundamental investigations some authors use electric heaters on the back side of the patterned microchannel plates for defined heat input (Kawamura *et al.* 2006, Kundu *et al.* 2006, 2007, Kwon *et al.* 2006, 2007, Shin *et al.* 2006, Tadd *et al.* 2005). **Figure 9** shows a typical example.

Special precautions are taken to achieve even flow distribution over a large number of parallel microchannels. This results in inlet ports (**Figure 12** shows two typical examples), as described by the majority of authors, and staggered rod bundles (Srinivas *et al.* 2004).

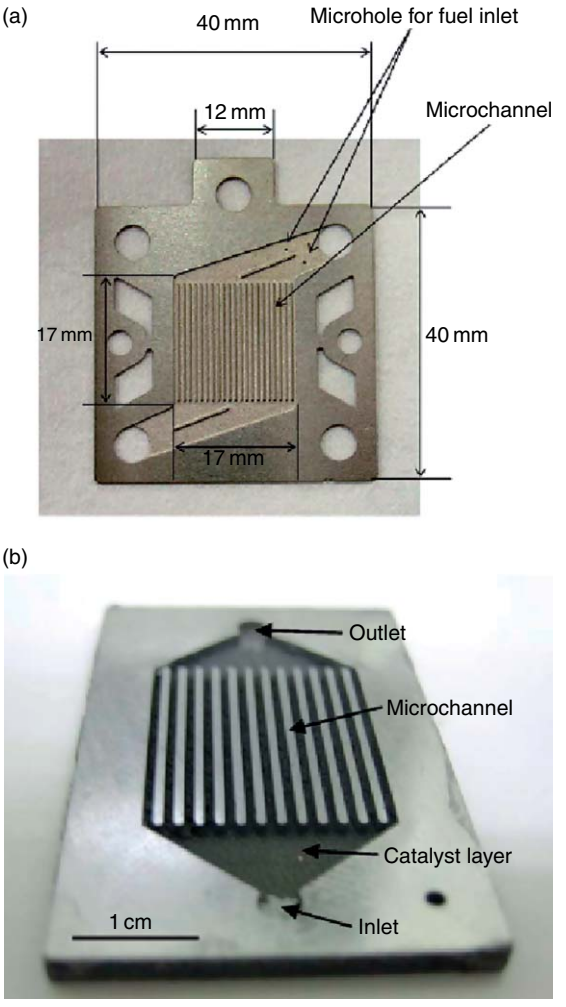


Figure 12 Two examples for an inlet port design to achieve even flow distribution to parallel microchannels. (a) (Source: Ryi S K, Park J S, Choi S H, Cho S H, Kim S H 2005 *Chem. Eng. J.* **113**, 47–53.) (b) The microchannel width is 500 μm and reactor chamber size is 1.5 cm × 2.0 cm × 0.1 cm. (Sources: Kim T, Kwon S 2006a *Chem. Eng. J.* **123**, 93–102; Kim T, Kwon S 2006b *Micromech. Microeng.* **16**, 1760–68.)

Some authors have worked with microtubes, reactors filled with catalyst powder (packed bed) (Shin *et al.* 2006), or foams with microcavities instead of microchannels (Bravo *et al.* 2004, Conant *et al.* 2006, Presting *et al.* 2004). One team even designed a radial flow reactor to minimize pressure drop in the packed bed for the volume increasing the reforming reaction (Pattekar and Kothare 2004, 2005). Simulation tools are used to predict flow regimes inside the microchannels and temperature distribution throughout the microchannel reactors (Kawamura *et al.* 2006, Mukherjee *et al.* 2007).

3.19.5.4 Materials and Fabrication

Materials typically used for microchannel reactors are metal foils and plates (stainless steel, Inconel, FeCrAlloy), silicon in combination with Pyrex glass, and ceramic (low-temperature cofired ceramic, LTCC). Different materials require different microreactor fabrication procedures.

3.19.5.4.1 Metal

(1) In metal sheets, microchannels can be created by wet chemical etching all the way through and then placing manifold sheets between the microchannel sheets. Finally, cover sheets at the bottom and the top allow tight assembly by use of screws (Park *et al.* 2004, 2005). This is depicted schematically in **Figure 13**.

(2) Another way is to create microchannels in metal plates by wet chemical etching, only partly into the metal material. The microchannel sheets are then placed between end plates and screwed together (Chen *et al.* 2002, 2007, Kundu *et al.* 2007, Ryi *et al.* 2005, 2006) (**Figure 14**).

(3) Some authors use wet chemical etching of microchannels in metal foils and subsequent laser

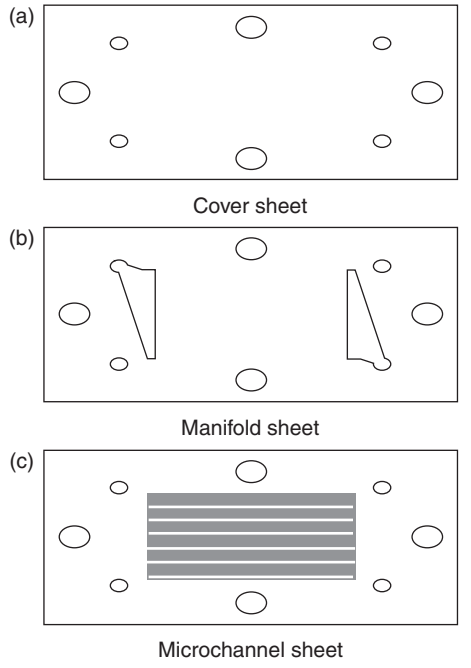


Figure 13 Microchannels created by wet etching all the way through metal sheets. (Source: Park G-G, Seo D J, Park S-H, Yoon Y-G, Kim C-S, Yoon W-L 2004 *Chem. Eng. J.* **101**, 87–92.)

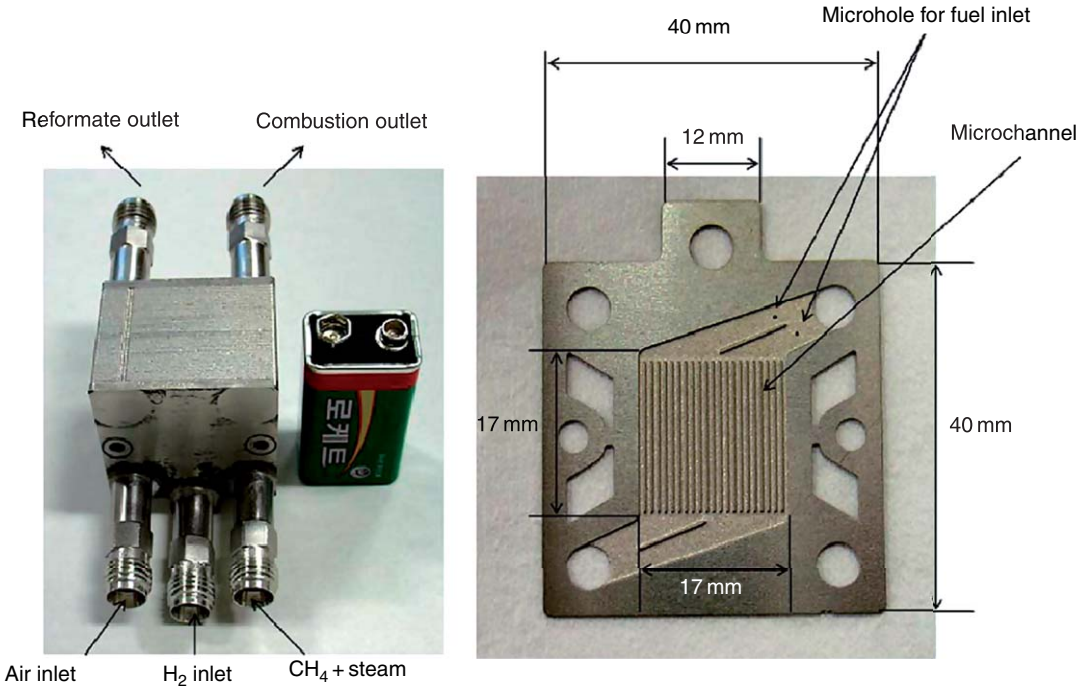


Figure 14 Example for microchannel plates and their assembly. (Source: Ryi S K, Park J S, Choi S H, Cho S H, Kim S H 2005 *Chem. Eng. J.* **113**, 47–53.)

welding as sealing technique (Kolb *et al.* 2007, Tonkovich *et al.* 2004).

(4) Other groups fabricate microchannel plates as described in (2) or by micromachining, then the plates are stacked together in a housing, and closed hermetically, for instance, by use of clamps (Aartun *et al.* 2004, Görke *et al.* 2005, Reuse *et al.* 2004) (**Figure 15**).

(5) Another approach is to use microreactor structures (for instance, foams) surrounded by a gasket and placed inside a metal gasket retainer sheet. This assembly is placed between separation sheets, with cover sheets on top and bottom. The whole array is finally screwed together (Tadd *et al.* 2005).

3.19.5.4.2 Silicon

(1) Microchannels in silicon wafer can be created by standard photolithography followed by deep reactive ion etching (DRIE). The channeled side of the wafer is then bonded with Pyrex glass by anodic bonding (Kawamura *et al.* 2006, Kundu *et al.* 2006, Park *et al.* 2006, Pattekar and Kothare 2005, Srinivas *et al.* 2004, Tanaka *et al.* 2004).

(2) Another method is wet etching of microchannels into a silicon wafer by KOH solution and subsequent anodic bonding with Pyrex glass (Kwon *et al.* 2006).

3.19.5.4.3 Ceramic (LTCC)

(1) Green tape is processed, with the microchannel structure created by mechanical punching, stacked, and laminated under elevated temperature and pressure (110–120 °C, 137–150 bar); LTCC stack can be sintered on a quartz glass plate to maintain uniform shape (Shin *et al.* 2006).

(2) According to Hallmark *et al.* (2006), a novel catalyst paste is applied to green tape using standard filling or screen printing techniques without any change to the ceramic tape process used in high-volume commercial foundries. In addition,

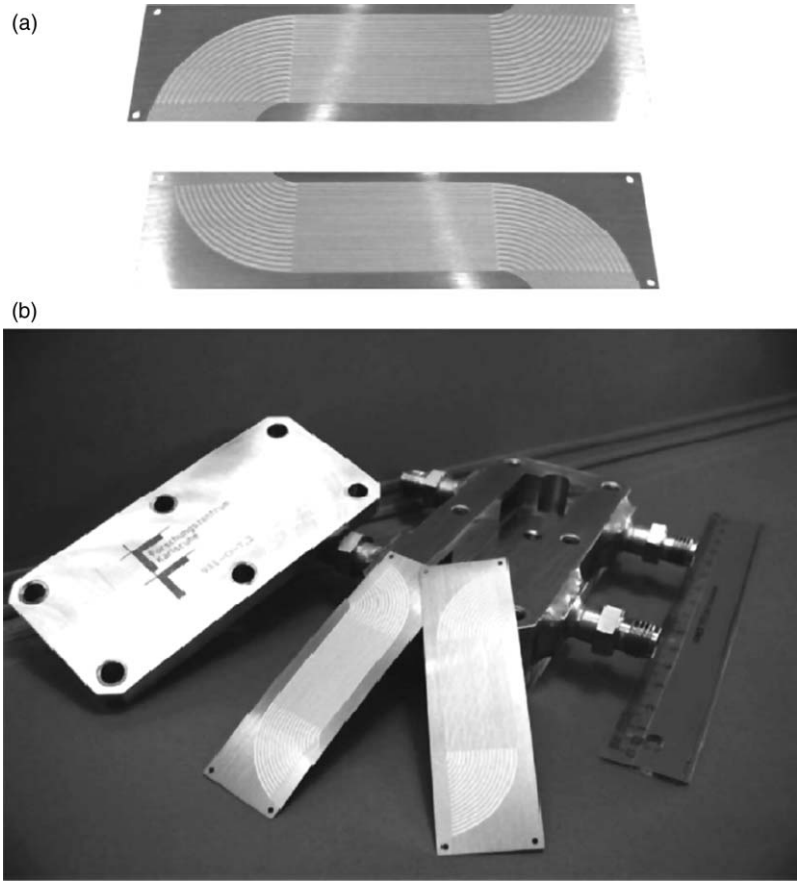


Figure 15 (a) Microstructured plates. (b) Housing of the two-passage microstructured reactor. (Source: Reuse P, Renken A, Haas-Santo K, Görke O, Schubert K 2004 *Chem. Eng. J.* **101**, 133–41.)

this process eliminates the coating fraction limitations inherent in slurry coating (Hallmark *et al.* 2006).

3.19.5.5 Balance of Plant

The balance of plant components were developed by many researchers during the development of microchannel reformers: catalytic burners (Park *et al.* 2005, Ryi *et al.* 2005), microheat exchangers (Ryi *et al.* 2006), evaporators (Yoshida *et al.* 2006), and microejectors (Tanaka *et al.* 2004). Also system insulation is of importance (Kawamura *et al.* 2006, Tanaka *et al.* 2004). Furthermore, system control is a topic that a few researchers address. According to Kothare (2006), new hardware architecture and high-performance control algorithms that can be embedded and integrated with microchemical systems need to be developed.

References

Aartun I, Gjervan T, Venik H, Görke O, Pfeifer P, Fathi M, Holmen A, Schubert K 2004 Catalytic conversion of propane to hydrogen in microstructures reactors. *Chem. Eng. J.* **101**, 93–9

Ahmed S, Krumpelt M 2001 Hydrogen from hydrocarbon fuels for fuel cells. *Int. J. Hydrogen Energy* **26**, 291–301

Amador C, Gavriilidis P, Angeli P 2004 Flow distribution in different microreactor scale-out geometries and the effect of manufacturing tolerances and channel blockage. *Chem. Eng. J.* **101**, 379–90

Bravo J, Karim A, Conant T, Lopez G P, Datye A 2004 Wall coating of a $CuO/ZnO/Al_2O_3$ methanol steam reforming catalyst for micro-channel reformers. *Chem. Eng. J.* **101**, 113–21

Chen G, Yuan Q, Li A 2002 Microchannel reactors for methanol autothermal reforming reaction. *Chin. J. Catal.* **23**, 1–2

Chen G, Yuan Q, Li H, Li S 2004 CO selective oxidation in microchannel reactor for PEM fuel cell. *Chem. Eng. J.* **101**, 101–6

Chen G, Li S, Yuan Q 2007 Pd-Zn/Cu-Zn-Al catalysts prepared for methanol oxidation reforming in microchannel reactors. *Catal. Today* **120**, 63–70

Conant T, Karim A, Rogers S, Samms S, Randolph G, Datye A 2006 Wall coating behavior of catalyst slurries in non-porous ceramic microstructures. *Chem. Eng. Sci.* **61**, 5678–85

Coutanceau C, Koffi R K, Léger J-M, Marestin K, Mercier R, Nayoze C, Capron P 2006 Development of materials for mini DMFC working at room temperature for portable applications. *J. Power Sources* **160**, 334–9

Delsman E R, de Croon M H J M, Kramer G J, Cobden P D, Hofmann Ch, Cominos V, Schouten J C 2004 Experiments and modelling of an integrated preferential oxidation-heat exchanger microdevice. *Chem. Eng. J.* **101**, 123–31

Ducrée J 2005 Centrifugal microfluidics for lab-on-a-chip and micro process engineering. Ph.D. thesis, University of Freiburg

Edlund D 2005 Development of a compact 250 W portable power supply for commercial and military markets. *2005 Fuel Cell Seminar*, Palm Springs, CA, USA, November 14–18, 2005, Abstract 554

Ehrich H, Kolb G, Jähnisch K 2006 Hydrogen generation in micro-structured reactors by reformation of bioethanol. *Abstracts IMRET9*, Potsdam/Berlin, GermanySeptember 6–8, 2006, pp. 228–9

Gavriilidis A, Angeli P, Cao E, Yeong K, Wan Y 2002 Technology and applications of microengineered reactors. *Trans. Inst. Chem. Engrs* **80(Part A)**, 3–30

Görke O, Pfeifer P, Schubert K 2005 Highly selective methanation by the use of a microchannel reactor. *Catal. Today* **110**, 132–9

Guo Z, Fagrhi A 2006 Miniature DMFCs with passive thermal-fluids management system. *J, Power Sources* **160**, 1142–55

Hallmark J, Samms S, Castellano C, Liu Y 2006 Development of a ceramic wall-coated methanol fuel processor utilizing novel co-fireable steam-reformation catalyst. *Abstr. 2006 Fuel Cell Seminar*, Honolulu, HI, USANovember 13–17, 2006, pp. 67–70

Hebling C, Yu H M 2005 Fuel cells: Micro systems. *Encycloped. Mater. Sci. Technol.*, 1–13

Hessel V, Löwe H, Müller A, Kolb G 2005 *Chemical Micro Process Engineering*. Wiley-VCH, Weinheim, Germany

Holladay J D, Jones E O, Phelps M, Hu J 2002 Microfuel processor for use in a miniature power supply. *J. Power Sources* **108**, 21–7

Holladay J D, Jones E O, Dagle R A, Xia G G, Cao C, Wang Y 2004a High efficiency and low carbon monoxide micro-scale methanol processors. *J. Power Sources* **131**, 69–72

Holladay J D, Wainright J S, Jones E O, Gano S R 2004b Power generation using a mesoscale fuel cell integrated with a microscale fuel processor. *J. Power Sources* **130**, 111–18

Hu J, Wang Y, VaderWiel D, Chin C, Palo D, Rozmiarek R, Dagle R, Cao J, Holladay J, Baker E 2003 Fuel processing for portable power applications. *Chem. Eng. J.* **93**, 55–60

Karnik S V, Hatalis M K, Kothare M V 2003 Palladium-based micro-membrane for water gas shift reaction and hydrogen gas separation. *J. Microelectromech. Syst.* **12(1)**, 93–100

Kasper M 1999 *Mikrosystementwurf*. Springer Verlag, Berlin, p. 5

Kawamura Y, Ogura N, Yamamoto T, Igarashi A 2006 A miniaturized methanol reformer with Si-based microreactor for a small PEMFC. *Chem. Eng. Sci.* **61**, 1092–101

Kelley S C, Deluga G A, Smyrl W H 2002 Miniature fuel cells fabricated on silicon substrates. *Mater. Interf. Electrochem. Phenom.* **48**, 1071–82

Kim T, Kwon S 2006a Catalyst preparation for fabrication of a MEMS fuel reformer. *Chem. Eng. J.* **123**, 93–102

Kim T, Kwon S 2006b *Micromech. Microeng.* **16**, 1760–68

Kolb G, Schürer J, Tiemann D, Wichert M, Zapf R, Hessel V, Löwe H 2007 Fuel processing in integrated micro-structured heat-exchanger reactors. *J. Power Sources*, doi:10.1016/j.jpowsour.2007.01.006

Kothare M V 2006 Dynamics and control of integrated microchemical systems with application to micro-scale fuel processing. *Comput. Chem. Eng.* **30**, 1725–34

Kundu A, Jang J H, Lee H R, Kim S H, Gil J H, Jung C R, Oh Y S 2006 MEMS-based micro-fuel processor for application in a cell phone. *J. Power Sources* **162**, 572–8

Kundu A, Park J M, Ahn J E, Park S S, Shul Y G, Han H S 2007 Micro-channel reactor for steam reforming of methanol. *Fuel* **86**, 1331–6

Kwon O J, Hwang S-M, Ahn J-G., Kim J J 2006 Silicon-based miniaturized-reformer for portable fuel cell applications. *J. Power Sources* **156**, 253–9

Kwon O J, Hwang S-M, Chae J H, Kang M S, Kim J J 2007 Performance of a miniaturized silicon reformer-PrOx-fuel cell system. *J. Power Sources* **165**, 342–6

Ledjeff-Hey K, Kalk T, Mahlendorf F, Niemzig O, Trautmann A, Roes J 2000 Portable PEFC generator with propane as fuel. *J. Power Sources* **86**, 166–72

Litterst C, Eccarius S, Hebling C, Zengerle R, Koltay P 2006 Increasing μDMFC efficiency by passive CO_2 bubble removal and discontinuous operation. *J. Micromech. Microeng.* **16**, S248–53

Liu X, Suo C, Zhang Y, Wang X, Sun C, Li L, Zhang L 2006 Novel modification of Naflon® 117 for a MEMS-based micro direct methanol fuel cell (μDMFC). *J. Micromech. Microeng.* **16**, S226–32

Madou M 1997 *Fundamentals of Microfabrication*. CRC Press, New York

Modroukas D, Modi V, Fréchette L G 2005 Micromachined silicon structures for free-convection PEM fuel cells. *J. Micromech. Microeng.* **15**, S193–201

Mukherjee S, Hatalis M K, Kothare M V 2007 Water gas shift reaction in a glass microreactor. *Catal. Today* **120**, 107–20

Palo D R, Holladay J D, Rozmiarek R T, Guzman-Leong C E, Wang Y, Chin Y-H, Dagle R A, Baker E G 2002 Development of a soldier-portable fuel cell system; Part I: a bread-board methanol fuel processor. *J. Power Sources* **108**, 28–34

Park D-E, Kim T, Kwon S, Kim C-K, Yoon E 2007 Micromachined methanol steam reforming system as a hydrogen supplier for portable proton exchange membrane fuel cells *Sens. Actuators A*: *Physical* 135(1), 58–66

Park G-G, Seo D J, Park S-H, Yoon Y-G, Kim C-S, Yoon W-L 2004 Development of microchannel methanol steam reformer. *Chem. Eng. J.* **101**, 87–92

Park G-G, Yim S-D, Yoon Y-G, Kim C-S, Seo D-J, Eguchi K 2005 Hydrogen production with integrated microchannel fuel processor using methanol for portable fuel cell systems. *Catal. Today* **110**, 108–13

Patil A S, Dubois T G, Sifer N, Bostic E, Gardner K, Quah M, Bolton C 2004 Portable fuel cell systems for America's army: technology transition to the field. *J. Power Sources* **136**, 220–5

Pattekar A V, Kothare M V 2004 A microreactor for hydrogen production in micro fuel cell applications. *J. Microelectromech. Syst.* **13(1)**, 7–18

Pattekar A V, Kothare M V 2005 A radial microfluidic fuel processor. *J. Power Sources* **147**, 116–27

Peterson R B 2003 Miniature and microscale energy systems. In: Fahgri M, Sunden B (eds.) *Heat and Fluid Flow in Microscale and Nanoscale Structures*. WIT Press, Southampton, pp. 1–37

Presting H, Konle J, Starkov V, Vyatkin A, König U 2004 Porous silicon for micro-sized fuel cell reformer units. *Mater. Sci. Eng. B* **108**, 162–5

Reuse P, Renken A, Haas-Santo K, Görke O, Schubert K 2004 Hydrogen production for fuel cell application in an autothermal micro-channel reactor. *Chem. Eng. J.* **101**, 133–41

Ryi S K, Park J S, Choi S H, Cho S H, Kim S H 2005 Novel micro fuel processor for PEMFCs with heat generation by catalytic combustion. *Chem. Eng. J.* **113**, 47–53

Ryi S K, Park J S, Cho S H, Kim S H 2006 Fast start-up of micro-channel fuel processor integrated with an igniter for hydrogen combustion. *J. Power Sources* **161**, 1234–40

Shah K, Shin W C, Besser R S 2003 Novel microfabrication approaches for directly patterning PEM fuel cell membranes. *J. Power Sources* **123(2)**, 172–181

Shin Y, Kim O, Hong J-C, Oh J-H, Kim W-J, Haam S, Chung C-H 2006 The development of micro-fuel processor using low temperature co-fired ceramic (LTCC). *Int. J. Hydrogen Energy* **31**, 1925–33

Srinivas S, Dhingra A, Im H, Gulari E 2004 A scalable silicon microreactor for preferential oxidation: performance comparison with a tubular packed bed microreactor. *Appl. Catal. A Gen.* **274**, 285–93

Tadd A R, Gould B D, Schwank J W 2005 Packed-bad versus microreactor performance in autothermal reforming of isooctane. *Catal. Today* **110**, 68–75

Tanaka S, Chang K S, Min K B, Satoh D, Yoshida K, Esashi M 2004 MEMS-based components of a miniature fuel cell/fuel reformer system. *Chem. Eng. J.* **101**, 143–9

Thormann J, Pfeifer P, Schubert K, Kunz U 2006 Reforming of diesel fuel in micro reactor for APU systems. *Abstracts IMRET9*, Potsdam/Berlin, Germany, September 6–8, 2006, pp. 226–7

Tonkovich A Y, Perry S, Wang Y, Qiu D, LaPLante T, Rogers W A 2004 Microchannel process technology for compact methane steam reforming. *Chem. Eng. Sci.* **59**, 4819–28

Yu J, Cheng P, Ma Z, Yi B 2003 Fabrication of miniature silicon wafer fuel cells with improved performance. *J. Power Sources* **124**, 40–6

Ziegler C 2005 Modeling and simulation of the dynamic behavior of portable proton exchange membrane fuel cells. URN: urn:nbn:de:bsz:352-opus-17126\\

Züttel A 2003 Materials for hydrogen storage. *Mater. Today* **6(9)**, 24–33

Relevant Websites

www.frost.com :World Advances in Microfuel Cell Technology (Technical Insights) 2005

URL: http://www.ub.uni-konstanz.de/kops/volltexte/2006/1712/

Biographies

Thomas Aicher obtained his M.Sc. and Ph.D. in Chemical Engineering from University of Karlsruhe, Germany. His dissertation deals with free and forced convection heat transfer in tube bundle heat exchangers and was completed in 1997. For the next four-and-a-half years he worked for the American Engineering, Procurement and Contracting (EPC) company, Fluor Daniel, Inc. in Aliso Viejo, CA, USA. His assignments included process studies, as well as front end and detail engineering for refineries, power plants, and gas treatment plants. In the beginning of 2002 he joined Fraunhofer ISE in Freiburg, Germany, where he is head of the group Hydrogen Production and Storage. He has published 25 technical and scientific papers on heat transfer, drying technology, gas processing, and reforming technology.

Lisbeth Rochlitz studied Environmental Engineering and Process Engineering with a focus on energy technology and water and sewage-water technology at the Brandenburg Technical University in Cottbus, Germany, and the Beijing Institute of Technology, P.R. China. She obtained her M.Sc. in 2002, and in 2003 she started working at Fraunhofer ISE in Freiburg, Germany. From 2004 through 2006 she held a scholarship from the German Federal Environmental Foundation to carry out her Ph.D. research in the field of microreforming during which she gave several oral presentations at conferences.

Christopher Hebling studied physics at the University of Freiburg with a focus on elementary particle physics. The Ph.D. dissertation, carried out at the Fraunhofer Institute for Solar Energy Systems ISE, dealt with the simulation, realization and characterization of crystalline silicon thin-film solar cells on insulating substrates. The Ph.D. degree in physics was received with summa cum laude in 1998 from the department of physics at the University of Konstanz.

In January 1999 Christopher Hebling founded the Micro Energy Technology group at Fraunhofer ISE which is investigating small energy converters as a supplement or a substitute to batteries. The technologies are device-integrated high efficiency solar cell modules, organic solar cells as well as fuel cells in the low power range. The target market are portable and remotely located electronic appliances with a power demand from some mW to some 100 W.

Since 2001 he is head of the department 'Energy Technology' which is mainly involved in hydrogen generation technologies such as reforming and electrolysis as well as fuel cell technology. The department employs about 70 engineers, scientists and students. He has published about 25 reviewed scientific papers in the field of crystalline silicon solar cells and micro fuel cell systems. Dr. Christopher Hebling is 44 years old, is married and has three children.

3.20 Molecular Machine-Based NEMS

Tony Jun Huang, Department of Engineering Science and Mechanics, The Pennsylvania State University, University Park, PA, USA

3.20.1 Introduction

Machines are extensively used in our daily lives. The *Oxford English Dictionary* defines a machine as an apparatus for applying mechanical power, having several parts each with a definite function (Hawkins 1979). The concept of a machine can be extended to the molecular level. Although there are many chemical compounds whose structure and/or shape can be modified by an external stimulus, the term molecular machine will only be used for systems whose component parts undergo movements of relatively large amplitudes. Furthermore, systems in which the molecular movements are not controlled by some easily identifiable and well-characterized external stimulus will not be considered to be molecular-level machines. For example, due to thermal energy, molecules in solution are almost never stationary and are consequently not considered to be molecular machines because such thermal movement is not controllable. Therefore, a molecular machine can be defined as an assembly of a discrete number of molecular components designed to perform mechanical-like movements (output) as a consequence of appropriate external stimuli (input) (Balzani *et al.* 2000, 2003).

The concept of a machine at the molecular level is not new. Our bodies can be viewed as very complex ensembles of molecular-level machines (Boyer 2002, Goodsell 1996). Natural molecular machines are individual protein molecules that are ultimately responsible for essentially all active biological motions. Important examples of natural molecular machines are myosin, kinesin, and ATPase. They perform vital, life-sustaining tasks, some of which include muscle contraction, cell division, intracellular transport, and genomic transcription. The idea of constructing artificial molecular-level machines, however, is quite a recent one. In 1959, Richard Feynman, one of the greatest physicists and teachers of the 20th century, originated the idea of artificial molecular machines in his historic address to the American Physical Society, "There is plenty of room at the bottom" (Feynman 1960). The Nobel laureate contemplated, "What are the possibilities of constructing molecular-scale mechanical machines ··· What would be the utility of such machines? Who knows? I cannot see exactly what would happen, but I can hardly doubt that when we have some control of the arrangement of things on a molecular scale we will get an enormously greater range of possible properties that substances can have, and of the different things we can do." In the early 1980s, the earliest examples of synthetic molecular-level machines were reported and they were based on the

photoisomerization of azobenzene (Shinkai and Manabe 1984). Since then, the development of artificial molecular machines has become one of the most intriguing topics in the field of nanoscience and nanotechnology.

This chapter commences with a review of biological molecular machines and their applications in nanoelectromechanical systems (NEMS). Subsequently, various artificial molecular machines will be introduced. Next, the author will describe the research he and his colleagues have conducted in the past few years on moving artificial molecular machines from solution to solid supports to NEMS device settings. Finally, the chapter concludes with a perspective for future developments.

3.20.2 Biological Molecular Machines and Their Applications in NEMS

Biological molecular machines are molecular proteins that convert the chemical energy released by hydrolysis of adenosine 5′-triphosphate (ATP) into mechanical energy. Ubiquitous in biological systems, molecular machines are responsible for the directed transport of macromolecules, membranes, or chromosomes within the cytoplasm, thus playing a critical role in cell behavior and architecture (Boyer 2002, Goodsell 1996, Howard 1997, Vall and Milligan 2000). Over the past decade, interest in motor molecules has blossomed. Atomic structures of various motors have been solved by X-ray crystallography; new motors have been discovered; forces, steps, and speeds have been measured at the molecular level; hybrid devices integrating bottom-up-based biomolecular motors and top-down-based nanoengineered structures have been realized. These scientific breakthroughs have not only led to a much improved understanding of cell motility but have also profound impact in the field of nanoscience and nanotechnology.

3.20.2.1 Myosin and Kinesin

Enzymes such as myosin, kinesin, and their relatives are linear motors that convert the energy of ATP hydrolysis into mechanical work along polymer substrates – myosin along actin filaments in muscles and cells and kinesin along microtubules (**Figure 1**) (Cross and Duncan 1996, Howard *et al.* 1989, Kull *et al.* 1996). In a kinesin molecule, two heads, each about 8 nm long, are joined at the coiled-coil neck. The neck then connects to a long coiled coil that terminates in the tail region and binds to the organelle cargo. Myosin's heads are about twice the size of kinesin's and are composed of two major domains: the motor domain that binds to actin and nucleotides and the regulatory domain that is an 8-nm-long α-helix stabilized by two calmodulin-like light chains. A long coiled-coil rod connects to the regulatory domain and oligomerizes to form the thick filament in muscle.

The developments in DNA/RNA sequences have helped reveal the similarities between myosin and kinesin (Kull *et al.* 1996). The core 180 amino acids within the myosin and kinesin head structures have the same fold. Furthermore, the preserved

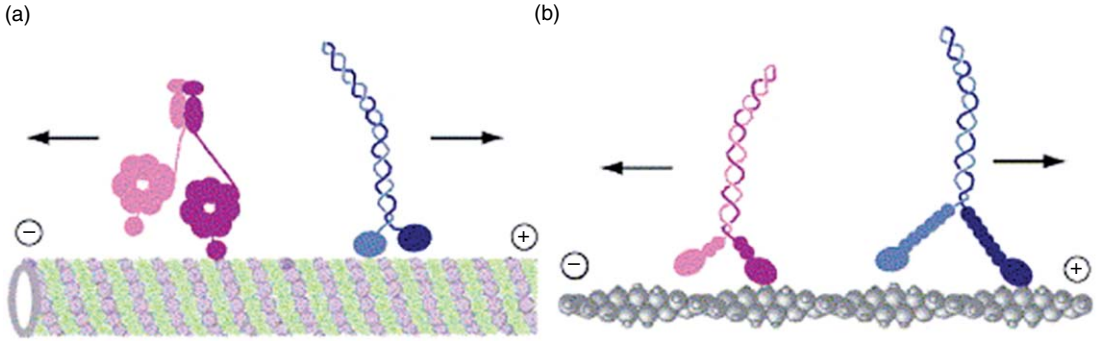


Figure 1 Linear motors that walk along cytoskeletal filaments: (a) kinesin and dynein motor that move to the plus and minus end, respectively, of a microtubule; and (b) myosin V and myosin VI that move to the plus (barbed) and minus (pointed) end, respectively, of an actin filament. (Source: Lipowsky R, Klumpp S 2005 Life is motion: Multiscale motility of molecular motors. *Physica A* **352**, 53–112.)

sequence direction and order of the α-helices and β-sheets within the core suggest that these two biomotors evolved from a common ancestor. Yet despite these structural similarities, there are profound functional differences between kinesin and myosin. Conventional kinesin operates alone or in small numbers to transport membrane-bound organelles over large distances (up to a millimeter) along microtubules, whereas muscle myosin operates in huge arrays of up to a billion molecules in a large muscle fiber and moves relatively short distances (up to about 1 μm) along actin filaments (**Figure 2**) (Hackney 1996, Lipowsky and Klumpp 2005). Myosin provides the power for all of our voluntary motions (running, walking, lifting, etc.) as well as for involuntary muscles (e.g., a beating heart).

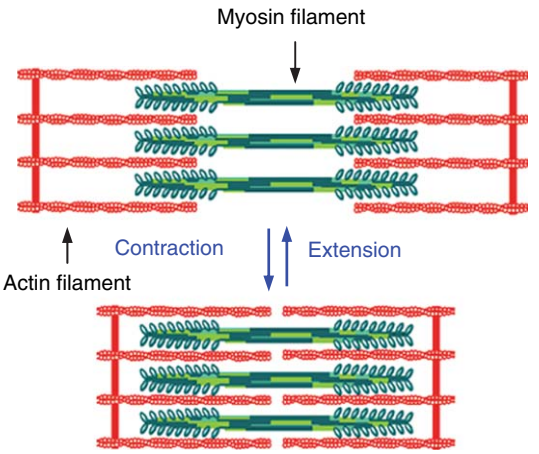


Figure 2 The sliding of myosin along actin filaments causes muscle contraction and extension.

Recent years have witnessed a strong interest in the exploration of biological molecular machines for NEMS applications. Limberis and Stewart (2000) demonstrated that kinesin can be engineered and used to transport silicon microchips along microtubules. Negatively charged microtubules bound electrostatically to a positively charged coverslip and were aligned in a microfluidic channel. Next, genetically modified kinesin motors attached with silicon microchips were added into the microfluidic channel. The microchips were observed to move in various ways (**Figure 3**), depending on how they were bound to microtubules. If a microchip moved along parallel microtubules, it traveled along linearly, with a maximum speed of approximately 0.8 μm s^{-1} (**Figure 3a**). When it interacted with antiparallel microtubules, it rotated (**Figure 3b**). A microchip with one end fixed flipped over (**Figure 3c**). This work proved that kinesin motors had enough force to transport cargo, such as silicon microchips that are much larger than kinesin molecules, thus paving the way for future development of biomotor-based NEMS.

To improve the efficiency of kinesin-based NEMS, the direction of the microtubules must be controlled and aligned. Hiratsuka *et al.* (2001) developed a lithographic process to restrict the microtubules onto fabricated tracks and control the direction of kinesin's mechanical motion. The experiment used two concentric circles with arrowhead patterns facing opposite directions (**Figure 4**). Without the arrowheads in the pattern, the microtubules would travel either clockwise or counterclockwise in either circle. However, once

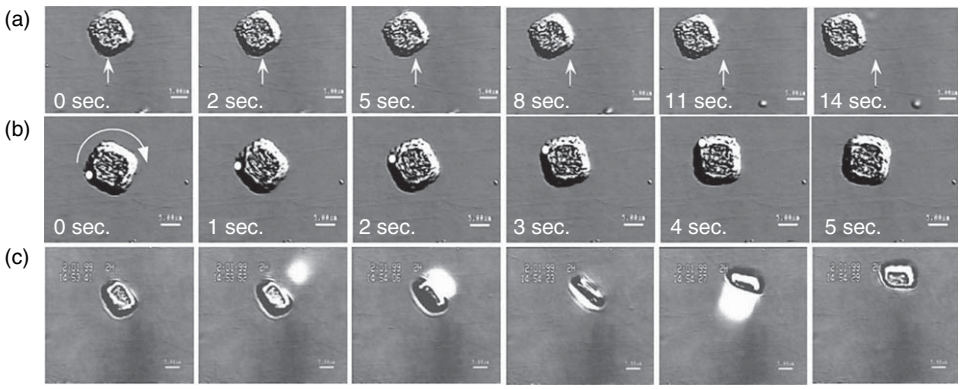


Figure 3 Kinesin-powered silicon microchips. Bar, 5 μm. (a) The chip moves on parallel microtubules at 0.8 μm s^{-1}. The arrow is a fixed reference point. (b) The chip rotates clockwise on antiparallel microtubules. A reference dot was placed on one corner of the chip. (c) The chip flips perpendicular to the surface. (Source: Limberis L, Stewart R J 2000 Toward kinesin-powered microdevices. *Nanotechnology* **11**, 47–51.)

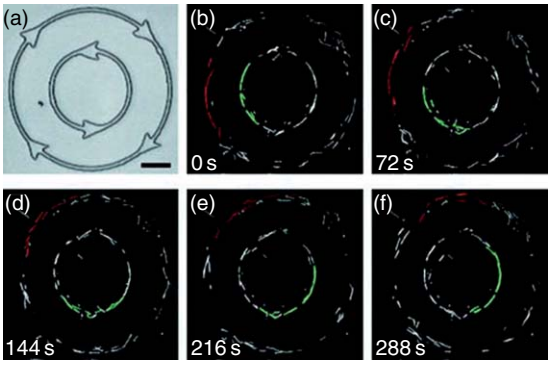


Figure 4 Unidirectional movement of microtubules along circular tracks. (a) Image of the arrowhead pattern. (b–f) Photographs demonstrating the microtubule rotational movement were taken at 72-s intervals beginning 20 min after the addition of adenosine 5′-triphosphate (ATP). Some filaments were dyed to facilitate the tracking of microtubules: the red microtubules are moving clockwise and the green microtubules are moving counterclockwise. (Source: Hiratsuka Y, Tada T, Oiwa K, Kanayama T, Uyeda T Q P 2001 Controlling the direction of kinesin-driven microtubule movements along microlitho-graphic tracks. *Biophys. J.* **81**, 1555–61.)

the pattern of arrowheads was added, within a few minutes the microtubules would travel in the same direction that the arrows pointed toward. As microtubules enter the arrowhead from the correct direction, they could easily pass through; however, if microtubules entered from the reverse direction, they would usually bump against the wall at the base of the arrow and turn around. Four different versions of the arrowhead pattern were tested; the most efficient pattern had a probability of a 73% chance of reversing microtubules in the wrong direction.

The alignment of microtubules for guiding the long-range transportation of kinesin motors has also been investigated. Platt *et al.* reported the alignment of magnetic nanoparticle functionalized microtubules in magnetic fields (**Figure 5**). In their study, microtubules were labeled with cobalt ferrite ($CoFe_2O_4$) nanoparticles through the biotin–neutravidin binding. When placed above a magnet, magnetized microtubules were aligned to match the orientation of magnetic line. It was also shown that microtubules could respond to the changing of magnetic orientation and reorient them within several seconds.

Recently, Van den Heuvel *et al.* (2006) reconstituted the kinesin–microtubule transport system in enclosed fluidic channels and realized active electrical control of the direction of individual kinesin-propelled microtubule filaments at Y-junctions. Using this technique, they demonstrated molecular sorting of differently labeled microtubules (**Figure 6**). The steering of microtubules is attributed to electric field-induced bending of the leading tip. From measurements of the orientation-dependent electrophoretic motion of individual, freely suspended microtubules, they estimate the net applied force on the tip to be in the range of piconewtons.

Transportation is not the only attribute of biological molecular machines. Hess *et al.* (2002) have developed a forcemeter assembled from kinesin and microtubules (**Figure 7**). One microtubule with

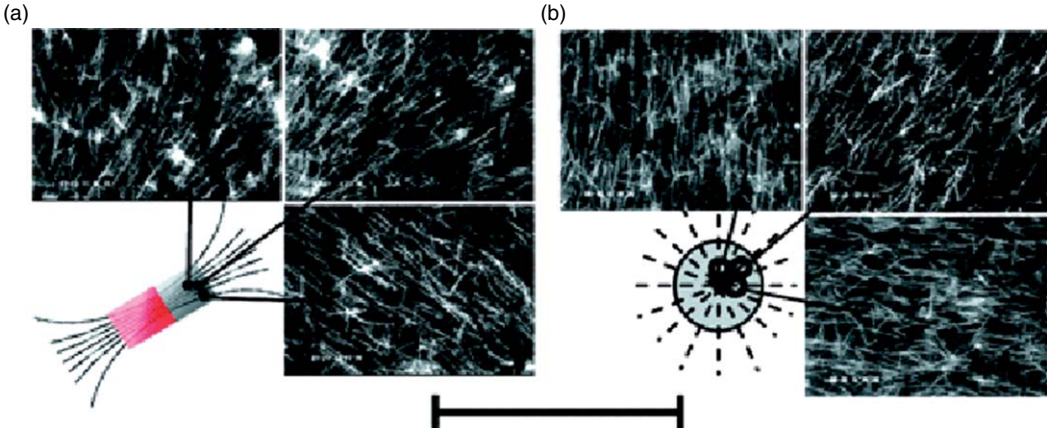


Figure 5 Fluorescence microscopy images of 37% biotinylated microtubules aligned with (a) a bar magnet for 3 min and (b) a cone magnet for 10 min. Scale bar = 116 μm. Also shown are the relative geometries of the magnets and magnetic field lines. (Source: Platt M, Muthukrishnan G, Hancock W O, Williams M E 2005 Millimeter scale alignment of magnetic nanoparticle functionalized microtubules in magnetic fields. *J. Am. Chem. Soc.* **127**, 15686–7.)

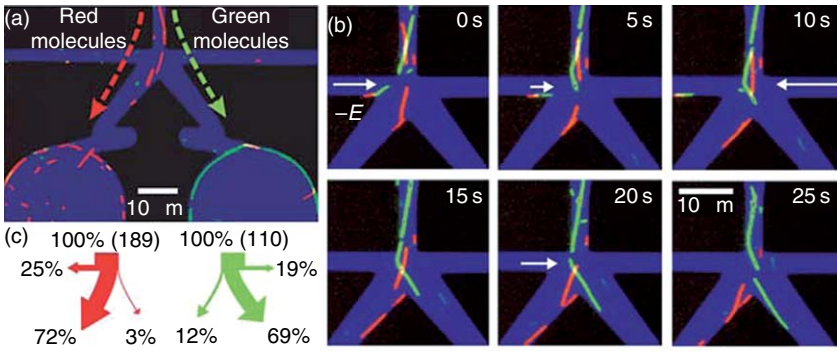


Figure 6 Demonstration of molecular sorting. (a) Color image of a mixture of red- and green-labeled microtubules approaching a Y-junction. Electrical force is used to steer microtubules carrying green and red fluorophores into the right and left reservoirs, respectively. (b) Example of successful sorting events for a green- and a red-labeled microtubule. (c) Steering efficiencies for 189 red and 110 green microtubules. (Source: Van den Heuvel M G L, De Graaff M P, Dekker C 2006 Molecular sorting by electrical steering of microtubules in kinesin-coated channels. *Science* **312**, 910–14.)

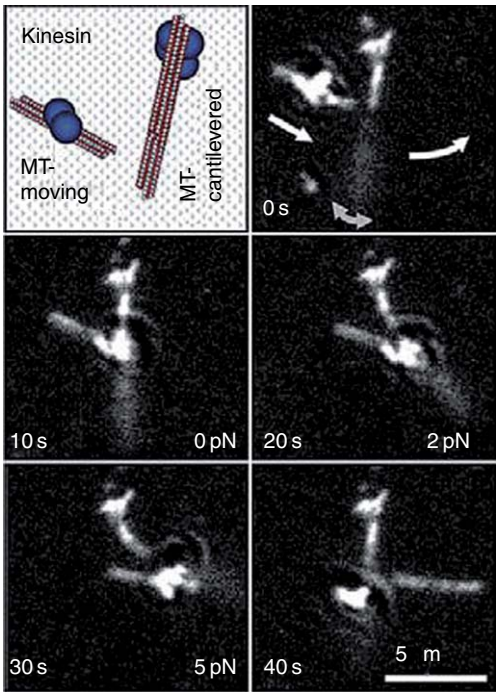


Figure 7 The cantilevered microtubule (MT) is attached on one end to a clump of beads, whereas the other end is suspended above the surface. Swiveling due to Brownian motion causes the smeared-out appearance of the free end. The moving microtubule is propelled by kinesin motor proteins adsorbed to the surface and carries a clump of beads. The snapshots show how the moving microtubule approaches (0 s), makes contact (10 s), bends (20 and 30 s), and releases (40 s) the cantilevered microtubule. (Source: Hess H, Howard J, Vogel V 2002 A piconewton forcemeter assembled from microtubules and kinesins. *Nano Lett.* **2**, 1113–15.)

known stiffness serves as a cantilever. Force is applied by another microtubule propelled by kinesin motor proteins adsorbed to a surface. The two microtubules are attached by a link whose strength is to be determined. The moving microtubule continues to travel while bending the molecular cantilever until the link ruptures. The force on the link can be calculated based on the bending of the cantilevered microtubule, the distance traveled by the second microtubule, and the angle between the two microtubules. Since this forcemeter measures forces at the range of piconewtons, it is ideal for measuring the strength of biological receptor/ligand pairs, such as streptavidin/biotin.

3.20.2.2 ATPase

ATP synthase (ATPase) is a ubiquitous protein that functions as a proton pump and the primary source of ATP for most eukaryotic cells (Allison 1998, Block 1997, Elston *et al.* 1998, Uptain *et al.* 1997, Yin *et al.* 1995). It manufactures ATP from ADP and phosphate by using the energy derived from a transmembrane protonmotive gradient. It can also reverse itself and hydrolyze ATP to pump protons against an electrochemical gradient. ATPase carries out both its synthetic and its hydrolytic cycles by a rotary mechanism. The general structure of an F_0F_1-ATPase is shown in **Figure 8**. It consists of two portions: a soluble component, F_1, containing the catalytic sites, and a transmembrane component, F_0, comprising the proton channel. The entire F_0F_1 structure is arranged as a counterrotating rotor and stator assembly. The stator portion consists of the

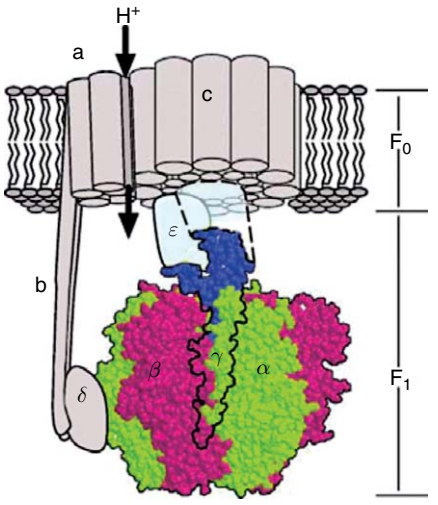


Figure 8 Schematic drawing of an F_0F_1-ATP synthase. (Source: Allison W S 1998 F_1-ATPase: A molecular motor that hydrolyzes ATP with sequential opening and closing of catalytic sites coupled to rotation of its subunit. *Acc. Chem. Res.* **31**, 819–26.)

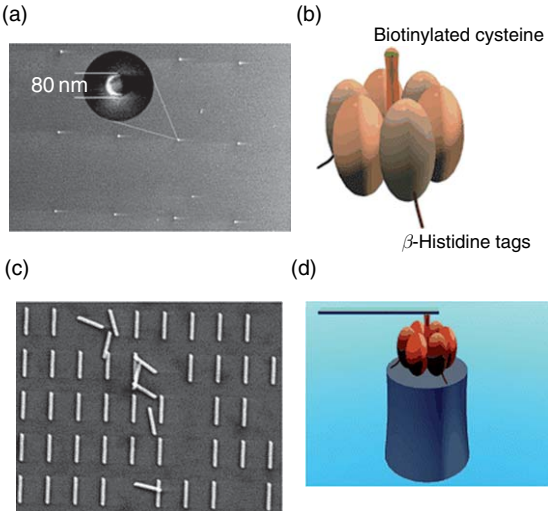


Figure 9 F_1-ATPase-powered nanomechanical device. The device consists of (a) nickel post (height 200 nm, diameter 80 nm), (b) F_1-ATPase motor, (c) nanopropellers (length 750–1400 nm, diameter 150 nm), and (d) assembled device. (Source: Soong R K, Bachand G D, Neves H P, Olkhovets A G, Craighead H G, Montemagno C D 2000 Powering an inorganic nanodevice with a biomolecular motor. *Science* **290**, 1555–8.)

catalytic sites contained in the $\alpha_3\beta_3$ hexamer, together with subunits a, b_2, and δ. The rotor consists of 9–12 c-subunits arranged in a ring and connected to the γ- and ε-subunits that form the central shaft.

Recently, the rotation of F_1-ATPase was directly observed during ATP hydrolysis by attaching a fluorescent actin filament to the γ-subunit and immobilizing the $\alpha_3\beta_3$ subunit on a coverslip (Noji *et al.* 1997). In another experiment (Soong *et al.* 2000), F_1-ATPase was integrated with inorganic nanoengineered systems to form hybrid organic–inorganic nanomechanical devices powered by biological molecular motors (**Figure 9**). The device consisted of three parts: (1) nanofabricated nickel posts; (2) recombinant F_1-ATPase motors that were engineered to adhere with the nanofabricated structures; and (3) engineered nanopropellers. To integrate these components into one functional unit, the F_1-ATPase biomolecular motors were biotinylated and then attached to the nickel posts using histidine tags engineered into the β-subunit. In addition, the nickel nanopropellers were linked to the γ-subunit using streptavidin/biotin. The rotation of the propellers was tested with and without ATP to prove the dependency on the F_1-ATPase motor. With the ATP, only five of the 400 propellers rotated continuously. Approximately 80% of the propellers did not move at all. It is thought that the propellers may have been attached to move than one motor or attached to the substrate in addition to the motor. Rotational velocity of the propellers varied depending on the point of attachment and the length of the propeller. Even though the device needs further improvement, it laid the foundation for the future development of NEMS using F_1-ATPase as the active nanostructure.

3.20.3 Artificial Molecular Machines

Biological molecular motors are intrinsically designed for the physiological conditions (37°C, pH 7.0) within the human body. These restrictions, in addition to the difficulties in modifying their structure and functionality, hinder integration with man-made devices. In this regard, the future of molecule-driven actuators lies in the development of artificial molecular motors as opposed to biomolecular motors. The ability to synthesize analogical molecular versions of macroscopic machinery is considered to be an essential step toward the construction of molecular devices capable of performing mechanical work and storing/transmitting information (Balzani *et al.* 2003). Like their macroscopic counterparts, artificial molecular machines are characterized by the (1) form of energy input enabling activation, (2) kind of movement executed by their components, (3) way in which their operation can be controlled and monitored, (4) possibility to repeat the

operation at will, (5) timescale needed to complete a cycle of operation, and (6) function performed. Recent advances in synthetic chemistry have yielded analogs of rotors, gears, switches, shuttles, turnstiles, and ratchets (Balzani *et al.* 2003). This section focuses on some of the most successful artificial molecular machines: rotaxanes, catenanes, and pseudorotaxanes.

Catenanes and rotaxanes along with their mechanically threaded supramolecular progenitors, pseudorotaxanes, are composed of mutually recognizable and intercommunicating ring and dumbbell-shaped components (Anelli *et al.* 1991, 1997, Ashton *et al.* 1991, Bissell *et al.* 1994, Brouwer *et al.* 2001, Martínez-Díaz *et al.* 1997, Raehm *et al.* 1999, Stoddart and Tseng 2002). When stimulated by light, electricity, or chemical reagents, mechanically interlocked molecules experience relative internal motions (rotary and linear), with their components behaving like the moving parts of macroscopic machines (**Figure 10**). These mechanically interlocked components recognize each other by means of donor/acceptor ([$\pi\cdots\pi$]) and additional ([C–H$\cdots$O] and [C–H$\cdots\pi$]) interactions that can be switched off by either oxidizing the donors or reducing the acceptors. The redox chemistry is usually reversible, and consequently, the interactions can be turned on again at will. Through proper selection of the redox-controllable recognition sites within the components, these interlocked molecules can be induced to switch chemically, electrochemically, or photochemically between two or more mechanically different states.

A rotaxane is a molecule composed of a macrocyclic and a dumbbell-shaped component (**Figure 10a**). The macrocycle encircles the linear rod-like portion of the dumbbell-shaped component and is trapped mechanically around it by two sizable stoppers. Thus, the two components cannot dissociate from one another, even though they are not covalently bonded to each other. The structural formula and graphical representation of a bistable rotaxane **R1** is shown in **Figure 11**. **R1** consists of a tetracationic $CBPQT^{4+}$ ring component and a linear rod section that contains two different stations: a tetrathiafulvaline (TTF) unit and a dioxynaphthalene (DNP) unit. The **R1**'s starting state has its TTF unit encircled by the positively charged ring. Upon oxidation, the TTF unit becomes dicationic (TTF^{2+} and experiences electrostatic repulsion with the ring, causing the ring to shuttle toward the DNP unit in the oxidized state. Conversely, the reduction of the TTF^{2+} back to a neutral TTF unit causes **R1** to return to its starting state. This reversible bistable rotaxane via ring translation can therefore be equated as a linear molecular motor.

By contrast, a catenane is a molecule composed of two interlocked macrocyclic components. The two macrocycles are not covalently bonded to each other, but rather mechanically linked, preventing their dissociation. The incorporation of two different recognition sites within the asymmetric macrocycle (**Figure 10b**) allows the symmetric macrocyle component to preferentially reside around one of these two positions. By switching on or off the properties at one of the two recognition sites of the asymmetric macrocycle, the relative position of the two species can be reversibly controlled. The complexation/decomplexation of metal ions or neutral organic molecules, protonation/deprotonation, and oxidation/reduction processes can all be exploited to reversibly alter the stereoelectronic properties of one

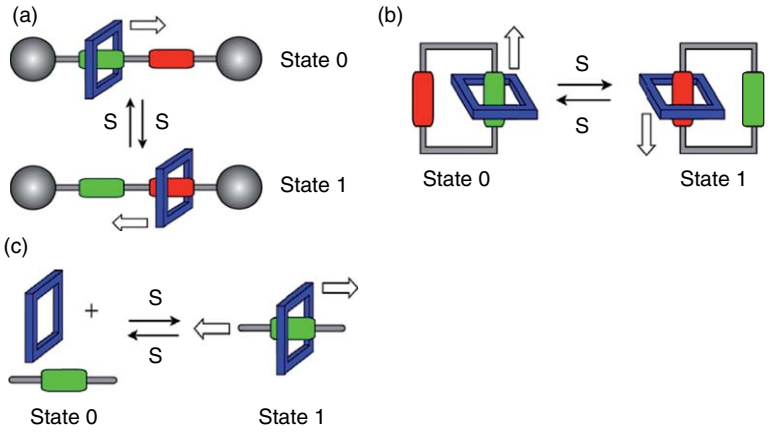


Figure 10 Schematic of a bistable (a) rotaxane, (b) catenane, and (c) pseudorotaxane.

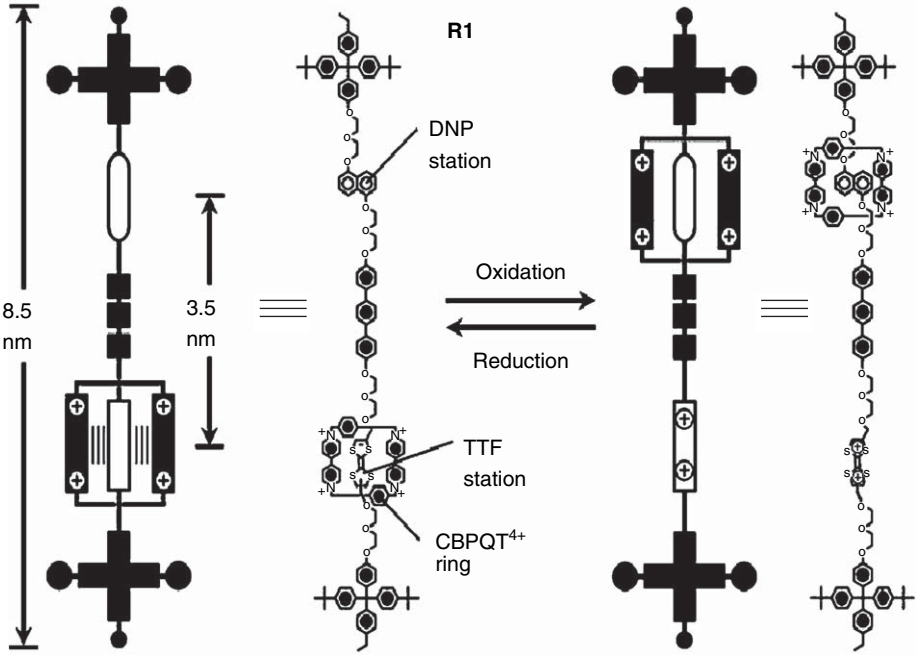


Figure 11 Molecular structure and schematic representations of a redox-controllable rotaxane molecular motor **R1**.

of the two recognition sites, thus affecting the ability to sustain noncovalent bonds. A pseudorotaxane is rotaxane's mechanically threaded supramolecular progenitor and can be reversibly dissociated into a free ring-type host and a free thread-type guest, thus giving rise to a dethreading/rethreading motion (**Figure 10c**).

Artificial molecular machines such as rotaxanes, catenanes, and pseudorotaxanes are integral components in developing functional nanosystems for nanotechnology. For example, as actuation materials, bistable rotaxanes have at least six advantages: (1) They can generate large strains up to 42%, while the strains generated by the gold-standard actuation materials, such as piezoelectric materials, are typically 0.1–0.2% (**Figure 12**) (Gad-El-Hak 2001, Kovacs 1998, Liu 2006, Madou 2002). (2) They have high force density, e.g., a bistable rotaxane generates a 100-pN force (Brough *et al.* 2006, Liu *et al.* 2005), while a kinesin biomotor, which is much larger than a bistable rotaxane, can only generate 6 pN (Howard 1997, Kull *et al.* 1996). (3) They can undergo controlled mechanical motion for various external stimuli (chemical (Bissell *et al.* 1994, Martínez-Díaz *et al.* 1997), electricity (Brouwer *et al.* 2001, Raehm *et al.* 1999), and light (Ashton *et al.* 1991)), while traditional actuation materials and biomotors must both rely on a single stimulus. (4) They can be

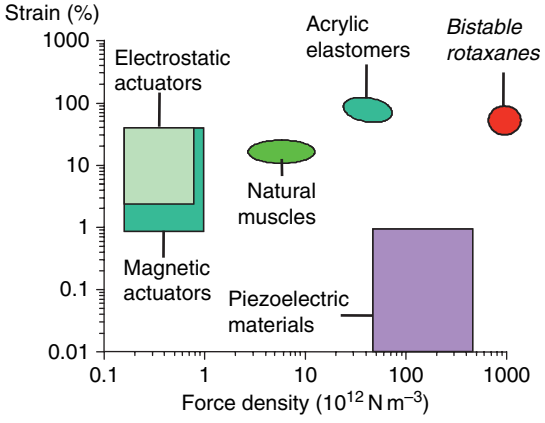


Figure 12 Comparison of different actuation materials.

made metastable, with the molecules persisting in their actuated state long after the stimulus has been removed, and thus saving energy. (5) They can be customized and optimized, therefore conferring the flexibility necessary for a multitude of engineering applications. For example, they can be derivatized with disulfide tethers to facilitate the formation of self-assembled monolayers (SAMs) (Tseng *et al.* 2004) or prepared with amphiphilic properties to facilitate the formation of Langmuir–Blodgett (LB) films (Lee *et al.* 2004, Petty 1996), both of which are key bottom-up nanomanufacturing technologies for

Table 1 Comparison between biomotors and bistable rotaxanes

	Biological molecular motors	*Bistable rotaxanes*
Surviving condition	Physiological conditions only (37°C, pH 7.0)	T −30 to +100°C, pH 5.0–9.0
Flexibility	Difficult to modify structures/functions	Easy to modify
Stimuli	Chemical only	Chemical, electricity, or light
Frequency	~1 Hz	Up to 1 MHz
Size (nm^3)	30–100	8
Force (pN)	~6 (kinesin)	40–100

the simultaneous self-organization of a multitude of molecules. (6) Bistable rotaxanes can survive in a wide range of temperatures (−30°C to 100°C) and pH values (5.0–9.0), while biomotors are restricted to physiological conditions ($T \sim 37°C$, $pH \sim 7.0$). These advantages indicate that although biomotors are perfect actuators in human bodies, artificial molecular motors such as bistable rotaxanes are more suitable for the development of molecular motor-based engineering applications (**Table 1**).

3.20.4 Moving Artificial Molecular Machines from Solution to Solid Supports

Thus far, it has been demonstrated in solution by numerous researchers that artificial molecular machines can exhibit substantial relative movements of their rings with respect to their dumbbell-shaped components (Balzani *et al.* 2000, 2003, Goodsell 1996). A solution, however, contains a large number of randomly distributed molecules that cannot be addressed individually or expressed coherently as they would be required to do in a device setting. In order to explore the full potential of these molecular-level machines and to develop functioning mechanical devices, molecules must be interfaced with the macroscopic world by imposing upon them a high degree of orientational control. In the ideal interface between artificial molecular machines and a solid support, the molecules would behave coherently, either in parallel or in series, and be addressable on the nanometer scale. More importantly, molecular machines' switching behavior demonstrated in solution still exists when molecules are in condensed phases on solid supports. This section will focus upon the author's and his colleagues' work on moving artificial molecular machines from solution state to solid state (Huang *et al.* 2004b, 2006).

3.20.4.1 Switching Artificial Molecular Machines in Solution

Since rotaxanes and catenanes were invented, UV/Vis spectroscopy has been established as an efficient tool to characterize the rotaxanes' chemical states in solution (Goodsell 1996). The UV/Vis spectrum (**Figure 13**, *curve a*) of a starting-state bistable rotaxane (**Figure 11**) revealed a charge transfer (CT) absorption band centered at 846 nm, characteristic of the translational isomer in which the TTF recognition site was encircled by the ring component. Furthermore, no absorption band was observed in the 500–560 nm region for the CT interaction that would result from another suspected translation isomer in which the DNP recognition site is encircled by the ring.

Chemical switching of the rotaxane molecular motor (8.0×10^{-4} M in acetonitrile) can also be monitored by UV/Vis spectroscopy (Tseng *et al.* 2003). $Fe(ClO_4)_3$ was used as the oxidant in MeCN. As shown in **Figure 13** (*curve b*), addition of the equivalent of one oxidant leads to a decrease in the intensity of the

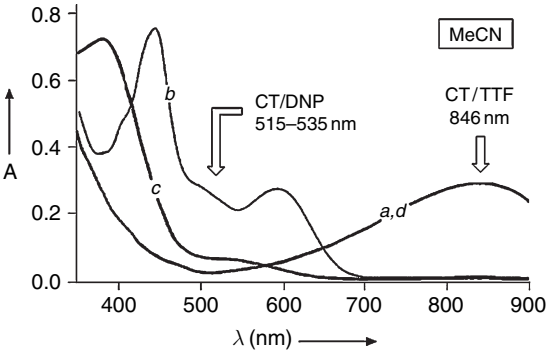


Figure 13 Absorption spectra (MeCN, 298 K) of a solution of bistable rotaxane **R1** (*curve a*) and of the same solution after the addition of 1 and 2 equivalents of $Fe(ClO_4)_3$ (*curves b* and *c*). Addition of ascorbic acid (2 equivalents) regenerates the original spectrum (*curve d*). (Source: Tseng H-R, Vignon S A, Stoddart J F 2003 Toward chemically controlled nanoscale molecular machinery. *Angew. Chem. Int. Ed.* **42**, 1491–5.)

CT band at 846 nm. Accompanying this decrease, an increase in the intensity of absorption bands centered at 450 and 600 nm, characteristic of the mono-oxidized form of the TTF unit, was observed. As a consequence of these spectral changes, a new CT interaction was observed at 515 nm as a shoulder-like absorption band. Further addition of the oxidant (up to 2 equivalents) caused the disappearance of the absorption band of mono-oxidized TTF and formation of a new band at 375 nm, which is attributed to the bis-oxidized TTF unit. A clear CT band (at 530 nm) corresponding to the encirclement of DNP ring system by the cyclophane ring can now be seen (**Figure 13**, *curve c*). Furthermore, addition of 2 equivalents of ascorbic acid as the reductant leads to restoration of the original spectrum (**Figure 12**, *curve d*).

3.20.4.2 Depositing Artificial Molecular Machines on Solid Substrates

We have used the LB technique (Petty 1996) to deposit artificial molecular machines onto solid substrates. In a typical LB experiment, amphiphilic molecules, with hydrophobic and hydrophilic ends, are dissolved in an organic solvent and the organic material is spread onto the air–water interface. After the organic solvent evaporates, a monolayer is left on the water surface. Next, moving barriers compress the films to a more organized and closely packed state. Films on the subphase, usually water, are called Langmuir monolayers, or Langmuir films. Monolayers transferred from the liquid–air interface to a solid substrate are called LB films. Atomic force microscopy (AFM) has been used to investigate the rotaxane LB film on solid supports. All monolayers were transferred to mica. As shown in **Figure 14**, domains formed at the air–water interface can be transferred to solid substrates, and at higher transfer pressure, domains are more compact. At surface pressures higher than 35 mN m^{-1} (Kovacs 1998), monolayers start to form folds in some regions and partial collapse can happen. AFM studies have confirmed that the optimal deposition pressure is around 30 mN m^{-1}.

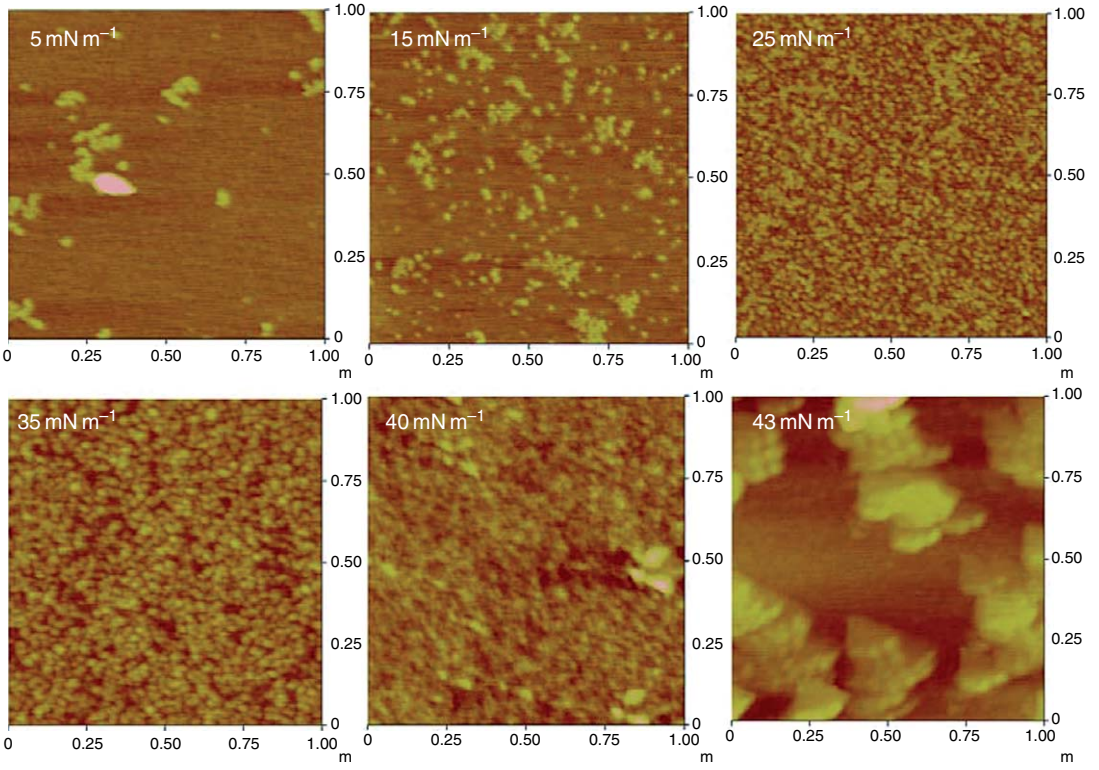


Figure 14 Atomic force microscopic (AFM) images of rotaxane Langmuir–Blodgett (LB) films on a mica surface at different surface pressures. (Source: Huang T J, Flood A H, Brough B, Liu Y, Bonvallet P A, Kang S, Chu C, Guo T F, Lu W X, Yang Y, Stoddart J F, Ho C M 2006 Understanding and harnessing biomimetic molecular machines for NEMS actuation materials. *IEEE Trans. Autom. Sci. Eng.* **3**, 254–9.)

3.20.4.3 Operating Artificial Molecular Machines in Condensed Phases on Solid Substrates

We have demonstrated, for the first time, from analyses of X-ray photoelectron spectra that redox-controllable molecular machines in the shape of amphiphilic, bistable rotaxanes are mechanically switchable in closely packed Langmuir films with chemical reagents. Most importantly, mechanical switching has been proven to occur in LB bilayers while mounted on solid substrates. The results not only constitute a proof-of-principle but also provide the impetus to continue forward and develop solid-state, artificial molecular machine-based NEMS.

3.20.4.3.1 Molecules' structural formulas

The structural formula and graphical representation of the amphiphilic, bistable rotaxane **R2** are shown in Figure 15. The molecule consists of a tetracationic ring component (dark blue) and a linear rod section that contains two different stations, a TTF unit (green) and a DNP unit (red), separated by a rigid

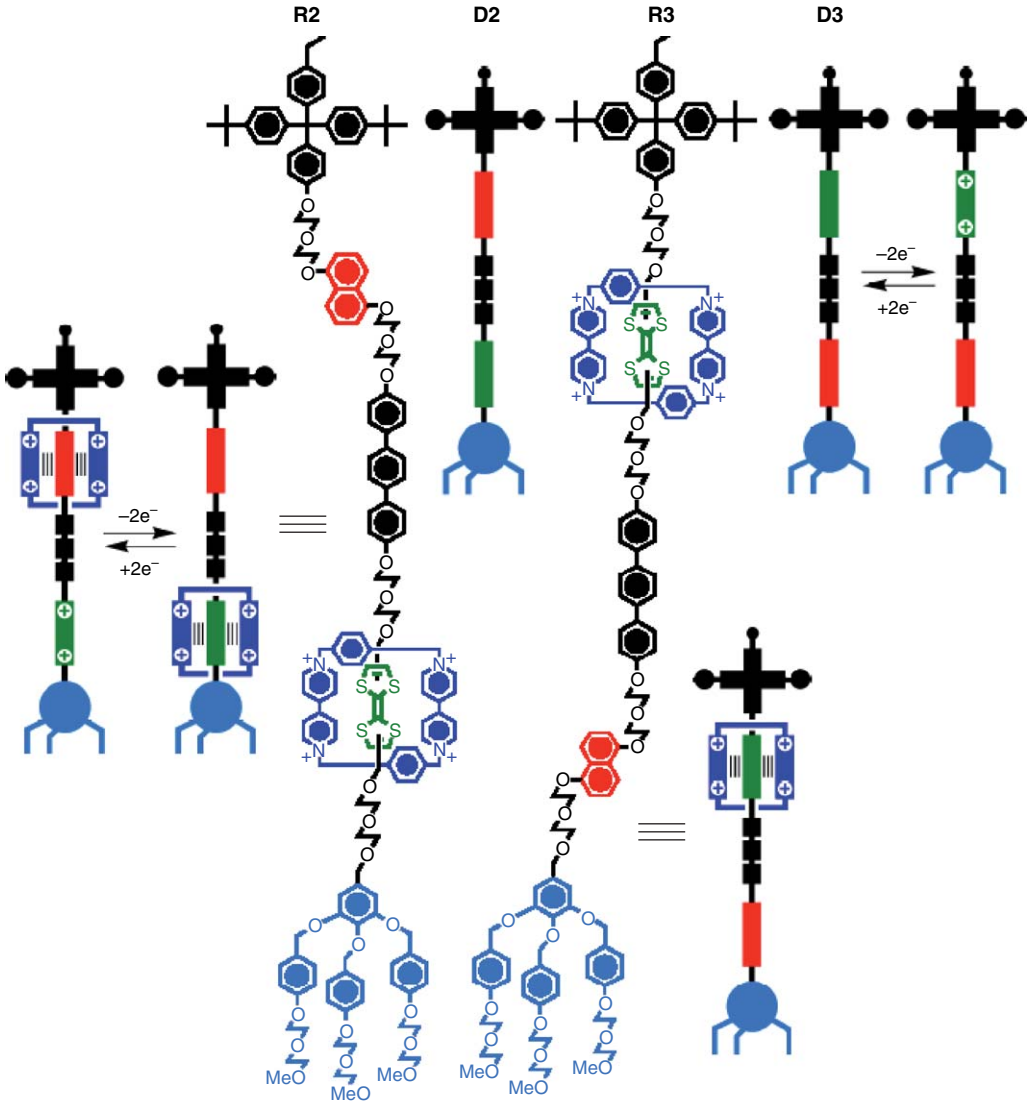


Figure 15 Structural formulas and idealized graphical representations of the rotaxane **R2** and control compounds – the dumbbell **D2** of **R2** and the constitutionally isomeric rotaxane **R3** and its dumbbell **D3**. (Source: Huang T J, Tseng H-R, Sha L, Lu W, Brough B, Flood A H, Yu B-D, Celestre P C, Chang J P, Stoddart J F, Ho C-M 2004b Mechanical shuttling of linear motor-molecules in condensed phases on solid substrates. *Nano Lett.* **4**, 2065–71.)

terphenylene spacer (black). A hydrophilic (light blue) and hydrophobic (black) stopper is incorporated at each end of the rod, enabling the molecules to form LB films. The starting-state **R2** of this bistable rotaxane has its TTF unit encircled by the tetracationic ring. Upon oxidation, the TTF unit becomes dicationic (TTF^{2+}) and experiences coulombic repulsion with the tetracationic ring, causing the ring to shuttle toward the DNP unit in the oxidized state. Conversely, the reduction of the TTF^{2+} dication back to a neutral TTF unit causes the rotaxane to return to its starting state. Thus, this rotaxane, with these two reversibly switchable redox states involving ring translation, can be likened to a linear molecular motor.

Figure 15 also shows the structural formulas and graphical representations of other compounds, the dumbbell **D2** of **R2** and the constitutionally isomeric rotaxane **R3** and its dumbbell **R3**, used in the control experiments. The difference between **R2** and **R3** relates only to the relative positioning of the TTF and DNP recognition sites with respect to the two different stoppers. The dumbbells **D2** and **D3** are both neutral compounds in which the rings have been removed from the rotaxanes **R2** and **R3**, respectively.

3.20.4.3.2 X-ray photoelectron spectroscopy (XPS) analysis

The LB technique of generating monolayer films allows the transfer of organized monolayers from an air–water interface to a solid substrate. It takes advantage of the amphiphilic nature of rotaxanes, allowing the construction of highly ordered molecular assemblies. In addition, artificial molecular machines in LB films are considered to be in a liquid-crystal-like environment, which maintains the mechanical switching properties seen in solution. The combination of the ordered structure in solid state and switching behavior in solution state makes the LB technique an ideal method for establishing the interface between rotaxane and solid supports.

XPS measurements (Compton 2001, Woodruff and Delchar 1994) have been carried out on Langmuir films of both amphiphilic, bistable rotaxanes (**R2** and **R3**) and their respective dumbbells (**D2** and **D3**). The XPS data were obtained on a VG ESCALAB electron spectrometer using A1 Kα X-ray radiation (1486.6 eV) with a pass-energy of 20 eV at a take-off angle of 90° (normal to the film surface). The photoemission of surface carbon (C 1s) at 284.6 eV was used as the reference binding energy. Langmuir monolayers and LB films were fabricated using a KSV 2000 Alternate Trough System. All monolayers were transferred to hydrophilic silicon dioxide substrates at a surface pressure of 25 mN m^{-1} as LB films. Before monolayer depositions, the silicon dioxide substrates were prepared as follows: (1) Piranha ($H_2SO_4/H_2O_2 = 5{:}1$) cleaning for 5 min, (2) deionized water cleaning for 5 min, and (3) air drying. Film thickness measurements were performed using a Gaertner L116B Ellipsometer with a refractive index of 1.46.

The photoemission intensity of each element depends on the depth at which the photoelectron is emitted and it attenuates exponentially with increased depth. Therefore, the XPS technique can be used to differentiate atoms at different depths within a film. The XPS signal intensity ($I_{A,Z}$) for a layer of atoms (A) buried at a distance Z underneath the film surface can be expressed as (Huang *et al.* 2004a):

$$I_{A,Z} = I_A^0 \cdot \exp\left(\frac{-Z}{\lambda(E_A)\cos\theta}\right)$$

where I_A^0 represents the signal intensity obtained from a layer of atoms located at the film surface, $\lambda(E_A)$ is the photoelectron attenuation length of A, which is a function of the kinetic energy of the emitted photoelectrons, and θ is the electron emission angle, which equals 0° in the present investigation. Thus, the photoemission intensity from a layer of atoms A embedded closer to the surface (smaller Z) will be stronger than that arising from a layer of the same atoms buried deeper (larger Z) in the film.

Molecular shuttling can be monitored by using XPS to track nitrogen (N), which is solely present on bistable rotaxane's only moving unit: the ring component. Langmuir monolayers were prepared for both the starting-state and oxidized rotaxanes (**R2** and **R3**), along with their dumbbell counterparts **D2** and **D3**, respectively. Next, monolayers were transferred to substrates as LB films for XPS analysis. **Figure 16a** shows the N 1s photoemission spectra of the starting-state and oxidized **R2**, along with its dumbbell counterparts **D2**. There is no nitrogen signal observed in the ringless dumbbell experiments, confirming that the peaks revealed in the **R2** scans are from the nitrogen atoms in the ring. Furthermore, a higher intensity of N 1s peak is observed for the oxidized rotaxane **R2** film than for the starting state,

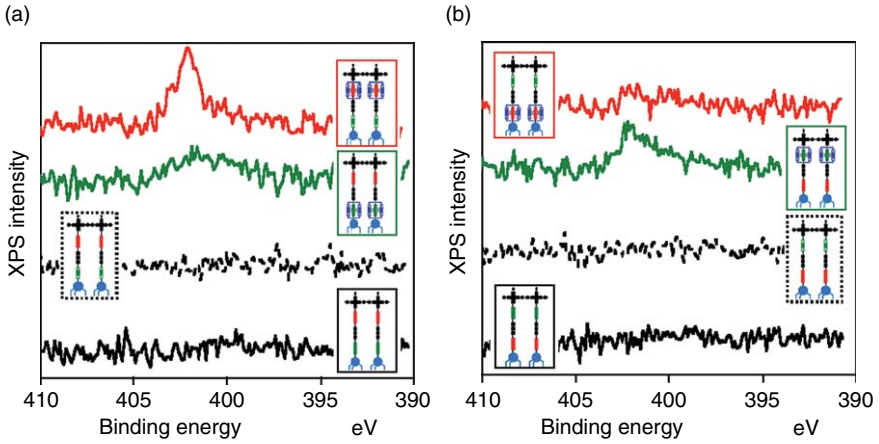


Figure 16 Nitrogen 1s X-ray photoemission spectra for (a) the starting-state rotaxane **R2** (solid green line) and its oxidized form (solid red line) and their respective dumbbell **D2** (black solid line) and its oxidized form (black dashed line) and (b) the starting-state rotaxane **R3** (solid green line) and its oxidized form (solid red line) and their respective dumbbells **D3** (black solid line) and its oxidized form (black dashed line). (Source: Huang T J, Tseng H-R, Sha L, Lu W, Brough B, Flood A H, Yu B-D, Celestre P C, Chang J P, Stoddart J F, Ho C-M 2004b Mechanical shuttling of linear motor-molecules in condensed phases on solid substrates. *Nano Lett.* **4**, 2065–71.)

an observation which supports the hypothesis that, in the oxidized molecule, the ring encircles the DNP recognition site, as opposed to the TTF^{2+} site, placing it closer to the film surface. Control experiments were conducted using the rotaxane **R3** in which the positions of the TTF and DNP units are the reverse of those in **R2**. **Figure 16b** shows clearly that the N 1s photoemission intensity was the highest from the starting-state rotaxane **R3** in which the ring encircles the TTF site, i.e., the one closer to the film surface. These control experiments ruled out the possibility of an oxidation-induced change in the N density and demonstrated the capability of XPS measurements to track the movement of the rings in LB films of redox-controllable rotaxanes.

Molecules of **R2** were *in situ* switched in highly packed Langmuir monolayers and transferred to substrates for XPS analysis. Once again, a much higher intensity of the N 1s peak is observed (**Figure 17a**) for the molecules *in situ* switched in highly packed Langmuir monolayers than that observed for the starting state of the bistable rotaxane, and the line

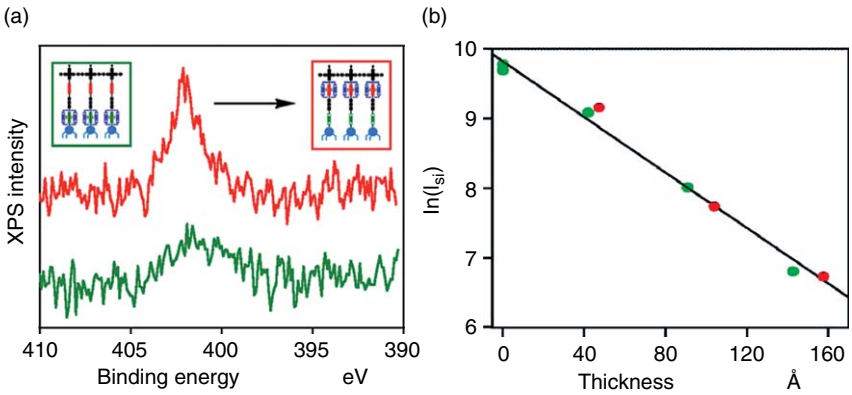


Figure 17 (a) The rotaxane **R2** (solid red line) having been switched (oxidized) in a highly packed monolayer compared with its starting-state form (solid green line). (b) The logarithms of Si 2p photoemission intensities versus film thickness measured by ellipsometry. The solid red dots represent samples composed of oxidized rotaxane molecules and the green solid dots represent samples composed of the starting-state molecules **R2**. (Source: Huang T J, Tseng H-R, Sha L, Lu W, Brough B, Flood A H, Yu B-D, Celestre P C, Chang J P, Stoddart J F, Ho C-M 2004b Mechanical shuttling of linear motor-molecules in condensed phases on solid substrates. *Nano Lett.* **4**, 2065–71.)

shape and intensity of N 1s remain similar to those recorded (**Figure 16a**) for the oxidized molecule. This datum proves that the molecules of the bistable rotaxanes are mechanically switchable in closely packed Langmuir films.

Next, the distance traveled by the ring component in the switchable rotaxane was quantified, thereby establishing that the ring moves completely from one station to the other while molecules are in closely packed Langmuir films. Quantitative XPS analysis indicated that the distance traveled by the ring in the *Z*-direction was 1.9 nm based on the data shown in **Figure 17b**. Molecular modeling indicates that the distance of the ring shuttling along the fully stretched 8.0 nm length of the rotaxane **R2** molecule is 3.7 nm, 46% of the total molecular length. Molecular modeling was performed in Chem3D by extending the molecule fully and then optimizing with the MM2 force field to obtain correct bond lengths and angles. The distance was measured from the center of the TTF unit to the center of the DNP unit. Since the film thickness was found to be only 4.4 nm by ellipsometry measurement, the translational value of 1.9 nm, or 42% of the total film thickness, is in good agreement with the modeling result.

To turn a rotaxane-based mechanical device into an engineering reality, it must be proven that these mechanically interlocked molecules maintain their shuttling properties while mounted on a solid substrate. Therefore, we extended our investigations into the realm of LB films. LB double layers of the starting-state rotaxane $\mathbf{3}^{4+}$ molecules were deposited on a silicon dioxide substrate at a surface pressure of 25 mN m^{-1} and the N 1s photoemission intensity was measured by XPS. The substrate coated with the LB double layers was then immersed in a solution containing oxidant and probed again by XPS. The oxidant solution was prepared by mixing $Fe(ClO_4)_3$ into deionized water to a concentration of 10^{-3} M. Typically, the sample was immersed in the oxidant solution for 30–40 min. Ellipsometry data show that the thickness of the films did not change while the samples were immersed in the oxidant solution. A double layer, instead of a single layer, was chosen because the surfaces of double-layer samples are hydrophilic, as opposed to the hydrophobic single layer, allowing the oxidant in an aqueous solution to penetrate the surface and oxidize the TTF unit. It is expected that only molecules in the top layer will be oxidized and switched. The observed difference (**Figure 18**) in the two N 1s spectra supports the hypothesis that mechanical switching, as characterized by ring movements in one of the two molecular layers, occurs within the film, even while the molecules are attached to the solid substrate.

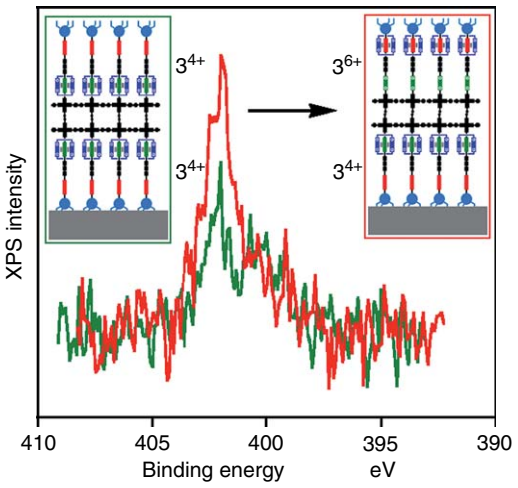


Figure 18 Nitrogen 1s photoemission spectra for the rotaxane **R3** in double Langmuir–Blodgett (LB) films prior to (solid green line) and after (solid red line) immersing samples into an aqueous solution of $Fe(ClO_4)_3$. (Source: Huang T J, Tseng H-R, Sha L, Lu W, Brough B, Flood A H, Yu B-D, Celestre P C, Chang J P, Stoddart J F, Ho C-M 2004b Mechanical shuttling of linear motor-molecules in condensed phases on solid substrates. *Nano Lett.* **4**, 2065–71.)

3.20.4.4 Remarks

Quantitative XPS analysis has proven that the ring component of a bistable amphiphilic rotaxane can be switched completely from one station to the other along the dumbbell component in self-organized, closely packed Langmuir monolayers. Moreover, XPS analysis also indicates that the mechanical operation of these linear motor molecules is maintained, even when they are closely packed within LB films mounted on solid supports. Isotherm analysis and other techniques employed also support this conclusion. This work marks a substantial step toward the development of motor molecule-based nanomechanical systems.

3.20.5 A Nanomechanical Device Based on Artificial Molecular Machines

Nanoscale actuators, capable of converting chemical or electrical energy into mechanical motion, are

needed for a wide range of applications. The bottom-up approach, which employs atoms and molecules both as the fundamental building blocks and as the working units, is potentially capable of delivering efficient operations at dramatically reduced scales in comparison with traditional microscale actuators. Artificial molecular machinery is an attractive means for performing controllable mechanical work at the nanoscale. Bistable rotaxanes hold particular promise in this regard. They have been likened to linear molecular motors on account of their ring component's ability to undergo controllable mechanical switching between two or more recognition sites via chemical, electrochemical, or photochemical stimulus. This section will describe the nanomechanical response of microcantilever beams coated with a SAM of artificial molecular motors following the cycled addition of chemical redox reagents (Huang *et al.* 2004a). A simple model that considers the mechanical movements of each molecule within the SAM verifies chemomechanical transduction as a likely mechanism for cantilever bending. In addition, this section will discuss the results of a structure–function study conducted through a series of control experiments and provide rational assessments to account for a wide range of alternative interaction mechanisms. The data provide compelling evidence for chemomechanical transduction as a general mode of operation for the generation of force from surface-bound linear molecular motors.

Expanding upon a series of bistable [2]rotaxanes, a bistable [3]rotaxane **R4** (**Figure 19**) was created as a molecular muscle to mimic the contraction and extension movements of skeletal muscle. This design takes advantage of well-established recognition chemistry that selectively positions the cyclobis(paraquat-*p*-phenylene) ($CBPQT^{4+}$) rings around the two TTF units of **R4**, as opposed to the two DNP units. Chemical oxidation of the TTF units to their dicationic forms (TTF^{2+}) drives the $CBPQT^{4+}$ rings to the DNP units. This power stroke arises primarily from electrostatic charge–charge repulsion between the $CBPQT^{4+}$ rings and the TTF^{2+} units. Upon reduction of the two TTF^{2+} stations back to their neutral form, the interring distance increases as the $CBPQT^{4+}$ rings return to the TTF stations by means of a thermally activated diffusive stroke. Thus, the cycle of contraction and extension within **R4** mimics the motion that takes place inside natural muscle fibers. The incorporation of a disulfide tether onto each $CBPQT^{4+}$ ring component provides an anchoring point by which the [3]rotaxane can be attached to a gold surface as a SAM. Oxidation of **R4** will generate a tensile stress upon a gold surface through the contractive action of its two disulfide-tethered $CBPQT^{4+}$ rings. If the substrate is sufficiently thin and flexible, such as a long cantilever beam, the cumulative effect of each individual molecular muscle will produce an upward mechanical bending of the beam. Correspondingly, reduction of the oxidized and contracted **R4** will return the $CBPQT^{4+}$ rings to the TTF stations and consequently relieve the stress upon the beam, resulting in a downward motion and a return to the beam's equilibrium position.

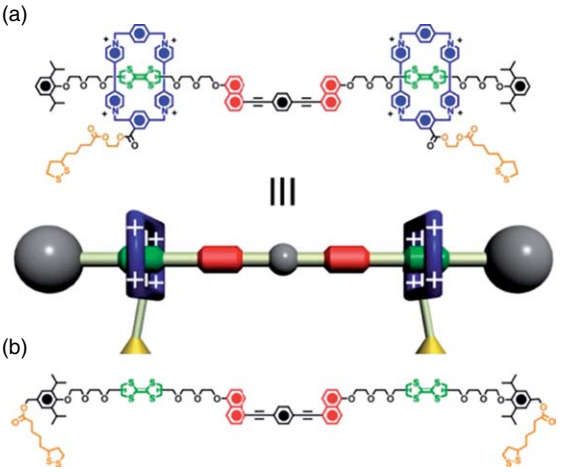


Figure 19 Molecular structure and schematic of (a) a bistable [3]rotaxane molecular muscle **R4** and (b) its dumbbell **D4**. (Source: Huang T J, Brough B, Ho C-M, Liu Y, Flood A H, Bonvallet P A, Tseng H-R, Stoddart J F, Baller M, Magonov S 2004a A nanomechanical device based on linear molecular motors. *Appl. Phys. Lett.* **85**, 5391–3.)

A single beam's deflection upon contraction of the molecular muscle was analyzed using a simple model for the [3]rotaxane **R4** bound to a section of the beam (**Figure 20a**). Oxidation produces a contraction of the interring distance and correspondingly exerts a force F plus a bending moment M upon the beam. Only the sections of the beam between the two moving rings will be subjected to the action of the bending moment M, whereas the other sections have a zero bending moment. The out-of-plane displacement w of a cantilever beam (**Figure 20b**) is governed by the Euler–Bernoulli beam equation

$$w = \frac{M_{\text{beam}} L^2}{2EI}$$

in which M_{beam} is the moment on the beam, L the total length of the cantilever beam, E the Young's modulus of the cantilever, and I the area moment of

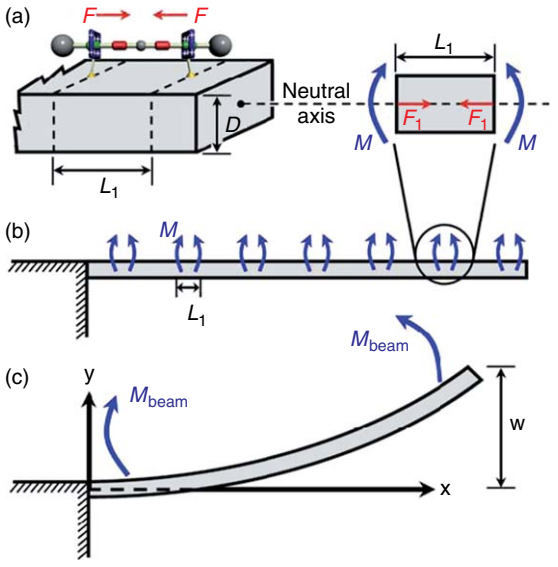


Figure 20 Free-body beam analysis. (a) Free-body diagram showing the contractive force *F* exerted within a single functional unit of the beam (*left*), its geometric translation **F1** to the neutral axis, and its associated upward bending moment *M* (*right*). (b) Overall diagram demonstrating the collective upward bending moments exerted by each of the functional units of the cantilever beam. (c) The bending moment M_{beam} for the entire cantilever beam produces an out-of-plane displacement *w*. (Source: Huang T J, Brough B, Ho C-M, Liu Y, Flood A H, Bonvallet P A, Tseng H-R, Stoddart J F, Baller M, Magonov S 2004a A nanomechanical device based on linear molecular motors. *Appl. Phys. Lett.* **85**, 5391–3.)

inertia of the beam's cross section. The parameter M_{beam} can be obtained by assuming that the force generated (40 pN) by a single molecule is solely due to electrostatic effects ($\varepsilon_{water} = 80$), the SAM covers 100% of the gold surface, and the molecules are idealized as randomly oriented noninteracting rigid rods. Based on this simplified model, the force generated by the molecular muscle can act against the spring-like restoring force of a cantilever beam (500 μm × 100 μm × 1 μm) to produce a theoretical beam displacement *w* of 48 nm.

This chemomechanical design was tested with a gold-coated silicon cantilever array that was coated with a SAM of **R4** (Figure 21a) and placed in a transparent fluid cell. The silicon cantilever arrays employed in this experiment incorporate cantilever beams with a length of 500 μm, a width of 100 μm, a thickness of 1 μm, and a spring constant of 0.02 N m^{-1}. They were fabricated using silicon-on-insulator (SOI) wafers. The gold coatings were prepared by thermally evaporating 1.5 nm of Ti, followed by 20 nm of Au onto the topside of the cantilever devices. Immediately afterward, the cantilever arrays were immersed in a 0.1 mM MeCN solution of the disulfide-tethered palindromic [3]rotaxane for 48 h in order to form a SAM of the compound. Following SAM formation, the cantilevers were rinsed with MeCN to remove any unbound material.

The experimental setup (Figure 22) is based on a Digital Instruments Scentris™ Research Tool. The cantilever array is housed in a 50-μl sample chamber covered by a transparent window in order to record the cantilevers' deflections by the beam-deflection method. Light from multiple superluminescent diodes emitting at 850 nm is coupled to an array of multimode fibers and projected onto the end of a single cantilever. Upon reflection, the light is collected by a linear position-sensitive detector (PSD), and the photocurrents are converted into voltages. The position of the beam is determined within ±0.5 nm or better. After amplification, the signals are digitized and stored. A time-multiplexing scheme is applied by switching individual light sources on and off to illuminate only one cantilever at a time. Continuous flow driven by gravitational force propels solutions of chemical reagents from different reservoirs through the sample chamber into a waste container at rates of 250–300 μL min^{-1}. The sample reservoirs were connected to a computer-controlled six-way valve in order to minimize the disturbance of the flow when switching between the different solutions. Addition of the oxidant solution was observed to cause the cantilever beams to bend to an upward limit of ~35 nm (Figure 21b, top series of traces). The entry of the reductant solution caused the beams to bend back downward to their starting positions. This behavior was observed for all four cantilever beams for 25 cycles (the first three complete cycles are shown here). The slight attenuation in beam deflection following each cycle is attributed to a gradual chemical and/or physical passivation of the SAM. Nevertheless, the movement of the cantilever beams is directly correlated with the cycling of the oxidant and reductant solutions and the experimental data (35 nm displacement) matches closely with the theoretical quantitative analysis (48 nm displacement).

In order to assert that the beams' bending is not a consequence of mundane conformational and/or electrostatic changes within the semirigid **R4** backbone, the disulfide-tethered dumbbell compound **D4** (Figure 19b) was synthesized. This control

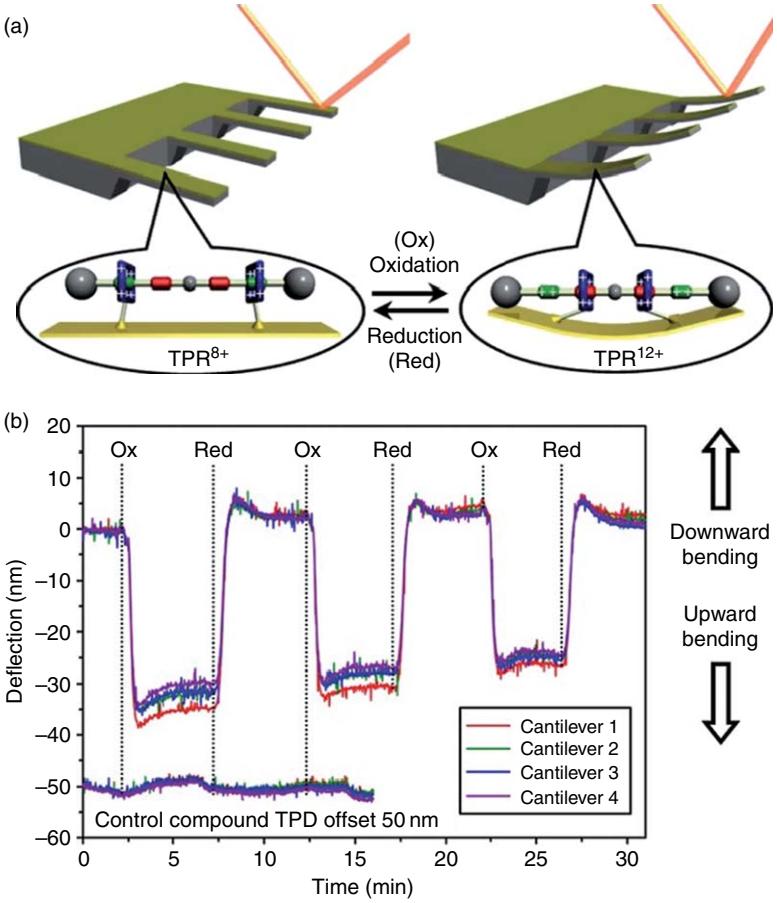


Figure 21 (a) Schematic of the proposed mechanism of the device operation. (b) The experimental data show bending of the four cantilever beams, as the aqueous oxidant (Ox) and reductant (Red) solutions are delivered to the sample cell. The top series of traces shows the deflection of the cantilever array coated with **R4**, whereas the bottom series of traces (offset by 50 nm) shows the limited movement of the cantilever array coated with the dumbbell control compound **D4**. (Source: Huang T J, Brough B, Ho C-M, Liu Y, Flood A H, Bonvallet P A, Tseng H-R, Stoddart J F, Baller M, Magonov S 2004a A nanomechanical device based on linear molecular motors. *Appl. Phys. Lett.* **85**, 5391–3.)

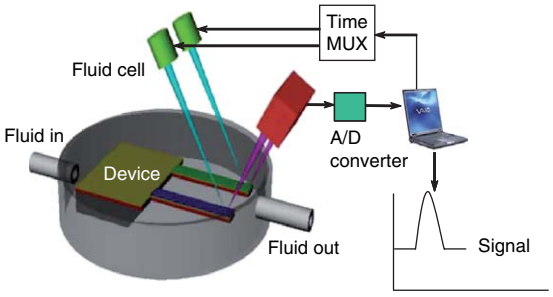


Figure 22 Schematic of the experimental setup. (Source: Huang T J, Brough B, Ho C-M, Liu Y, Flood A H, Bonvallet P A, Tseng H-R, Stoddart J F, Baller M, Magonov S 2005 A nano-chemo-mechanical actuator based on artificial molecular machines. *MEMS 2005*, Miami, FL, USA, January 30–February 3, 2005, pp. 871–4.)

compound contains pairs of TTF and DNP recognition sites with the same relative geometries that are present in **R4**. However, **D4** lacks the mechanically mobile $CBPQT^{4+}$ rings and the disulfide tethers are attached at different locations on the dumbbell's two stoppers. An array of cantilever beams coated with a SAM of **D4** bends only slightly (**Figure 21b**, bottom series of traces) following sequential injections of the same oxidant and reductant solutions used to study **R4**. This observation suggests that the presence of the mechanically active, disulfide-tethered $CBPQT^{4+}$ rings in $\mathbf{R1}^{8+}$ is essential for the redox-controlled bending of the cantilever beams.

Furthermore, control experiments and quantitative analysis (supporting online text) have established that pH variations (**Figure 23**), electrostatic charge

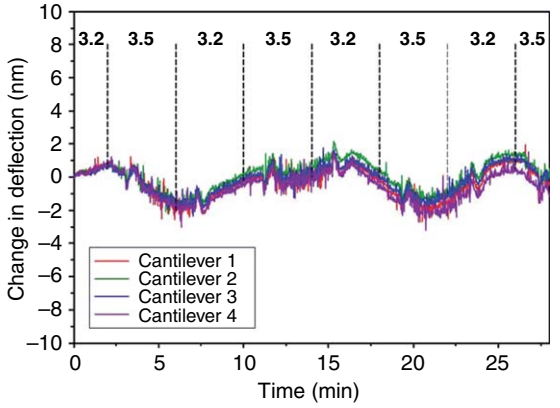


Figure 23 Control experiment demonstrating that the cantilever beam coated with the [3]rotaxane molecular muscle **R4** does not display any significant deflection when exposed to the same variation in pH (3.2 and 3.5) that accompanies the addition of oxidant and reductant, respectively. (Source: Huang T J, Brough B, Ho C-M, Liu Y, Flood A H, Bonvallet P A, Tseng H-R, Stoddart J F, Baller M, Magonov S 2004a A nanomechanical device based on linear molecular motors. *Appl. Phys. Lett.* **85**, 5391–3.)

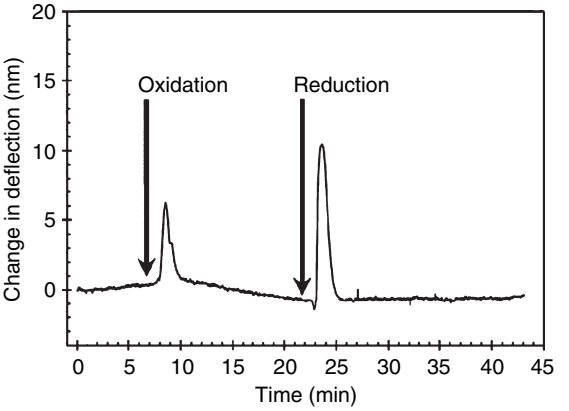


Figure 24 Control experiment demonstrating that refractive index has very limited effects on the cantilever beams' deflection. (Source: Huang T J, Brough B, Ho C-M, Liu Y, Flood A H, Bonvallet P A, Tseng H-R, Stoddart J F, Baller M, Magonov S 2004a A nanomechanical device based on linear molecular motors. *Appl. Phys. Lett.* **85**, 5391–3.)

repulsion, thermal contributions, and photothermal differences have an inconsequential effect on the bending of cantilever beams coated with SAMs of the [3]rotaxane molecular muscle. Changes in solution pH, for example, have been demonstrated to cause silicon cantilevers to bend (Fritz *et al.* 2000). A control experiment (**Figure 23**) with citrate buffer solutions, which have the same pH as the oxidant and reductant solutions (3.2 and 3.5, respectively), establishes that pH variations have an insignificant effect on the bending of the cantilever beams that are coated with the SAM of the [3]rotaxane molecular muscle. Furthermore, the direction of beam bending does not correlate in any repeatable way with the pH of the surrounding solution. Control experiments conducted in solution also suggest that the [3]rotaxane molecular muscle is stable in acidic solutions for extended periods of time, indicating that pH-controlled chemical decomposition does not affect significantly the molecular SAM that coats the cantilever beams. Specifically, the prototypical [3]rotaxane molecular muscle displays no apparent UV/Vis and ^{1}H NMR spectroscopic changes following exposure to trifluoroacetic acid over a 2-week period.

Another control experiment was conducted to rule out the effect of any change in the reflective index (**Figure 24**). Instead of focusing the laser beam on the tip of the cantilever, it was focused on the unbendable substrate. As different solutions were introduced into the sample chamber, only a change in the reflective index of the different solutions will cause a variation in the signal output. As shown in **Figure 24**, this variation has a minimal effect when changing between two solutions. After the flow has stabilized, there is very little difference in the signals between the addition of water, oxidant, or reductant.

Other factors such as charge repulsion and thermal effects may also contribute to beam bending (Varesi *et al.* 1997). Upon oxidation, the electronic charge of the disulfide-tethered palindromic [3]rotaxane molecular muscle changes from +8 to +12, giving rise to a change in the electrostatic interactions between neighboring surface-bound [3]rotaxane molecules. The proximity of two like-charged molecules directly on the cantilever surface could create a repulsive force, thus leading to the beam bending downward. Yet this is contrary to what is observed. Similarly, when considering thermal effects, the release of energy upon oxidation of the [3]rotaxane would cause the bimorph-like beams to bend downward on account of the higher thermal expansion coefficient of the upper gold surface in comparison with the lower silicon surface (Moulin *et al.* 1997). However, subtle thermal effects are negligible in aqueous solutions because of the high heat capacity of water. Finally, the photothermal bending effect (Barnes *et al.* 1994) initiated by the 850-nm detection laser beam is discounted, both on account of the water's high heat capacity and by the fact that the incident laser light with a spot size of 3 μm irradiates an insignificant amount (0.02%) of the cantilever's total surface area.

Given that all these factors do not contribute to a concerted motion of the cantilever, the deflections of the beams coated with SAMs of [3]rotaxane molecular muscle are consistent with our assignment of the movement to the redox-driven changes in the molecular monolayer. This result correlates directly with the collective effect of molecular-scale contractions and extensions of the interring distances in the disulfide-tethered palindromic [3]rotaxane molecules and supports the hypothesis that the cumulative nanoscale movements within surface-bound molecular muscles can be harnessed to perform larger-scale mechanical work.

It appears that **Figure 21** also indicates that as the oxidation–reduction cycle is repeated, the magnitude of the beam bending decreases. A 25-cycle experiment confirms this observation. As seen in **Figure 25**, the magnitude of the beam deflection decreases from 35 nm in cycle 1 to 3 nm in cycle 25. Interestingly, this exponential decrease in the cantilever's deflection matches qualitatively with a separately observed trend in which rotaxane degrades in solution following multiple oxidation–reduction cycles (**Figure 26**). The solution-phase decay was estimated from the percent recovery of the original UV/Vis absorption intensity of rotaxanes after one oxidation–reduction cycle. This correlation suggests that chemical degradation is probably responsible for the gradual decrease in the magnitude of the beams' bending. In order to develop more efficient and practical molecular machine-based nanomechanical actuators, electricity or light could be used instead of chemicals as a switching stimulus. This modification would allow molecules to remain in an environment without air, water, or solvents, thereby increasing the system's lifetime.

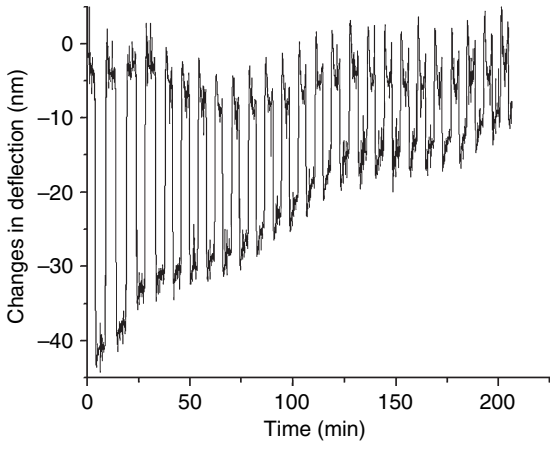


Figure 25 A 25-cycle experimental data showing that as the oxidation–reduction cycle is repeated, the magnitude of the beam bending decreases. (Source: Liu Y, Flood A H, Bonvallet P A, Vignon S A, Northrop B H, Tseng H-R, Jeppesen J O, Huang T J, Brough B, Baller M, Magonov S, Solares S D, Goddard W A, Ho C-M, Stoddart J F 2005 Linear artificial molecular muscles. *J. Am. Chem. Soc.* **127**, 9745–59.)

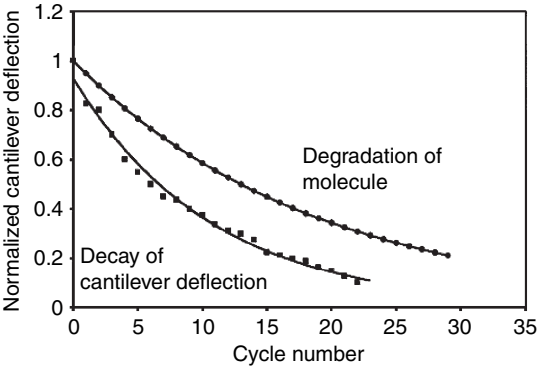


Figure 26 Normalized graph showing the qualitative correlation between the attenuation of the beam's deflection and the degradation of rotaxane.

3.20.6 Conclusions

Inspired by biological molecular machines, we are developing a new class of mechanical nanodevices that employ a group of artificial molecular machines called rotaxanes. As a substantial step toward this long-term objective, we have proven for the first time that rotaxanes are mechanically switchable in condensed phases on solid substrates. We have further developed a rotaxane-powered microcantilever actuator utilizing an integrated approach that combines bottom-up assembly of molecular functionality with top-down micro/nanofabrication. By harnessing the nanoscale mechanical motion from artificial molecular machines and eliciting a nanomechanical response in a microscale device, this system mimics natural skeletal muscle. We have every reason to believe that the chemically driven redox process we demonstrated thus far can be replaced with direct electrical or optical stimulation, a logical development that would establish a technological basis for the production of a new class of multiscale NEMS devices based on molecular mechanical motion.

Acknowledgments

This work was supported in part by the NSF NIRT grant (ECCS-0609128) and the start-up fund provided by the ESM Department, College of Engineering, Materials Research Institute, and Huck Institutes for the Life Sciences at The Pennsylvania State University. The author wishes to thank Xiaole Mao, Dana Mallen, John Waldeisen, and Jinjie Shi for their help with the manuscript.

References

Allison W S 1998 F_1-ATPase: A molecular motor that hydrolyzes ATP with sequential opening and closing of catalytic sites coupled to rotation of its subunit. *Acc. Chem. Res.* **31**, 819–26

Anelli P-L, Spencer N, Stoddart J F 1991 A molecular shuttle. *J. Am. Chem. Soc.* **13**, 5131–3

Anelli P-L, Asakawa M, Ashton P R, Bissell R A, Clavier G, Górski R, Kaifer A E, Langford S J, Mattersteig G, Menzer S, Philp D, Slawin A M Z, Spencer N, Stoddart J F, Tolley M S, Williams D J 1997 Toward controllable molecular shuttles. *Chem. Eur. J.* **3**, 1113–15

Ashton P R, Philp D, Spencer N, Stoddart J F 1991 The self-assembly of [n]pseudorotaxanes. *J. Chem. Soc. Chem. Commun.* **23**, 1677–9

Balzani V, Credi A, Raymo F M, Stoddart J F 2000 Artificial molecular machines. *Angew. Chem. Int. Ed.* **39**, 3348–91

Balzani V, Credi A, Venturi M 2003 *Molecular Devices and Machines – A Journey into the Nano World*. Wiley-VCH, Weinheim

Barnes J R, Stephenson R J, Welland M E, Gerber Ch, Gimzewski J K 1994 Photothermal spectroscopy with femtojoule sensitivity using a micromechanical device. *Nature* **372**, 79–81

Bissell R A, Córdova E, Kaifer A E, Stoddart J F A 1994 Chemically and electrochemically switchable molecular shuttle. *Nature* **369**, 133–7

Block S M 1997 Real engines of creation. *Nature* **386**, 217–19

Boyer P D 2002 A research journey with ATP synthase. *J. Biol. Chem.* **277**, 39045–61

Brough B, Northrop B H, Schmidt J J, Tseng H-R, Kendall N H, Stoddart J F, Ho C-M 2006 Evaluation of synthetic linear motor-molecule actuation energetics. *Proc. Natl. Acad. Sci. U S A* **103**, 8583–8

Brouwer A M, Frochot C, Gatti F G, Leigh D A, Mottier L, Paolucci F, Roffia S, Wurpel G W H 2001 Photoinduction of fast, reversible, translational motion in a hydrogen-bonded molecular shuttle. *Science* **291**, 2124–8

Compton P J 2001 Estimation of inelastic mean free paths for polymers and other organic materials: Use of quantitative structure–property relationships. *Surf. Interface Anal.* **31**, 23–34

Cross R, Duncan T 1996 Subunit rotation in F_0F_1-ATP synthases as a means of coupling proton transport through F_0 to the binding changes in F_1. *J. Bioenerg. Biomembr.* **28**, 403–8

Elston T, Wang H, Oster G 1998 Energy transduction in ATP synthase. *Nature* **391**, 510–13

Feynman R P 1960 There is plenty of room at the bottom. *Eng. Sci.* **23**, 22–36

Fritz J, Baller M K, Lang H P, Strunz T, Meyer E, Güntherodt H-J, Delamarche E, Gerber Ch, Gimzewski J K 2000 Stress at the solid–liquid interface of self-assembled monolayers on gold investigated with a nanomechanical sensor. *Langmuir* **16**, 9694–9696

Gad-El-Hak M 2001 *The MEMS Handbook*. CRC Press, Boca Raton

Goodsell D S 1996 *Our Molecular Nature: The Body's Motors, Machines, and Messages*. Copernicus, New York

Hackney D D 1996 The kinetic cycles of myosin, kinesin, and dynein. *Annu. Rev. Physiol.* **58**, 731–50

Hawkins J M 1979 *The Oxford Paperback Dictionary*. Oxford University Press, Oxford

Hess H, Howard J, Vogel V 2002 A piconewton forcemeter assembled from microtubules and kinesins. *Nano Lett.* **2**, 1113–15

Hiratsuka Y, Tada T, Oiwa K, Kanayama T, Uyeda T Q P 2001 Controlling the direction of kinesin-driven microtubule movements along microlitho-graphic tracks. *Biophys. J.* **81**, 1555–61

Howard J 1997 Molecular motors: Structural adaptations to cellular functions. *Nature* **389**, 561–7

Howard J, Hudspeth A J, Vale R D 1989 Movement of microtubules by single kinesin molecules. *Nature* **342**, 154–8

Huang T J, Brough B, Ho C-M, Liu Y, Flood A H, Bonvallet P A, Tseng H-R, Stoddart J F, Baller M, Magonov S 2004a A nanomechanical device based on linear molecular motors. *Appl. Phys. Lett.* **85**, 5391–3

Huang T J, Tseng H-R, Sha L, Lu W, Brough B, Flood A H, Yu B-D, Celestre P C, Chang J P, Stoddart J F, Ho C-M 2004b Mechanical shuttling of linear motor-molecules in condensed phases on solid substrates. *Nano Lett.* **4**, 2065–71

Huang T J, Brough B, Ho C-M, Liu Y, Flood A H, Bonvallet P A, Tseng H-R, Stoddart J F, Baller M, Magonov S 2005 A nano-chemo-mechanical actuator based on artificial molecular machines. *MEMS 2005*, Miami, FL, USA, January 30–February 3, 2005, pp. 871–4

Huang T J, Flood A H, Brough B, Liu Y, Bonvallet P A, Kang S, Chu C, Guo T F, Lu W X, Yang Y, Stoddart J F, Ho C M 2006 Understanding and harnessing biomimetic molecular machines for NEMS actuation materials. *IEEE Trans. Autom. Sci. Eng.* **3**, 254–9

Kovacs G T A 1998 *Micromachined Transducers Sourcebook*. McGraw-Hill, New York

Kull F J, Sablin E P, Lau R, Fletterick R J, Vale R D 1996 Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature* **380**, 550–5

Lee I C, Frank C W, Yamamoto T, Tseng H-R, Flood A H, Stoddart J F, Jeppesen J O 2004 Langmuir and Langmuir–Blodgett films of amphiphilic bistable rotaxanes. *Langmuir* **20**, 5809–28

Limberis L, Stewart R J 2000 Toward kinesin-powered microdevices. *Nanotechnology* **11**, 47–51

Lipowsky R, Klumpp S 2005 Life is motion: Multiscale motility of molecular motors. *Physica A* **352**, 53–112

Liu C 2006 *Foundations of MEMS*. Pearson Education, New Jersey

Liu Y, Flood A H, Bonvallet P A, Vignon S A, Northrop B H, Tseng H-R, Jeppesen J O, Huang T J, Brough B, Baller M, Magonov S, Solares S D, Goddard W A, Ho C-M, Stoddart J F 2005 Linear artificial molecular muscles. *J. Am. Chem. Soc.* **127**, 9745–59

Madou M 2002 *Fundamentals of Microfabrication*. CRC Press, New York

Martínez-Díaz M-V, Spencer N, Stoddart J F 1997 The self-assembly of a switchable [2]rotaxane. *Angew. Chem. Int. Ed. Engl.* **36**, 1904–7

Moulin M, Stephenson R J, Welland M E 1997 Micromechanical thermal sensors: Comparison of experimental results and simulations. *J. Vac. Sci. Technol. B* **15**, 590–6

Noji H, Yasuda R, Yoshida M, Kinosita K 1997 Direct observation of the rotation of F_1-ATPase. *Nature* **386**, 299–302

Petty M C 1996 *Langmuir–Blodgett Films: An Introduction*. Cambridge University Press, Cambridge, UK

Platt M, Muthukrishnan G, Hancock W O, Williams M E 2005 Millimeter scale alignment of magnetic nanoparticle functionalized microtubules in magnetic fields. *J. Am. Chem. Soc.* **127**, 15686–7

Raehm L, Kern J M, Sauvage J-P 1999 A transition metal containing rotaxane in motion: Electrochemically induced pirouetting of the ring on the threaded dumbbell. *Chem. Eur. J.* **5**, 3310–17

Shinkai S, Manabe O 1984 Photocontrol of ion extraction and ion transport by photofunctional crown ethers. *Top. Curr. Chem.* **121**, 67–104

Soong R K, Bachand G D, Neves H P, Olkhovets A G, Craighead H G, Montemagno C D 2000 Powering an inorganic nanodevice with a biomolecular motor. *Science* **290**, 1555–8

Stoddart J F, Tseng H-R 2002 Chemical synthesis gets a fillip from molecular recognition and self-assembly processes. *Proc. Natl. Acad. Sci. U S A* **99**, 4797–800

Tseng H-R, Vignon S A, Stoddart J F 2003 Toward chemically controlled nanoscale molecular machinery. *Angew. Chem. Int.* **42**, 1491–5

Tseng H-R, Wu D, Fang N, Zhang X, Stoddart J F 2004 The metastability of an electrochemically controlled nanoscale machine on gold surfaces. *Chem. Phys. Chem.* **5**, 111–16

Uptain S, Kane C, Chamberlin M 1997 Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.* **66**, 117–72

Vall R D, Milligan R A 2000 The way things move: Looking under the hood of molecular motor proteins. *Science* **288**, 88–95

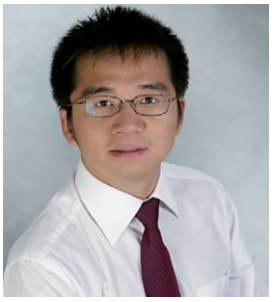
Van den Heuvel M G L, De Graaff M P, Dekker C 2006 Molecular sorting by electrical steering of microtubules in kinesin-coated channels. *Science* **312**, 910–14

Varesi J, Lai J, Perazzo T, Shi Z, Majumdar A 1997 Photothermal measurements at picowatt resolution using uncooled micro-optomechanical sensors. *Appl. Phys. Lett.* **71**, 306–8

Woodruff D P, Delchar T 1994 *A Modern Techniques of Surface Science*. Cambridge University Press, Cambridge, UK

Yin H, Wang M D, Svoboda K, Landick R, Block S M, Gelles J 1995 Transcription against an applied force. *Science* **270**, 1653–7

Biography



Tony Jun Huang received his Ph.D. degree in Mechanical and Aerospace Engineering from the University of California, Los Angeles (UCLA) in 2005, and his B.S. and M.S. degrees in Energy and Power Engineering from Xi'an Jiaotong University, Xi'an, China, in 1996 and 1999, respectively. He is the James Henderson Assistant Professor in the Department of Engineering Science and Mechanics at The Pennsylvania State University. His research interests include Biomedical NanoElectroMechanicalSystems (BioNEMS), Molecular Mechanics, Nanomaterials/Nanodevices, and Microfluidics. During his young career, he has already received many awards and honors including the 2006 Rustum and Della Roy Innovation in Materials Research Award, by *The Pennsylvania State University*; 'One of the five most intriguing articles in the third quarter of 2005', by *CAS Science Spotlight*; the James Henderson Endowed Professorship, by *The Pennsylvania State University*; Outstanding Ph.D. Award, by *UCLA School of Engineering.*